Quantum-Inspired Spectral-Spatial Pyramid Network for Hyperspectral Image Classification

Jie Zhang[1], Yongshan Zhang[2,1*], Yicong Zhou[1*]
[1] Department of Computer and Information Science, University of Macau, Macau, China
[2] School of Computer Science, China University of Geosciences, Wuhan, China
jiezh1997@gmail.com, yszhang.cug@gmail.com, yicongzhou@um.edu.mo

Abstract

Hyperspectral image (HSI) classification aims at assigning a unique label for every pixel to identify categories of different land covers. Existing deep learning models for HSIs are usually performed in a traditional learning paradigm. Being emerging machines, quantum computers are limited in the noisy intermediate-scale quantum (NISQ) era. The quantum theory offers a new paradigm for designing deep learning models. Motivated by the quantum circuit (QC) model, we propose a quantum-inspired spectral-spatial network (QSSN) for HSI feature extraction. The proposed QSSN consists of a phase-prediction module (PPM) and a measurement-like fusion module (MFM) inspired from quantum theory to dynamically fuse spectral and spatial information. Specifically, QSSN uses a quantum representation to represent an HSI cuboid and extracts joint spectral-spatial features using MFM. An HSI cuboid and its phases predicted by PPM are used in the quantum representation. Using QSSN as the building block, we further propose an end-to-end quantum-inspired spectral-spatial pyramid network (QSSPN) for HSI feature extraction and classification. In this pyramid framework, QSSPN progressively learns feature representations by cascading QSSN blocks and performs classification with a softmax classifier. It is the first attempt to introduce quantum theory in HSI processing model design. Substantial experiments are conducted on three HSI datasets to verify the superiority of the proposed QSSPN framework over the state-of-the-art methods.

1. Introduction

Hyperspectral images (HSIs) are always represented by three-order tensors that collected by remote sensors to record the characteristics reflected by land covers. Each HSI contains two spatial dimensions and one spectral dimension to reflect abundant spectral information and spatial context, making it different from color images. HSI classification aims to determine land-cover categories of areas represented by every pixel according to rich spatial and spectral information [21]. It has a broad range of applications, including target detection [42], mining [1] and agriculture [8]. Based on the recorded HSIs, samples in different categories may be extremely imbalanced in some data sets. This makes HSI classification very challenging [20, 49].

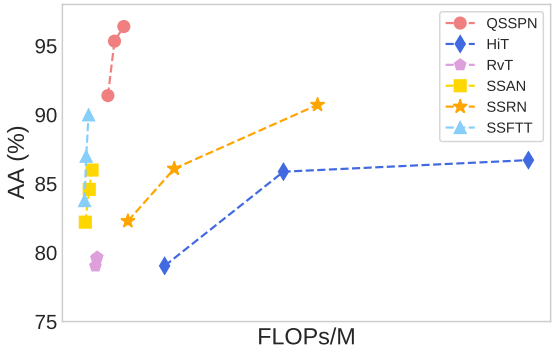


Figure 1. Performance comparison between the proposed QSSPN framework and existing methods. Average accuracies on Indian Pines dataset are presented.

In recent decades, a large number of HSI classification models have been proposed. Existing models can be divided into traditional learning methods and deep learning methods. Traditional classification models are simply adopted for HSIs, such as support vector machine (SVM) [27] and K-nearest neighbours (KNN) [19]. To achieve better results, advanced traditional learning models, e.g., extended morphological profile (EMP) [3] and extended multiattribute profile (EMAP) [7], first conduct feature extraction and then perform classification. These methods are two-step learning models that cannot achieve satisfactory performance. Deep learning models are promising to simultaneously conduct feature extraction and classification

*Corresponding authors.

in an end-to-end manner [12, 34]. Benefiting from the developed technique, deep learning models for HSI classification are in blossom. A wide variety of models rely on convolutional neural networks (CNNs). Most existing models are designed to separately extract spatial and spectral features in different branches [18, 47]. To enrich the development of deep learning models, attention mechanic [26, 46], graph convolution network [45] and transformer learning [36, 43, 48] have been explored in recent works. These works make significant improvement in architecture design and classification performance. However, they separately aggregate spatial and spectral information in different modules, resulting in model redundancy and inefficient exploration of the correlation between spatial and spectral information. Besides, they are all developed from traditional learning paradigm.

Different from traditional computers, quantum computers are emerging machines to perform quantum algorithms [2]. Quantum computing utilizes quantum theory to process data in quantum device [14, 28]. There are several models of quantum computation, such as quantum circuits, quantum annealing and adiabatic quantum computation. It has been proved that quantum computers outperform classical computers in solving certain problems [11, 33]. For example, Shor's algorithm can be used to solve the integer factorization problem much faster than algorithms running on the classical computer [33]. In the noisy intermediate-scale quantum (NISQ) era [31], quantum computers cannot perform complex quantum algorithms for practical applications. Fortunately, quantum computation provides a new mathematical formalism for computing. As a new learning paradigm, quantum machine learning (QML) adopts quantum computation to enhance classical machine learning models [4, 32]. Recently, quantum theory has been adopted in classical algorithms and deep learning models, with an expectation of improving computation efficiency and outcome quality [9, 38].

Motivated by quantum theory, in this paper, we first introduce a quantum-inspired spectral-spatial network (QSSN) for HSI feature extraction. The proposed QSSN includes a phase-prediction module (PPM) and a measurement-like fusion module (MFM). Instead of fusing spatial and spectral information in different modules, we extract joint spatial-spectral features in the same operation. A small data cuboid taken from an HSI is represented by quantum-inspired state representation, and the amplitudes in a state vector are values in normalized input cuboid. The corresponding phases are predicted by PPM, dynamically modulating the spatial and spectral correlations. MFM simultaneously aggregates spatial and spectral information to generate a feature cuboid. Additionally, we design a quantum-inspired spectral-spatial pyramid network (QSSPN) framework with multiple cascaded QSSN blocks and a softmax classifier for HSI classification. Deep pyramid structure of QSSPN gradually decreases the channels of feature cuboids and extracts robust and expressive features from original data for classification. This is a simple but efficient end-to-end framework taking advantage from both quantum theory and deep neural network.

The contributions of this article are summarized as follows.

- As far as we know, QSSN and QSSPN are the first proposed quantum-inspired method for HSI feature extraction and classification. It shows that this quantum-inspired framework is promising.
- Motivated by QC in quantum theory, we propose a QSSN for HSI feature extraction. QSSN aggregates spatial-spectral information simultaneously using quantum-like representations and operations.
- Based on QSSN, we develop a QSSPN for HSI classification. In QSSPN, QSSN blocks stacked in a pyramid manner that can extract robust and discriminative features for classification. As shown in Fig. 1, QSSPN achieves the best classification accuracy with low model complexity.

The rest of this paper is organized as follows: Sec. 2 reviews the related work about deep learning-based HSI classification and quantum-inspired deep learning methods. Sec. 3 introduces the detailed framework of QSSN and describes the proposed QSSPN for HSI classification. Sec. 4 presents the experimental results and analysis. Finally, Sec. 5 draws the conclusion.

2. Related Work

Deep learning-based HSI Classification. Deep learning models adapted from color image processing show priority in HSI processing [21, 29]. Autoencoder-based [15] models are unsupervised networks that are widely researched in HSI classification [5, 6]. These methods cannot efficiently capture the spatial-spectral correlations. Inspired by human vision system, local connectivity allows CNN to extract spatial information efficiently [23]. CNN-based HSI classification methods are widely researched [29, 44, 47]. Zhong *et al*. proposed 3D CNN with residual connection to improve the network trainability and classification accuracy [47]. Attention mechanic [41] is introduced for learning correlations among neighboring pixels and bands [26, 35, 46]. Mei *et al*. designed two attention-based branches to extract discriminant feature vectors from spatial and spectral aspects [26]. Frameworks based on graph convolution network and transformer have been also investigated in recent years. These methods show priorities in capturing subtle spectral discrepancies [36] and learning structural information [45].

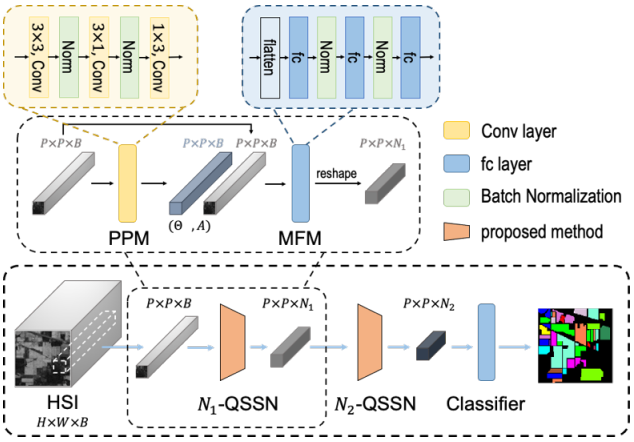


Figure 2. Illustration of the proposed QSSN and QSSPN framework. The yellow and blue boxes at the top are phase-prediction module (PPM) and measurement-like fusion module (MFM), respectively. The middle box shows the QSSN framework. The QSSPN framework is presented at the bottom.

Quantum-Inspired Computation. The mathematical formalism of quantum theory has been adopted in various tasks, including natural language processing (NLP) [25], computer vision (CV) [38], information retrieval (IR) [39, 40] and multimodal analysis [9, 10, 24]. Van *et al*. pioneeringly exploited the quantum probability for object representation in IR tasks [40]. To achieve better outcome quality and interpretability, Li *et al*. introduced a quantum-inspired network in NLP models for semantic matching [25]. Quantum-motivated strategies are also exploited in multimodal information processing. Li *et al*. investigated multimodal feature fusion methods via quantum-like neural network for conversation emotion recognition [24]. Gkoumas *et al*. introduced a video semantic recognition network via fusing multimodal information at the decision level [10]. Moreover, the quantum-inspired methods are also developed in image processing tasks. Tang's work utilizes the wave-like function to represent an image, using phase to model the relationship between patches [38]. The proposed wave-multilayer perceptron achieves promising performance on various vision tasks, such as image classification and objection detection.

3. Method

This section presents the QSSPN framework for HSI classification based on QSSN. We first briefly discuss the QC model. Then, we describe the proposed QSSN model to extract joint spatial-spectral features. At last, using QSSN as the building block, we introduce the proposed QSSPN framework for HSI classification. Details of QSSN and QSSPN are shown in Fig. 2.

3.1. Preliminaries

Most quantum computers are constructed based on the QC model [2, 28]. The computation of QC includes state initialization, unitary evolution and measurement. The quantum device is first set to the initial state. Scientists manipulate the quantum state by applying unitary operators to quantum device. In this process, the initial state of the system is mapped to the final state, storing processing results in quantum device. For capturing classical information from quantum device, measurements are adopted to extract processing results from the final state to the classical representation. Mathematically, the state of quantum system can be represented by a state vector in N-dimensional Hilbert space $\mathcal{H}^N$. We adopt the Dirac notations in quantum theory to denote the states of quantum system [28]. The notations $|\cdot\rangle$ and $\langle\cdot|$ denote a Ket to represent a complex-valued unit vector and a Bra to indicate its conjugate transpose. Specifically, given a set of basis state vectors $\{|\phi_n\rangle\}_{n=1}^{N}$, we can represent the states of quantum system $|\psi\rangle$ as follows*:

$$|\psi\rangle = \sum_{n=0}^{N-1} A_n e^{i\theta_n} |\phi_n\rangle , \tag{1}$$

where A_n is a probability amplitude satisfying $\sum_N |A_n|^2 = 1$, i is the imaginary unit and θ_n indicates the phase. A simple example of basis state vectors for $\mathcal{H}^N$ is standard unit basis. Detailedly, the standard basis of $\mathcal{H}^2$ is formed by column vectors $|0\rangle_2 = [1, 0]^T$ and $|1\rangle_2 = [0, 1]^T$. Here, we use the notation

$$|n\rangle_N , n = 0, 1, \ldots, N-1 \tag{2}$$

to represent the standard basis of $\mathcal{H}^N$. Accordingly, $\langle n|_N$ denotes the conjugate transpose of $|n\rangle_N$. All components of $|n\rangle_N$ are 0 except the n-th component is 1. In quantum theory, Eq. (1) represents the superposition of basis states.

The reversible *unitary* operator U describes the state's evolution in QC, mapping the initial states $|\psi_{initial}\rangle$ to final states $|\psi_{final}\rangle$. These operators manipulate the phases θ_n and amplitudes A_n by applying specific physical operations to the quantum system. In QC, projection-valued measure (PVM) is a function that can collapse the system state from superposition of multiple basis states to one basis state. Mathematical formulation of PVM is a set of projection operators $\{\Pi_m = |\phi_m\rangle\langle\phi_m|\}_{m=1}^{M}$. According to the *Born rule* [28], Π_m projects the state $|\psi_{final}\rangle$ to the corresponding basis state $|\phi_m\rangle$ with a probability:

$$P\left(|\phi_m\rangle\right) = Tr(\Pi_m \rho) = \langle\phi_m|\psi_{final}\rangle \langle\psi_{final}|\phi_m\rangle, \tag{3}$$

where $\rho = |\psi_{final}\rangle \langle\psi_{final}|$ is the density matrix representation of $|\psi_{final}\rangle$ and $\sum_M P(|\phi_m\rangle) = 1$. $|\cdot\rangle \langle\cdot|$ denotes the

*Only pure states are considered in this paper.

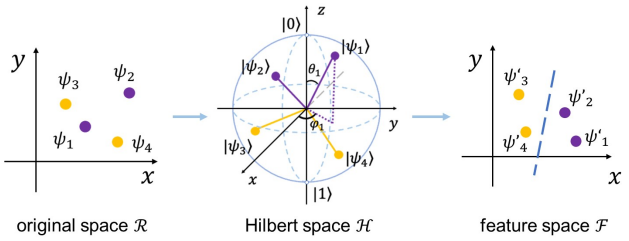


Figure 3. Illustration of data mapping in QSSN. The state representation $|\psi\rangle$ encodes data from original space to Hilbert space: $\mathcal{R} \rightarrow \mathcal{H}$. The MFM outputs a feature vector mapped from $|\psi\rangle$ to feature space: $\mathcal{H} \rightarrow \mathcal{F}$.

outer product and $\langle\cdot|\cdot\rangle$ denotes the inner product of $\langle\cdot|$ and $|\cdot\rangle$. Eq. (3) can be rewritten as:

$$P(|\phi_m\rangle) = \langle\phi_m| \rho |\phi_m\rangle = (\langle\psi_{final}|\phi_m\rangle)^2 \quad (4)$$
$$= \sum_{n_1}^{N-1} \sum_{n_2}^{N-1} A_{n_1} A_{n_2} e^{i(\theta_{n_1} - \theta_{n_2})} \langle\phi_m|n_1\rangle\langle n_2|\phi_m\rangle.$$

Here we neglect the subscript N of $|n\rangle_N$ for simplicity. In Eq. (4), the size of $|n\rangle$ is N. For the same measurement operator $\Pi_m \sim |\phi_m\rangle$, the output $P(|\phi_m\rangle)$ depends on all pairs of A_{n_1} and A_{n_2} and their correlations $(\theta_{n_1} - \theta_{n_2})$. PVM yields values $P_{res} = [P(|\phi_1\rangle), P(|\phi_2\rangle), \ldots, P(|\phi_M\rangle)]$ corresponding to the selected projection operators $\{\Pi_m\}_{m=1}^{M}$, reflecting the state of quantum system.

We loosely describe the quantum mechanic related to our method. A rigorous definition of quantum computing can be found in [28].

3.2. Quantum-Inspired Spectral-Spatial Network

Let $X \in \mathcal{R}^{H \times W \times B}$ be the HSI with H pixels in height, W pixels in width, and B spectral bands. Taking the (i, j)-th pixel as center, information in a small cuboid $x_{i,j} \in \mathcal{R}^{P \times P \times B}$ with $P \times P$ pixels are considered for finding the joint spectral-spatial feature representation corresponding to the center pixel [21]. We pad X with 0 so that we can successfully take cuboids for edge pixels. As shown in the middle box of Fig. 2, QSSN is composed of PPM and MFM to learn spectral-spatial feature representations for HSI classification.

Phase-Prediction Module (PPM). For HSI analysis, it is vital to take full advantage of spectral and spatial information. PPM utilizes spectral and spatial information to dynamically generate phases $\Theta_{i,j} \in R^{P \times P \times B}$ for all elements $x'_{i',j',k'}$ in $x_{i,j}$. Phases play an important role in encoding data $x_{i,j}$ in Hilbert space. We hope that phases can correctly model the spectral-spatial correlations of all elements. Besides, we hope that samples encoded in Hilbert space have small intra-class distance and large inter-class distance. Hence, PPM should be able to capture discriminative information from original data and assign appropriate phase $\theta_{i',j',k'}$ for each element $x'_{i',j',k'}$. Since 2D-CNN shows its priority in HSI analysis [29], we utilize the CNN blocks to construct PPM. PPM is a three-layer CNN. For predicting phase for each $x'_{i',j',k'}$, the shapes of input and output of each convolution layer are the same. This means that the kernels of each layer have B channels and zero-padding methods are adopted in each layer. The first CNN layer in PPM extracts features in 3×3 pixel windows. The following asymmetric convolution layers [37] with 3×1 and 1×3 kernel shapes fuse the information in a smaller region in the vertical and horizontal direction to predict the phases $\Theta_{i,j}$. Before we encode each HSI cuboid into complex finite-dimensional Hilbert space $\mathcal{H}^K$ $(K = P \times P \times B)$, we reshape the HSI cuboid $x_{i,j}$ and the corresponding phase cuboid $\Theta_{i,j}$ as follow:

$$x^{vec} = [x'_0, \ldots, x'_{i'+j' \times P + k' \times P \times P}, \ldots, x'_{K-1}],$$
$$\Theta^{vec} = [\theta_0, \ldots, \theta_{i'+j' \times P + k' \times P \times P}, \ldots, \theta_{K-1}], \quad (5)$$

where $i', j' \in [0 : P-1]$ and $k' \in [0 : B-1]$. Then phase vector x^{vec} and original data vector x^{vec} are used in the state vector representation $|x_{i,j}\rangle$ as follows:

$$|x_{i,j}\rangle = \sum_{n=0}^{K-1} A_n e^{i\theta_n} |n\rangle_K . \quad (6)$$

In Eq. (6), A_n can be obtained as:

$$A_n = \sqrt{\frac{x_n^{vec}}{\sum_{n^*=0}^{K-1} x_{n^*}^{vec}}}, \quad (7)$$

where x_n^{vec} indicates the n-th element in x^{vec}. In Eq. (6), $x_{i,j}$ can be represented by a complex vector $|x_{i,j}\rangle = \left[A_0 e^{i\theta_0}, A_1 e^{i\theta_1}, \ldots, A_{K-1} e^{i\theta_{K-1}}\right]$ in $\mathcal{H}^K$. As for classification task, we hope to encode samples in the space that has large inter-class distance and small intra-class distance, as shown in Fig. 3. A simple strategy to encode $x_{i,j}$ in $|x_{i,j}\rangle$ is neglecting the imaginary parts $e^{i\theta}$ (meaning 'no phase') or only using the same phases in spatial dimension (neglecting 'spectral phase') or spectral dimension (neglecting 'spatial phase'). Based on PPM, we adopt the imaginary parts to dynamically fuse spectral-spatial information in MFM. Compared to these simple strategies, our strategy is able to fully exploit the spatial and spectral correlations.

Measurement-like Fusion Module (MFM). Motivated by PVM, we design a measurement-like operation (MO) to generate low-dimensional feature representations [49]. Recalling that in Eq. (3), for simplicity, the mathematical formulation of measurement can be described as $PVM(\Pi_m, \rho)$: using a density matrix ρ to represent the state of quantum system $\rho = |\psi_{final}\rangle \langle\psi_{final}|$ and then applying the matrix manipulation (matrix product and trace)

with operators $\{\Pi_m\}_{m=1}^{M}$. For each Π_m, we get the probability $P(|\phi_m\rangle)$. Motivated by quantum measurement, we first use the density matrix representation to denote the state vector by

$$D_{i,j} = |x_{i,j}\rangle\langle x_{i,j}| = \sum_{n=0}^{K-1}\sum_{n'=0}^{K-1} A_n A_{n'} e^{i(\theta_n-\theta_{n'})} |n'\rangle \langle n|, \quad (8)$$

where $|n'\rangle\langle n|$ represents a square matrix with $P \times P \times B$ rows and columns that all components are 0 except the (n, n')-th component is 1. Each term in the density matrix $D_{i,j}$ is the multiplication of A_n and $A_{n'}$ with a coefficient $e^{i(\theta_n-\theta_{n'})}$. Empirically, this is similar to fusing information of each pair of elements x_n^{vec} and $x_{n'}^{vec}$. Phases θ_n and $\theta_{n'}$ are predicted by PPM. We hope the imaginary part $e^{i(\theta_n-\theta_{n'})}$ can dynamically modulate the correlations between x_n^{vec} and $x_{n'}^{vec}$. In each pair, the location and band of x_n^{vec} and $x_{n'}^{vec}$ may be different. Therefore, the density matrix can simultaneously fuse spatial and spectral information. Recalling that in Eq. (3), PVM generates numerical result $P(|\phi_m\rangle)$ corresponding to Π_m. Note that PVM has corresponding operations applied on physic devices. There are constraints of operations Π_m in their mathematical formulation in quantum theory. We use an abstract representation of PVM as $PVM(\Pi_m, \rho) \rightarrow PVM(Operator_m, \rho)$. To learn a low-dimensional representation, we design an MFM with a deep learning framework to relieve the limitations and perform similar functions in PVMs. In MFM, we use fully-connected layers FC_m with depth 3 as $Operator_m$. The MO generates a feature point as follows:

$$fea_m = MO(FC_m, D_{i,j}), \quad (9)$$

where $D_{i,j}$ stores the fusion information of each element pair, x_n^{vec} and $x_{n'}^{vec}$. MO is an operation that feeds $D_{i,j}$ into the operator FC_m to efficiently aggregate spatial and spectral information. Given M operators $\{FC_m\}_{m=1}^{M}$, we can get an M-dimensional feature vector $F^{vec} = [fea_1, fea_2, \ldots, fea_M]$. Aggressive dimension reduction of feature cuboids may cause bottlenecks of deep learning models [37]. Thus, M should be carefully designed for preserving important factors in data.

Following the steps above, QSSN extracts an M-dimensional feature vector from input cuboid $x_{i,j}$ with the size of $P \times P \times B$. Furthermore, we introduce a new notation N-QSSN to denote the learned vector representation F^{vec} with the size of $(P \times P \times N) \times 1$. The output feature vector F^{vec} is reshaped into a 3-D feature cuboid F^{cub} with the size of $P \times P \times N$. In N-QSSN, the size of M to denote the number of MO is determined by

$$M = P \times P \times N. \quad (10)$$

The outcome of N-QSSN F^{cub} has same spatial resolution as $x_{i,j}$. The input HSI cuboid $x_{i,j}$ has abundant band information but little spatial information. A few neighboring pixels are considered during feature extraction. Empirically, models are prone to ignore the spatial correlations. The N-QSSN preserves the spatial resolution in output, expecting QSSN to learn the location-sensitive information. Besides, it is possible to construct deep models based on the N-QSSN blocks.

3.3. QSSN-based Pyramid Framework

It has been proved that deep models have high-quality outcome in theory [16, 17] and practice [12, 34]. Recalling that QSSN consists of three convolution layers and three fully-connected layers. It can be treated as a building block to construct a deeper framework. To extract more robust and discriminative features, we use QSSN to develop a hierarchical framework for HSI classification, called QSSPN. Detailedly, we can cascade multiple QSSN blocks to extract deep features from $x_{i,j}$, followed by a classifier to identify the land-cover category of the center pixel (i.e., the (i, j)-th pixel). An example of the framework of QSSPN is shown in Fig. 2. The example model consists of two QSSN blocks and a linear classifier. For the two QSSN blocks, N_1-QSSN and N_2-QSSN are stacked in a hierarchical manner. Specifically, for N_1-QSSN, we take an original cuboid with the size of $P \times P \times B$ as input to learn the feature cuboid with the size of $P \times P \times N_1$ as output. Sequently, for N_2-QSSN, we use the feature representation with the size of $P \times P \times N_1$ learned from N_1-QSSN as input to extract the feature cuboid with the size of $P \times P \times N_2$ as output. In our setting, $B > N_1 > N_2$, meaning that the channel of feature maps are gradually reduced from the previous QSSN to the latter QSSN. The reduction of the channel of feature cuboids enables QSSPN to show the pyramid-like structure. Empirically, extreme decrease of the dimensionality of each QSSN may result in discarding important information [37]. Thus, we gently reduce the channels of each QSSN to find expressive and robust representations by aggregating spatial and spectral information.

The output feature cuboids learned from the final QSSN are then flattened into a vector and fed into a classifier. A linear layer with a softmax function is adopted as a classifier in this work. Therefore, our proposed framework is end-to-end and easy to implement. For simplicity, we use QSSPN-L to indicate QSSPN with L QSSN layers. For example, the model with two QSSN layers shown in Fig. 2 can be denoted by QSSPN-2.

Table 1. Classification results and model analysis of different methods on HSI datasets.

Datasets	Metrics	SSAN	SSRN	RvT	HiT	SSFTT	QSSPN-1	QSSPN-2	QSSPN-3
IP	OA (%)	89.46	91.85	83.85	90.59	**96.35**	90.36	92.66	95.87
	AA (%)	85.99	81.51	79.67	86.71	89.99	89.93	93.34	**96.40**
	κ(%)	88.04	90.73	81.68	89.27	**95.82**	89.05	91.68	95.34
	Params.	148.83K	735.88K	10.78M	49.60M	148.50K	910.50K	1.13M	1.17M
	FLOPs	7.88M	212.48M	17.83M	345.88M	3.66M	34.54M	52.75M	55.59M
PU	OA (%)	99.15	99.63	97.37	99.43	99.52	99.38	99.44	**99.71**
	AA (%)	98.70	99.29	95.86	99.09	99.20	98.61	98.89	**99.43**
	κ(%)	98.87	99.51	96.52	99.24	99.36	99.17	99.25	**99.61**
	Params.	94.63K	396.99K	9.77M	42.41M	148.03K	609.16K	666.38K	700.52K
	FLOPs	5.57M	108.04M	16.83M	190.85M	3.66M	10.25M	14.89M	17.66M
SA	OA (%)	98.92	99.31	98.11	99.38	99.53	99.54	99.63	**99.66**
	AA (%)	99.33	99.70	98.83	99.70	99.72	99.61	99.65	**99.81**
	κ(%)	98.80	99.23	97.90	99.31	99.47	99.49	99.59	**99.63**
	Params.	149.71K	750K	10.82M	50M	148.50K	926.90K	1.16M	1.20M
	FLOPs	7.97M	216.84	17.80M	354.42M	3.66M	35.87M	54.82M	57.67M

Table 2. Settings of channel reduction in QSSPN-L

Models	Bands	N_1	N_2	N_3
QSSPN-1	B	30	-	-
QSSPN-2	B	$\lfloor B/2 \rfloor$ *	30	-
QSSPN-3	B	$\lfloor B/2 \rfloor$	50	30

* $\lfloor B/2 \rfloor$ represents the greatest integer less than or equal to $B/2$.

4. Experiments

4.1. Experiment Details

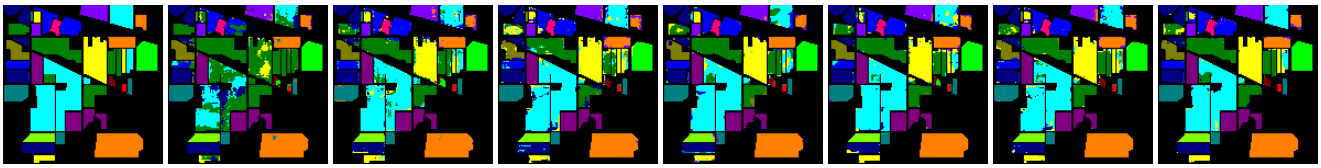
Datasets. We evaluate the proposed models on three HSI datasets, including Indian Pines (IP), Pavia University (PU) and Salinas (SA). The Indian Pines and Salinas datasets were collected by the airborne visible and infrared imaging spectrometer (AVIRIS) sensor. The Indian Pines dataset with 145×145 pixels and 200 valid spectral bands contains 16 land-cover categories, recording a scene of the physiognomy from the Indian Pines test site in Northwest Indiana. In this dataset, samples in different classes are extremely imbalanced. The covered area of Salinas Valley comprises 512×217 pixels and 204 spectral signatures in the Salinas dataset to describe 16 different ground truth categories. The Pavia University dataset was collected by the reflective optics system imaging spectrometer (ROSIS) sensor over the University of Pavia. There are 610×340 pixels with 103 bands to record nine classes of ground truths.

Compared methods. To validate the effectiveness of the proposed QSSPN framework, we utilize the following methods for comparison. We compare our model with advanced CNN-based and transformer-based models. The selected CNN-based models are spectral-spatial attention network (SSAN) [26] and spectral–spatial residual network (SSRN) [47]. As for the selected transformer-based models, there are robust vision transformer (RvT) [13], hyperspectral image transformer (HiT) [43] and spectral–spatial feature tokenization transformer (SSFTT) [36].

Setup. We adopt the Adam [22] optimizer with learning rate $1e^{-3}$ to update parameters of deep learning models in each training iteration. Models are trained through 200 epochs. The batch size is set to 64. Other settings of our proposed models are shown in Tab. 2. To make a fair comparison, we use the same input patch size $9 \times 9 \times B$ for all methods. For all datasets, we randomly select 10%, 5% and 85% labeled samples for model training, validation and testing, respectively. We train and test each model 10 times and report the average results. The performance is evaluated in terms of overall accuracy (OA), average accuracy (AA) and kappa coefficients (κ). Different from OA, AA helps to see if a model is strong or weak when classifying specific classes. All the experiments are performed with Pytorch [30] on NVIDIA 3090 GPU and 48GB RAM.

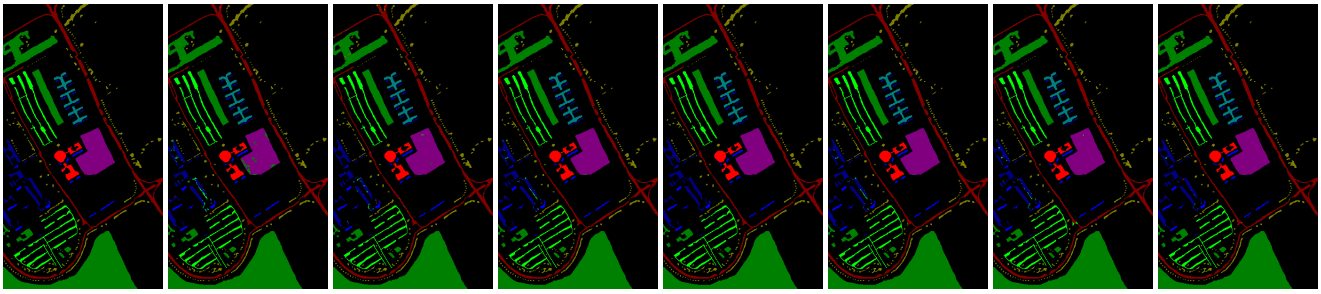
4.2. Classification Results

Tab. 1 reports the results of OA, AA and κ on three HSI datasets. Our proposed QSSPN-3 achieves the best AA on three dataset. Specifically, QSSPN-3 achieves 96.40% AA with only 1.17M Parameters on the Indian Pines dataset. It is much better than the existing methods. As shown in Fig. 4 (a), samples of different classes on the Indian Pines dataset are extremely imbalanced. Without data augmentation strategies, there are a few classes that only have less than 6 samples for training most of the time, such as 'Grass-pasture-mowed' and 'Oats'. While for some classes, such as classes 'Corn-notill' and 'Soybean-mintill', there are hundreds of samples. This makes it difficult for models to learn from data. Our methods achieve high classification accuracies in each class regardless of the distribution of training samples (96.40% AA). On the Salinas and Pavia University



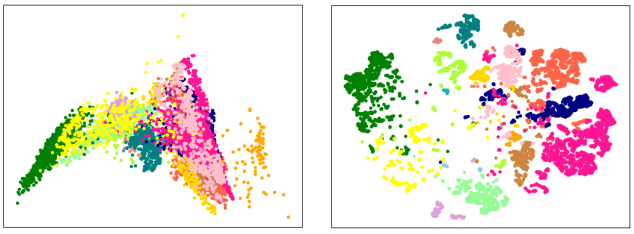
(a) Ground truth (b) RvT (c) HiT (d) SSAN (e) SSRN (f) QSSPN-1 (g) QSSPN-2 (h) QSSPN-3

Figure 4. Classification maps of different methods for the Indian Pines dataset.



(a) Ground truth (b) RvT (c) HiT (d) SSAN (e) SSRN (f) QSSPN-1 (g) QSSPN-2 (h) QSSPN-3

Figure 5. Classification maps of different methods for the Pavia University dataset.



(a) Original data (b) Spectral-spatial representation

Figure 6. Distribution of original data and representation extracted by QSSN blocks in QSSPN-3 for the Indian Pines dataset.

datasets, our methods obtain 0.1% and 0.3% improvement with respect to κ, respectively. QSSPN-1 with only one QSSN block achieves competitive performance with low computation complexity. It implies that our proposed QSSN framework can efficiently aggregate spatial-spectral information. Fig. 4 and Fig. 5 visualize the classification maps of our proposed methods on the Indian Pines and Pavia University datasets, along with results of best CNN-based and transformer-based models and the ground-truth maps.

To investigate the separability of data, we employ t-Distributed Stochastic Neighbor Embedding (t-SNE) to visualize the distribution of original data and spectral-spatial features extracted by QSSN blocks in QSSPN-3 in Fig. 6. As shown in Fig. 6, the extracted feature embedding are more separable. It means that QSSN can learn expressive representations for HSI classification. To verify the learnability of the proposed framework, we further evaluate the performance of different models on the Indian Pines dataset by varying training sizes of $\{15\%, 20\%, 25\%, 30\%\}$. Fig. 7 shows the OA values achieved by different models. All models achieve higher performance when the training size

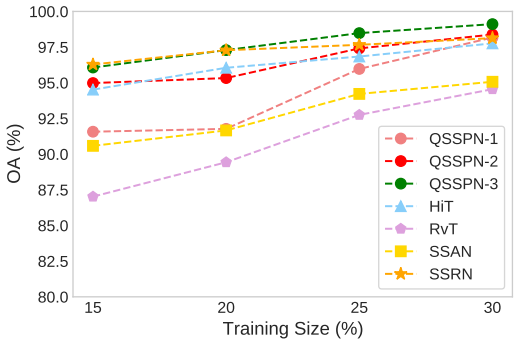


Figure 7. OA of different models for the Indian Pines dataset with different training sizes.

increases. Overall, QSSPN-3 surpasses other methods, achieving 99.11% OA with a 30% training size. It is a promising result, implying that our proposed framework cannot only learn from limited data, but also extract robust feature representations with sufficient training samples.

4.3. Ablation Study

Phases are used in representing a data cuboid by a state vector, dynamically modulating the spatial-spectral correlations between elements. To investigate the effectiveness of the phase module, we tabulate the classification results on the challenging Indian Pines dataset with QSSPN-2 for ablation study in Tab. 3. Without the phase information (i.e., the no phase mode, $\theta_{i',j',k'} = 0, i', j' \in [0 : P-1], k' \in [0 : B-1]$), the model has the worst performance compared with other models. The spatial phase mode (i.e., $\theta_{i',j',0} = \theta_{i',j',1} = \cdots = \theta_{i',j',B-1}$) considers spatial

Table 3. Effectiveness of phase on the Indian Pines dataset.

Mode	Params.	FLOPs	OA (%)	κ (%)
No phase	717.13K	18.92M	86.86	85.10
Spatial phase	708.74K	18.22M	89.18	87.74
Spectral phase	705.68K	17.97M	88.66	87.09
All	1.13M	52.75M	92.66	91.68

Table 4. Effectiveness of pyramid structure on the Indian Pines dataset.

L	Channel	Params.	FLOPs	OA (%)	κ (%)
1	10	571.18K	32.97M	82.72	80.45
1	30	910.5K	34.54M	90.36	89.05
1	50	1.26M	36.69M	91.63	90.49
2	30	1.14M	52.75M	91.93	90.84
3	30	1.17M	55.59M	92.93	91.96

* L indicates the QSSPN-L model.
* Channel indicates the dimension of the learned feature cuboid for classification.

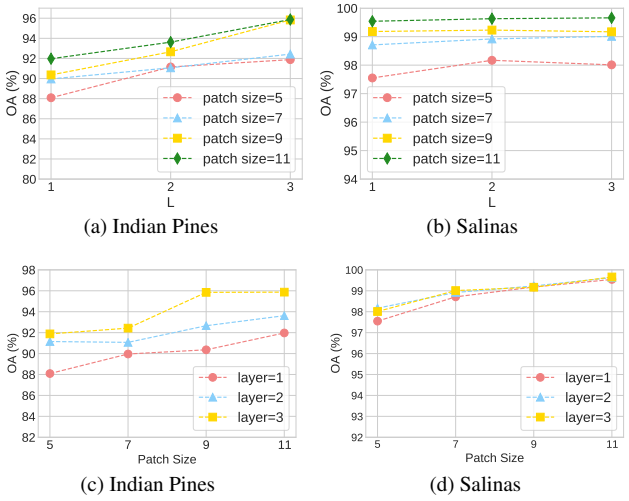
information and achieves better performance than the no phase mode. This implies that spatial phase can efficiently aggregate spatial information. However, this mode ignores abundant band information, making it perform worse than QSSPN. The spectral phase mode (i.e., $\theta_{0,0,k'} = \theta_{0,1,k'} = \cdots = \theta_{P-1,P-1,k'}$) only considers the spectral correlations and has similar performance as the spatial phase mode. The QSSPN uses joint spatial-spectral phases, incurring the best performance.

4.4. Parameter Sensitivity Study

To better understand the proposed QSSPN framework, we study the parameter sensitivity of patch size for HSI cuboid and model depth for pyramid structure. The parameter sensitivity studies are conducted on the Indian Pines dataset with QSSPN-2 unless being specifically stated.

Patch size for HSI cuboid. The patch size of the cuboid controls the size of input data. We investigate the effectiveness of patch size for classification. As shown in Figs. 8c and 8d, QSSPN achieves better accuracies with larger patches in most cases. We notice that the classification accuracy decreases when the size changes from 9 to 11 in Fig. 8c. It can be explained that as patch size increases, unrelated information and noise will increase. Besides, there is a trade-off between computation complexity and outcome quality. Large patches increase the computation cost with slight accuracy improvement.

Model depth for pyramid structure. The model structure plays an important role in deep learning models. We investigate QSSPN by varying the number of QSSN blocks and the channels of the learned feature cuboids. As shown in Figs. 8a and 8b, with the increase of QSSN blocks,



(a) Indian Pines (b) Salinas

(c) Indian Pines (d) Salinas

Figure 8. OA of QSSPN with different patch sizes and depths.

QSSPN achieves higher performance. Tab. 4 investigates the correlations between the channels of feature cuboid and the depth of the model. The extreme compression of feature channels incurs poor performance (*e.g.* 82.72% OA with 10 channels). Large feature cuboids achieve higher classification accuracy (*e.g.* 91.63% OA with 50 channels). However, the computation complexity greatly increases from 571.18K parameters to 1.26M parameters. With the same output feature channel, QSSPN-3 achieves higher accuracy compared to QSSPN-1 and QSSPN-2. QSSPN decreases the feature channels in a pyramid manner to avoid the bottleneck of deep learning models. This indicates that QSSPN is able to extract discriminative and robust spatial-spectral features with a deep pyramid structure.

5. Conclusion

In this paper, motivated by quantum theory, we first proposed a QSSN framework for HSI feature extraction. In QSSN, phases are predicted by PPM to represent the state representation for an HSI cuboid, and joint spatial-spectral features are extracted from state-like representation by MFM. Furthermore, we designed a QSSPN framework based on QSSN for HSI classification. This pyramid framework can extract expressive and robust features for classification. Extensive experiments have demonstrated the superiority of QSSPN. In our future work, we will investigate effective and efficient quantum-inspired methods for other applications.

Acknowledgements. This work was supported by the Science and Technology Development Fund, Macau SAR (File no. 0049/2022/A1), and by the University of Macau (File no. MYRG2022-00072-FST).

References

[1] Isabel Cecilia Contreras Acosta, Mahdi Khodadadzadeh, Laura Tusa, Pedram Ghamisi, and Richard Gloaguen. A machine learning framework for drill-core mineral mapping using hyperspectral and high-resolution mineralogical data fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(12):4829–4842, 2019. 1

[2] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando GSL Brandao, David A Buell, et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 2019. 2, 3

[3] Jón Atli Benediktsson, Jón Aevar Palmason, and Johannes R Sveinsson. Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):480–491, 2005. 1

[4] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, 2017. 2

[5] Yushi Chen, Zhouhan Lin, Xing Zhao, Gang Wang, and Yanfeng Gu. Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected topics in applied earth observations and remote sensing*, 7(6):2094–2107, 2014. 2

[6] Yushi Chen, Xing Zhao, and Xiuping Jia. Spectral–spatial classification of hyperspectral data based on deep belief network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6):2381–2392, 2015. 2

[7] Mauro Dalla Mura, Alberto Villa, Jon Atli Benediktsson, Jocelyn Chanussot, and Lorenzo Bruzzone. Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis. *IEEE Geoscience and Remote Sensing Letters*, 8(3):542–546, 2010. 1

[8] Caroline M Gevaert, Juha Suomalainen, Jing Tang, and Lammert Kooistra. Generation of spectral–temporal response surfaces by combining multispectral satellite and hyperspectral uav imagery for precision agriculture applications. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6):3140–3146, 2015. 1

[9] Dimitris Gkoumas, Qiuchi Li, Shahram Dehdashti, Massimo Melucci, Yijun Yu, and Dawei Song. Quantum cognitively motivated decision fusion for video sentiment analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 827–835, 2021. 2, 3

[10] Dimitris Gkoumas, Qiuchi Li, Shahram Dehdashti, Massimo Melucci, Yijun Yu, and Dawei Song. Quantum cognitively motivated decision fusion for video sentiment analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 827–835, 2021. 3

[11] Lov K Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 212–219, 1996. 2

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 2, 5

[13] Byeongho Heo, Sangdoo Yun, Dongyoon Han, Sanghyuk Chun, Junsuk Choe, and Seong Joon Oh. Rethinking spatial dimensions of vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 11936–11945, 2021. 6

[14] Jack D Hidary and Jack D Hidary. *Quantum computing: an applied approach*, volume 1. Springer, 2021. 2

[15] Geoffrey E Hinton and Richard Zemel. Autoencoders, minimum description length and helmholtz free energy. *Advances in neural information processing systems*, 6, 1993. 2

[16] Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural networks*, 4(2):251–257, 1991. 5

[17] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989. 5

[18] Wei Hu, Yangyu Huang, Li Wei, Fan Zhang, and Hengchao Li. Deep convolutional neural networks for hyperspectral image classification. *Journal of Sensors*, 2015, 2015. 2

[19] Kunshan Huang, Shutao Li, Xudong Kang, and Leyuan Fang. Spectral-spatial hyperspectral image classification based on knn. *Sensing and Imaging*, 17(1):1–13, 2016. 1

[20] Gordon Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE transactions on information theory*, 14(1):55–63, 1968. 1

[21] Maryam Imani and Hassan Ghassemian. An overview on spectral and spatial information fusion for hyperspectral image classification: Current trends and challenges. *Information fusion*, 59:59–83, 2020. 1, 2, 4

[22] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6

[23] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015. 2

[24] Qiuchi Li, Dimitris Gkoumas, Alessandro Sordoni, Jian-Yun Nie, and Massimo Melucci. Quantum-inspired neural network for conversational emotion recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 13270–13278, 2021. 3

[25] Qiuchi Li, Benyou Wang, and Massimo Melucci. Cnm: An interpretable complex-valued network for matching. *arXiv preprint arXiv:1904.05298*, 2019. 3

[26] Xiaoguang Mei, Erting Pan, Yong Ma, Xiaobing Dai, Jun Huang, Fan Fan, Qinglei Du, Hong Zheng, and Jiayi Ma. Spectral-spatial attention networks for hyperspectral image classification. *Remote Sensing*, 11(8):963, 2019. 2, 6

[27] Farid Melgani and Lorenzo Bruzzone. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on geoscience and remote sensing*, 42(8):1778–1790, 2004. 1

[28] Michael A Nielsen and Isaac Chuang. Quantum computation and quantum information, 2002. 2, 3, 4

[29] ME Paoletti, JM Haut, J Plaza, and A Plaza. Deep learning classifiers for hyperspectral imaging: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 158:279–317, 2019. 2, 4

[30] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019. 6

[31] John Preskill. Quantum computing in the nisq era and beyond. *Quantum*, 2:79, 2018. 2

[32] Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione. An introduction to quantum machine learning. *Contemporary Physics*, 56(2):172–185, 2015. 2

[33] Peter W Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM review*, 41(2):303–332, 1999. 2

[34] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 2, 5

[35] Le Sun, Yu Fang, Yuwen Chen, Wei Huang, Zebin Wu, and Byeungwoo Jeon. Multi-structure kelm with attention fusion strategy for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–17, 2022. 2

[36] Le Sun, Guangrui Zhao, Yuhui Zheng, and Zebin Wu. Spectral–spatial feature tokenization transformer for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2022. 2, 6

[37] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016. 4, 5

[38] Yehui Tang, Kai Han, Jianyuan Guo, Chang Xu, Yanxi Li, Chao Xu, and Yunhe Wang. An image patch is a wave: Phase-aware vision mlp. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10935–10944, 2022. 2, 3

[39] Sagar Uprety, Dimitris Gkoumas, and Dawei Song. A survey of quantum theory inspired approaches to information retrieval. *ACM Computing Surveys (CSUR)*, 53(5):1–39, 2020. 3

[40] Cornelis Joost Van Rijsbergen. *The geometry of information retrieval*. Cambridge University Press, 2004. 3

[41] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 2

[42] Shuo Yang and Zhenwei Shi. Hyperspectral image target detection improvement based on total variation. *IEEE Transactions on Image Processing*, 25(5):2249–2258, 2016. 1

[43] Xiaofei Yang, Weijia Cao, Yao Lu, and Yicong Zhou. Hyperspectral image transformer classification networks. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2022. 2, 6

[44] Shuyu Zhang, Meng Xu, Jun Zhou, and Sen Jia. Unsupervised spatial-spectral cnn-based feature learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–17, 2022. 2

[45] Yongshan Zhang, Yang Wang, Xiaohong Chen, Xinwei Jiang, and Yicong Zhou. Spectral-spatial feature extraction with dual graph autoencoder for hyperspectral image clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022. 2

[46] Xiangtao Zheng, Hao Sun, Xiaoqiang Lu, and Wei Xie. Rotation-invariant attention network for hyperspectral image classification. *IEEE Transactions on Image Processing*, 2022. 2

[47] Zilong Zhong, Jonathan Li, Zhiming Luo, and Michael Chapman. Spectral–spatial residual network for hyperspectral image classification: A 3-d deep learning framework. *IEEE Transactions on Geoscience and Remote Sensing*, 56(2):847–858, 2017. 2, 6

[48] Zilong Zhong, Ying Li, Lingfei Ma, Jonathan Li, and Wei-Shi Zheng. Spectral–spatial transformer network for hyperspectral image classification: A factorized architecture search framework. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2021. 2

[49] Arthur Zimek, Erich Schubert, and Hans-Peter Kriegel. A survey on unsupervised outlier detection in high-dimensional numerical data. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 5(5):363–387, 2012. 1, 4