Deep Reinforcement Learning for Band Selection in Hyperspectral Image Classification

Lichao Mou, Sudipan Saha, *Member, IEEE*, Yuansheng Hua, *Student Member, IEEE*, Francesca Bovolo, *Senior Member, IEEE*, Lorenzo Bruzzone, *Fellow, IEEE*, and Xiao Xiang Zhu, *Fellow, IEEE*

***Abstract*— Band selection refers to the process of choosing the most relevant bands in a hyperspectral image. By selecting a limited number of optimal bands, we aim at speeding up model training, improving accuracy, or both. It reduces redundancy among spectral bands while trying to preserve the original information of the image. By now, many efforts have been made to develop unsupervised band selection approaches, of which the majorities are heuristic algorithms devised by trial and error. In this article, we are interested in training an intelligent agent that, given a hyperspectral image, is capable of automatically learning policy to select an optimal band subset without any hand-engineered reasoning. To this end, we frame the problem of unsupervised band selection as a Markov decision process, propose an effective method to parameterize it, and finally solve the problem by deep reinforcement learning. Once the agent is trained, it learns a band-selection policy that guides the agent to sequentially select bands by fully exploiting the hyperspectral image and previously picked bands. Furthermore, we propose two different reward schemes for the environment simulation of deep reinforcement learning and compare them in experiments. This, to the best of our knowledge, is the first study that explores a deep reinforcement learning model for hyperspectral image analysis, thus opening a new door for future research and showcasing the great potential of deep reinforcement learning in remote sensing applications. Extensive experiments are carried out on four hyperspectral data sets, and experimental results demonstrate the effectiveness of the proposed method. The code is publicly available.**

***Index Terms*— Deep Q-network, deep reinforcement learning, hyperspectral band selection, hyperspectral image classification, neural network, unsupervised learning.**

Manuscript received October 5, 2020; revised January 18, 2021; accepted February 25, 2021. This work was supported in part by the European Research Council (ERC) through the European Union's Horizon 2020 Research And Innovation Programme under grant Agreement ERC-2016-StG-714087 (*So2Sat*), in part by the Helmholtz Association through the Framework of Helmholtz AI under Grant ZT-I-PF-5-01 (Local Unit "Munich Unit @Aeronautics, Space and Transport (MASTr)" and Helmholtz Excellent Professorship "Data Science in Earth Observation—Big Data Fusion for Urban Research"), and in part by the German Federal Ministry of Education and Research (BMBF) in the framework of the International Future AI Lab "AI4EO—Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond" under Grant 01DD20001. *(Corresponding author: Xiao Xiang Zhu.)*

Lichao Mou, Yuansheng Hua, and Xiao Xiang Zhu are with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), 82234 Weßling, Germany, and also with the Data Science in Earth Observation (SiPEO; former: Signal Processing in Earth Observation), Technical University of Munich (TUM), 80333 Munich, Germany (e-mail: lichao.mou@dlr.de; yuansheng.hua@dlr.de; xiaoxiang.zhu@dlr.de).

Sudipan Saha is with the Data Science in Earth Observation (SiPEO; former: Signal Processing in Earth Observation), Technical University of Munich (TUM), 80333 Munich, Germany (e-mail: sudipan.saha@tum.de).

Francesca Bovolo is with the Fondazione Bruno Kessler, 38123 Povo, Italy (e-mail: bovolo@fbk.eu).

Lorenzo Bruzzone is with the Department of Information Engineering and Computer Science, University of Trento, 38122 Trento, Italy (e-mail: lorenzo.bruzzone@ing.unitn.it).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TGRS.2021.3067096.

Digital Object Identifier 10.1109/TGRS.2021.3067096

I. Introduction

In remote sensing, spectral sensors are widely used for Earth observation tasks, such as land cover classification [1]–[15], anomaly detection [16]–[20], and change detection [21]–[33]. A hyperspectral image often comprises hundreds of spectral bands within and beyond the visible spectrum. Such an image can be deemed as a hypercube, providing rich spectral information that helps to identify various land covers. Hyperdimensionality also raises some issues, e.g., a high level of redundancy among spectral bands, high computational overheads, and large storage requirements. Therefore, it is beneficial to reduce data redundancy.

In the literature, two kinds of methodologies, namely, feature extraction [34], [35] and band selection [36]–[55], are commonly used to reduce redundancy in hyperspectral images. The former transforms the original hyperspectral data into a lower dimension via a linear or nonlinear mapping. For example, Wang and Chang [34] make use of independent component analysis (ICA) to extract features from a hyperspectral image in an unsupervised way. Bandos *et al.* [35] investigate a supervised feature extraction approach based on the linear discriminant analysis (LDA). Moreover, several works put effort into using manifold learning algorithms, e.g., Laplacian eigenmaps (LEs) [56], locally linear embedding (LLE) [57], and isometric feature mapping (Isomap) [58], to learn low-dimensional features by taking advantage of the underlying geometric structure of hyperspectral data. On the other hand, band selection refers to the process of choosing a cluster of informative spectral bands and discarding ones that are often not discriminative enough for the considered problem. Unlike feature extraction, band selection can keep the physical meaning of the original hyperspectral images and be better interpreted for certain tasks [52]. Hence, in this article, we are interested in hyperspectral band selection. Band selection is applicable to tasks as diverse as hyperspectral image classification, change detection, and anomaly detection. In this

work, we use classification tasks to validate the effectiveness of selected bands.

From the perspective of the availability and use of labeled data, band selection methods are grouped into the following three categories: unsupervised, semisupervised, and supervised. Semisupervised and supervised models exploit labeled samples to learn a band selection strategy. Such labeled data, however, are not often available in practical remote sensing applications. Hence, unsupervised band selection is more desirable in the community. In this direction, the existing methods can be approximately sorted into the following categories.

1) *Ranking-Based Methods:* These methods aim at seeking an effective criterion to measure the significance of each spectral band and prioritize all bands. Afterward, top-ranked bands are selected. Some representative ranking-based band selection methods are [36]–[38].
2) *Searching-Based Methods:* The searching-based band selection approaches usually have two components: an objective function and a sequential search algorithm. The former is a criterion that the latter seeks to minimize over all feasible band subsets by adding or removing bands from a candidate set. The searching-based methods have two variants: sequential forward selection and sequential backward selection. The works in [39] and [40] are both representative works in this direction.
3) *Clustering-Based Methods:* In these methods, all spectral bands are first grouped into several clusters via a clustering algorithm. Afterward, the most representative band is selected from each cluster. Representative clustering-based band selection methods include [41]–[44].
4) *Others:* Some hybrid approaches, e.g., combining ranking and clustering [46]–[48], are proposed for band selection tasks. Furthermore, sparse learning, low rank representation, and deep learning also provide new insights [45], [49], [50].

In essence, hyperspectral band selection can be treated as a combinatorial optimization problem. The aforementioned methods that use exact and heuristic algorithms have proven to be effective for such a task. However, these heuristic algorithms are devised based on domain knowledge from human experts by trial and error. Hence, we are curious as to whether this heuristic design procedure for unsupervised band selection tasks can be automated using artificial intelligence techniques. If feasible, there would be much to be gained. Reinforcement learning systems are trained from their own experience, in principle, allowing them to operate in tasks where human expertise is lacking and, thus, being suitable for discovering new band selection methods without any hand-engineered reasoning. Recently, deep reinforcement learning, introducing deep learning into reinforcement learning, has demonstrated breakthrough achievements in various fields [59]–[63]. In this article, we propose a framework that can solve the problem of unsupervised band selection using deep reinforcement learning.

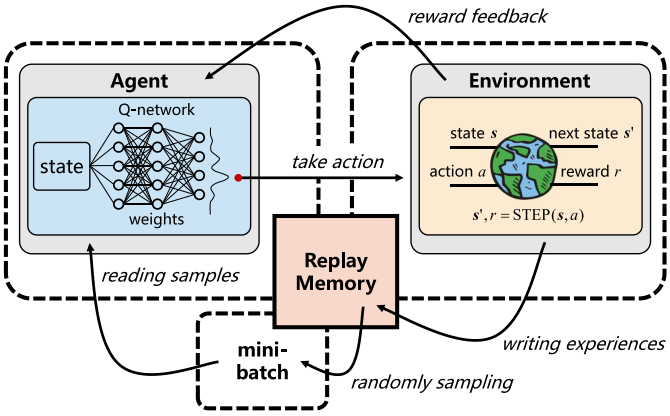


Fig. 1. Overview of the proposed deep reinforcement learning model for unsupervised hyperspectral band selection. In the training phase, an intelligent agent (Q-network) interacts with a tailored environment in order to learn a band-selection policy by trial and error. Specifically, the Q-network takes as input the state representation encoding selected bands and outputs a vector whose each component is a Q value for each band. In the test phase, the agent selects bands according to the learned policy.

This work's novel contributions are in the following aspects.

1) We cast the problem of unsupervised hyperspectral band selection as a Markov decision process of an agent and then solve this problem with a deep reinforcement learning algorithm. To the best of our knowledge, this is the first study that makes use of deep reinforcement learning for the task of band selection.
2) We propose an effective solution to parameterize the Markov decision process for optimal band selection. More specifically, for the agent, we devise the set of actions, the set of states, and an environment simulation tailored for this task.
3) We present and discuss two instantiations of the reward scheme of the environment simulation, namely, information entropy and correlation coefficient, for unsupervised hyperspectral band selection.
4) We train a deep reinforcement learning model using the Q-network to learn a band-selection policy whose effectiveness has been validated extensively with various data sets and classifiers.

We organize the remainder of this article as follows. Hyperspectral band selection is detailed in Section I. Section II introduces the proposed model. Section III tests the proposed model and presents experimental results and the discussion. Finally, this article is concluded in Section IV.

II. Methodology

Let us consider a hyperspectral image with L-bands. Our goal is to select K optimal bands to reduce redundancy. The number of all possible combinations is $\binom{L}{K}$. Suppose that $L = 200$ and $K = 30$; the number is about 4×10^{35}. In this work, we first formulate the task as a Markov decision process, as detailed in Section II-A. Afterward, a deep reinforcement learning model is used to solve this problem (see Section II-B). Section II-C discusses implementation details.

A. Problem Formulation: Band Selection as a Markov Decision Process

We view the task of hyperspectral band selection as a sequential forward search (SFS) process, i.e., a sequential decision-making problem of an agent, which interacts with a tailored environment (see Fig. 1). To be more specific, the agent needs to decide which spectral band it should pick at each time step so that it can find an optimal combination of K-bands in K steps, and during this procedure, the agent explores the environment through actions and observes rewards and states. In this article, we cast this problem as a Markov decision process that offers a formal framework for modeling the procedure of sequential decision-making when outcomes are partially uncertain.

A 5-tuple $(\mathcal{A}, \mathcal{S}, P, R, \gamma)$ is often used to define a Markov decision process [64]. Here, $\mathcal{A}$ denotes the set of all actions, $\mathcal{S}$ is the set of all countable or uncountable states, $P : \mathcal{S} \times \mathcal{A} \to \mathcal{P}(\mathcal{S})$ represents the Markov transition function, $R : \mathcal{S} \times \mathcal{A} \to \mathcal{P}(\mathbb{R})$ is the distribution of immediate rewards of state-action pairs, and $\gamma \in (0, 1)$ denotes a discount factor. In specific, upon taking an action $a \in \mathcal{A}$ at a state $\boldsymbol{s} \in \mathcal{S}$, the probability distribution of the next state can be defined by $P(\cdot|\boldsymbol{s}, a)$, and $R(\cdot|\boldsymbol{s}, a)$ depicts the distribution of the immediate reward for the chosen action. In what follows, we detail how we parameterize the Markov decision process for our case.

Action: The action of the agent in our case is to choose a spectral band from the hyperspectral image at each time step. The complete set of all actions $\mathcal{C}$ is identical to the set of bands, i.e., $\mathcal{C} = \{1, 2, \ldots, L\}$. Let $\mathcal{B}$ be a set consisting of actions that have been taken before. Then, the actual set of actions for the current time step is $\mathcal{A} = \mathcal{C} \setminus \mathcal{B}$. During the training phase, an action a, $a \in \mathcal{A}$, is taken by the agent and subsequently sent to the environment, and the latter receives the action, evaluates it, and gives the agent a positive or negative reward. In the test phase, the agent acts according to a learned policy to sequentially select bands.

State: The state $\boldsymbol{s}$ in our case is represented as the action history of the agent and is denoted as a L-dimensional vector with multihot encoding that records which actions have been taken (i.e., which spectral bands have been chosen) in the past. For example, $s_i = 1$ means that the ith band has been picked in previous time steps, while $s_i = 0$ represents that it is still selectable. Taking the action history as the state implies dependencies among spectral bands, which helps to select the next band. Note that there exists a one-to-one correspondence between $\boldsymbol{s}$ and $\mathcal{B}$.

Transition: The transition function P deems the next state as a possible outcome of taking an action at a state. In this work, the transition function is deterministic, which means that the next state is specified for each state-action pair. Specifically, P updates the state by changing the action history as follows:

$$\mathcal{B}' = \mathcal{B} \cup \{a\} \tag{1}$$

where $\mathcal{B}'$ represents the set of selected bands associated with the next state $\boldsymbol{s}'$.

Reward: The reward function R should be in proportion to the advancement that the agent makes after picking a specific band. In this work, we discuss two ways to instantiate our reward scheme and measure the improvement from one state to another in our setup. They are detailed as follows.

1) *Information Entropy:* The information entropy is capable of measuring the information amount of a random variable quantitatively. Hence, we make use of it to evaluate the richness of spectral information of bands. More specifically, denoting $\boldsymbol{x}_i \in \mathbb{R}^N$ as the ith band vector, we calculate the mean information entropy of selected bands as follows:

$$\mathrm{MIE}(\boldsymbol{s}) = -\frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} \sum_{n=1}^{N} P(x_i^n) \log_2 P(x_i^n) \tag{2}$$

where $\mathcal{B}$ is associated with the state $\boldsymbol{s}$. When the agent takes an action a and moves from state $\boldsymbol{s}$ to $\boldsymbol{s}'$, the reward $R(\boldsymbol{s}, \boldsymbol{s}')$ can be calculated as follows:

$$R(\boldsymbol{s}, \boldsymbol{s}') = \mathrm{MIE}(\boldsymbol{s}') - \mathrm{MIE}(\boldsymbol{s}) . \tag{3}$$

2) *Correlation Coefficient:* The correlation coefficient measures how strong the relationship between two variables is. Here, we use it to estimate intraband correlations among selected bands. There are several types of correlation coefficients, and we exploit a commonly used one, Pearson's correlation, also known as, Pearson's R, to calculate the mean correlation coefficient for $\boldsymbol{s}$ as follows:

$$\mathrm{Corr}(\boldsymbol{s}) = \frac{1}{|\mathcal{B}|^2} \sum_{i \in \mathcal{B}} \sum_{j \in \mathcal{B}} \frac{E[(\boldsymbol{x}_i - \mu_{\boldsymbol{x}_i})(\boldsymbol{x}_j - \mu_{\boldsymbol{x}_j})]}{\sigma_{\boldsymbol{x}_i} \sigma_{\boldsymbol{x}_j}} . \tag{4}$$

Then, the agent can be rewarded by the following formula:

$$R(\boldsymbol{s}, \boldsymbol{s}') = \mathrm{Corr}(\boldsymbol{s}) - \mathrm{Corr}(\boldsymbol{s}') . \tag{5}$$

Intuitively, (3) and (5) tell that the reward is positive if the quality of selected bands is improved from state $\boldsymbol{s}$ to state $\boldsymbol{s}'$ and negative otherwise. Driven by this reward scheme, the agent pays a penalty for choosing a noninformative band and is rewarded to add a band that results in an increase in the informative content of the whole set of selected bands. We quantitatively compare the above two instantiations of the reward scheme in Section III-C. The information entropy and correlation coefficient are two commonly used metrics to assess the quality of bands selected by an unsupervised band selection model [42], [50], which is the reason why we consider them as the reward scheme. Furthermore, we believe that more alternatives are possible and may improve results in the future.

B. Deep Reinforcement Learning for Band-Selection Policy

In Section II-A, we discuss the parameterization of the Markov decision process for our task. By doing so, the band selection task is transformed into a sequential decision-making problem. Next, we show how we use deep reinforcement learning to learn a band-selection policy in this setup.

The policy we seek is an action-value function, denoted by $Q(\boldsymbol{s}, a)$,[1] which specifies the action a to be taken when the

[1]"Q" in reinforcement learning is an abbreviation of the word "Quality."

Algorithm 1: Training

randomly initialize Q-network weights $\boldsymbol{w}$;
initialize replay memory $\mathcal{M}$;
initialize the complete set of all actions $\mathcal{C}$;
while *not converged* **do**
  initialize state: $\boldsymbol{s} = \vec{0}$;
  empty the set of chosen bands: $\mathcal{B} = \emptyset$;
  for $t = 1$ *to* K **do**
    compute the actual set of actions: $\mathcal{A} = \mathcal{C} \setminus \mathcal{B}$;
    simulate one step with the ϵ-greedy policy π_ϵ:
    $a = \pi_\epsilon(\boldsymbol{s})$; $\boldsymbol{s}', r = \text{STEP}(\boldsymbol{s}, a)$;
    $\mathcal{B} \leftarrow \mathcal{B} \cup \{a\}$;
    add the experience $(\boldsymbol{s}, a, r, \boldsymbol{s}')$ into $\mathcal{M}$;
    $\boldsymbol{s} \leftarrow \boldsymbol{s}'$;
  end
  randomly sample a mini-batch $\mathcal{B}$ from $\mathcal{M}$;
  for *all* $(\boldsymbol{s}, a, r, \boldsymbol{s}') \in \mathcal{B}$ **do**
    calculate the learning target according to Eq. (8):
    $y = r + \gamma \max_{a'} Q(\boldsymbol{s}', a'; \boldsymbol{w})$;
  end
  carry out a gradient descent step on $\mathcal{L}$ w.r.t. $\boldsymbol{w}$
  according to Eq. (9):
  $\nabla_{\boldsymbol{w}} \mathcal{L} = \mathbb{E}_{(s,a,r,s')}[(y - Q(\boldsymbol{s}, a; \boldsymbol{w}))] \nabla_{\boldsymbol{w}} Q(\boldsymbol{s}, a; \boldsymbol{w})$;
  update Q-network weights.
end

current state is $\boldsymbol{s}$. Based on this function, the agent chooses the action that is associated with the maximum reward value. That is to say, in our task, bands with high information entropy or low correlation are expected to be chosen. Q-learning [65], a classical reinforcement learning algorithm, is often employed to approximate $Q(\boldsymbol{s}, a)$ by iteratively updating the action-selection policy using the Bellman equation

$$Q(\boldsymbol{s}, a) = r + \gamma \max_{a'} Q(\boldsymbol{s}', a') \tag{6}$$

where r denotes the immediate reward and the second term $Q(\boldsymbol{s}', a')$ is a future reward. In Q-learning, a lookup table, termed Q-table, serves as the Q-function for the agent to query the best action. However, this becomes impractical when action and state spaces are very large. To tackle such a problem, in this article, we exploit a network named Q-network to approximate the action-value function.

Q-network Architecture: The Q-network takes as input the state representation introduced in Section II-A and outputs a vector whose each component is a Q value for each action. A detailed description of the Q-network that we use is given as follows. The input consists of an L-dimensional vector. The first fully connected layer has $2L$ units, followed by rectifier linear units (ReLUs) [66]. The second fully connected layer has the same structure as the first layer, again followed by ReLUs. Finally, the last layer, a linear fully connected layer with L units, follows. The structure of the Q-network is outlined in Table I.

Q-Network Learning: The Q-network is learned by minimizing the following mean squared Bellman error:

$$\mathcal{L} = \mathbb{E}_{(s,a,r,s')}[(y - \underbrace{Q(\boldsymbol{s}, a; \boldsymbol{w})}_{\text{Prediction}})^2] \tag{7}$$

where $\boldsymbol{w}$ represents network weights, and y is the one-step ahead learning target

$$y = r + \gamma \max_{a'} Q(\boldsymbol{s}', a'; \boldsymbol{w}) . \tag{8}$$

From (8), it can be seen that the target is composed of the immediate reward r and a discounted future reward. Ideally, the prediction of the current action-selection policy is supposed to be very close to the target, i.e., we want the error to decrease. Hence, we carry out a gradient descent step on $\mathcal{L}$ with respect to $\boldsymbol{w}$ according to

$$\nabla_{\boldsymbol{w}} \mathcal{L} = \mathbb{E}_{(s,a,r,s')}[(y - Q(\boldsymbol{s}, a; \boldsymbol{w}))] \nabla_{\boldsymbol{w}} Q(\boldsymbol{s}, a; \boldsymbol{w}) . \tag{9}$$

In fact, the learning of the Q-network for estimating the action-value function tends to be unstable. Therefore, in deep Q-learning, several techniques are used to address this problem, and they are detailed in the following.

Experience Replay: Here, an experience refers to a 4-tuple $(\boldsymbol{s}, a, r, \boldsymbol{s}')$. Consecutively generated experiences in our model are highly correlated with each other, and this could result in unstable and inefficient learning that is also a notorious problem in Q-learning. One solution to make the learning converge is to collect and store experiences in a replay memory, and during the training phase of the Q-network, minibatches are randomly taken out from this replay memory and utilized for the Q-network training. This method has the following advantages.

1) One experience can be potentially used for many gradient descent steps, which improves data efficiency.
2) Randomizing experiences breaks correlations among consecutive samples, therefore, reduces the variance of gradient descent steps, and stabilizes the learning of the network.

Exploration-Exploitation: To train the Q-network, we use an ϵ-greedy policy, which means that the agent either chooses actions at will with a probability ϵ or takes the best actions relying on the already learned band-selection policy with a probability $1 - \epsilon$. The learning of the Q-network starts with a relatively large ϵ and then gradually decays it. The main idea behind this policy is that the agent is encouraged to try as many actions (i.e., various band combinations) as possible to begin with before it starts to see patterns. When it does not select actions at random, given a state, the agent is able to estimate the reward for each action. Thus, the best action leading to the highest reward can be picked. Moreover, note that the ϵ-greedy policy of our model is carried out on the actual action set $\mathcal{A}$, instead of the complete action set $\mathcal{C}$.

C. Implementation Details

In this work, we set the maximum size of the replay memory as 50 000 and make use of a batch size of 100. The ϵ-greedy policy starts with $\epsilon = 1$ and decreases until $\epsilon = 0.01$ in steps of 0.95. The weights of the Q-network are initialized

Algorithm 2: Environment Simulation (Based on Information Entropy)

function $s', r = \text{STEP}(s, a)$ **:**
get s' based on s and a;
if s *is* $\vec{0}$ **then**
  $r = -\sum_{n=1}^{N} P(x_a^n) \log_2 P(x_a^n)$;
else
  calculate r according to Eq. (3):
  $r = \text{MIE}(s') - \text{MIE}(s)$;
end

TABLE I
ILLUSTRATION OF THE Q-NETWORK THAT WE USE. TAKING THE INDIAN PINES DATA SET AS AN EXAMPLE

Layer	Input	Output	#Units	Connected to	Activation
fc1	(200,)	(400,)	400	input	ReLU
fc2	(400,)	(400,)	400	fc1	ReLU
fc3	(400,)	(200,)	200	fc2	linear

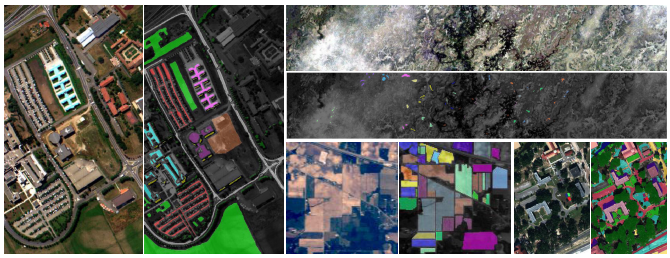


Fig. 2. From Left to Right and Top to Bottom: true-color composite images and ground-truth data of the Pavia University, Botswana, Indian Pines, and MUUFL Gulfport data sets.

randomly from a uniform distribution [67], and we note that outcomes are not sensitive to this initialization. For training the Q-network, Nesterov Adam [68] is chosen as the optimizer, and its parameters are set as recommended, i.e., $\beta_1 = 0.9$ and $\beta_2 = 0.999$. In addition, a learning rate of 1×10^{-4} is used. The training procedure and environment simulation of our model are shown in Algorithm 1 and 2. Once the agent is trained with Algorithm 1, it learns a band-selection policy that guides the agent to choose the band with the maximum estimated Q value at each step. It should be noted that, no matter in the training or the test phase, the agent is supposed to select spectral bands without duplicates in an episode. The code is publicly available.[2]

III. EXPERIMENTS AND ANALYSIS

A. Hyperspectral Data Set Description

1) Indian Pines: This scene was collected in June 1992 via NASA/JPL's Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor. It covers a geographical area in Northwestern Indiana, United States. This data set includes 145×145 pixels, and its spatial resolution is 20 m/pixel. There are totally 220 spectral bands, and their wavelength values range between 400 and 2500 nm. The ground truth provided by the data set involves 16 classes of interest, of which the majority of these classes are related to crops at variant growth stages (see Fig. 2). Before performing band selection algorithms, we remove 20 bands, i.e., 104–108, 150–163, and 220, as they are both water absorption ones, and as a result, 200 spectral bands are eventually used in total.

[2]https://github.com/lcmou/DRL4BS

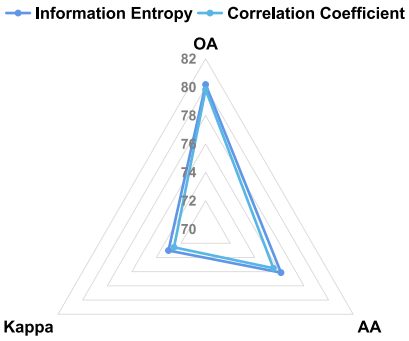


Fig. 3. Comparison of two reward schemes, namely, information entropy and correlation coefficient, on the Pavia University data set.

2) Pavia University: The second scene was captured through Reflective Optics Spectrographic Imaging System (ROSIS) on an aircraft operated by the German Aerospace Center (DLR) in 2002. It covers an area of the University of Pavia and is composed of 103 spectral bands in the wavelength range of 430–860 nm after discarding 12 noisy bands and 610×340 pixels. The spatial resolution of this scene is 1.3 m/pixel. Except for unknown pixels, nine land cover categories are labeled manually in the ground truth. Fig. 2 exhibits the true-color composite image and reference data of the Pavia University data set.

3) Botswana: The Botswana scene was collected by a hyperspectral sensor, Hyperion, on the NASA EO-1 satellite in May 2001. It covers a 7.7-km strip in the Okavango Delta, Botswana, and includes 1476×256 pixels. Its spatial resolution is 30 m/pixel; 242 spectral bands whose wavelength varies between 400 and 2500 nm are originally captured in this data set, but only 145 bands are used in our study after we remove noisy and uncalibrated bands. There are 14 classes representing different land covers included in the ground truth provided by the data set (see Fig. 2).

4) MUUFL Gulfport: The MUUFL Gulfport data set [69], [70] was acquired over the campus of the University of Southern Mississippi Gulf Park, Long Beach, Mississippi, in 2010. It includes coregistered hyperspectral and LiDAR data, but, in this work, we only use the hyperspectral data that originally contains 72 bands. However, due to noise, the first four and last four bands are omitted, bringing about an image with 64 bands. There is a total of 325×337 pixels, and the provided ground-truth map includes 11 classes. This data set can be used to evaluate the performance of different band selection methods under the circumstance where a hyperspectral data set has a limited number of bands.

Table II outlines the number of labeled samples and classes of each data set.

B. Experiment Setting

1) Evaluation: We use classification tasks to validate the effectiveness of selected bands. As to evaluation measurements, we make use of the following ones.

TABLE II
NUMBER OF LABELED SAMPLES IN THE INDIAN PINES, PAVIA UNIVERSITY, BOTSWANA, AND MUUFL GULFPORT DATA SETS

No.	Indian Pines		Pavia University		Botswana		MUUFL Gulfport	
	Class name	# Samples	Class name	# Samples	Class name	# Samples	Class name	# Samples
1	Corn-notill	1434	Asphalt	6631	Water	270	Trees	23246
2	Corn-min	834	Meadows	18649	Hippo grass	101	Mostly grass	4270
3	Corn	234	Gravel	2099	Floodplain grasses 1	251	Mixed ground surface	6882
4	Grass-pasture	497	Trees	3064	Floodplain grasses 2	215	Dirt/Sand	1826
5	Grass-trees	747	Metal sheets	1345	Reeds 1	269	Road	6687
6	Hay-windrowed	489	Bare Soil	5029	Riparian	269	Water	466
7	Soybean-notill	968	Bitumen	1330	Firescar 2	259	Building shadow	2233
8	Soybean-mintill	2468	Bricks	3682	Island interior	203	Buildings	6240
9	Soybean-clean	614	Shadows	947	Acacia woodlands	314	Sidewalk	1385
10	Wheat	212	-	-	Acacia shrublands	248	Yellow curb	183
11	Woods	1294	-	-	Acacia grasslands	305	Cloth panels	269
12	Buildings-grass-trees	380	-	-	Short mopane	181	-	-
13	Stone-steel-towers	95	-	-	Mixed mopane	268	-	-
14	Alfalfa	54	-	-	Exposed soils	95	-	-
15	Grass-pasture-mowed	26	-	-	-	-	-	-
16	Oats	20	-	-	-	-	-	-

1) *Overall Accuracy (OA):* This metric is calculated by summing the amount of correctly identified data and dividing by the total amount of data.
2) *Average Accuracy (AA):* This measurement is computed by averaging all per-class accuracies.
3) *Kappa Coefficient:* This coefficient evaluates the agreement between predictions and labels.

2) Band Selection Methods in Comparison: To evaluate the proposed approach, we compare it with several state-of-the-art band selection algorithms that are listed as follows.

1) *MVPCA [36]:* A ranking-based band selection method that uses an eigenanalysis-based criterion to prioritize spectral bands.
2) *ICA [37]:* A band selection approach that compares mean absolute ICA coefficients of individual spectral bands and picks independent ones including the maximum information.
3) *IE [38]:* A ranking-based band selection algorithm in which band priority is calculated based on information entropy.
4) *MEV-SFS [40]:* A searching-based band selection method that combines maximum ellipsoid volume (MEV) [39] method with SFS. The MEV model deems an optimal band subset as a band combination with the maximum volume.
5) *OPBS [40]:* An accelerated version of MEV-SFS that takes advantage of a relationship between orthogonal projections (OPs) and the ellipsoid volume of bands to find out an optimal band combination.
6) *WaLuDi [41]:* A hierarchical clustering-based band selection method that uses Kullback–Leibler divergence as the criterion of the clustering algorithm.
7) *E-FDPC [42]:* A clustering-based band selection approach that makes use of an enhanced version of fast density peak-based clustering (FDPC) [71] algorithm by introducing an exponential learning rule and a parameter to control the weight between local density and intracluster similarity.
8) *OCF [43]:* A band selection method using an optimal clustering algorithm that is capable of achieving optimal clustering results for an objective function with a carefully designed constraint.
9) *ASPS [44]:* A clustering-based band selection method that exploits an adaptive subspace partition strategy.
10) *DARecNet [45]:* An unsupervised convolutional neural network (CNN) for band selection tasks. It employs a dual-attention mechanism, i.e., spatial position attention and channel attention, to learn to reconstruct hyperspectral images. Once the network is trained, bands are selected according to entropies of the reconstructed bands.
11) *DRL:* Our proposed deep reinforcement learning model for unsupervised hyperspectral band selection.

3) Classification Setting: We consider four commonly used classifiers in the remote sensing community to implement hyperspectral image classification. They are as follows.

1) *k-NN:* A k-nearest neighbors algorithm (the number of neighbors is set to 3).
2) *RF:* A random forest being made up of 200 decision trees.
3) *MLP:* A multilayer perceptron that consists of three fully connected layers. The first two layers contain 256 units, and their outputs are activated by Leaky RuLU. For the last layer, the number of units equals the number of classes, and the used activation function is softmax. In the learning phase, we select Adam as the optimizer and define the loss function as categorical cross entropy. The learning rate is set to 0.0005, and the training epochs are 2000 for the purpose of sufficient learning.
4) *SVM-RBF*[3]*:* A support vector machine (SVM) equipped with radial basis function (RBF) kernel. A fivefold cross-validation method is utilized to determine optimal hyperparameters, i.e., γ and C.

For both the Indian Pines and the Botswana data sets, we randomly select 10% samples from each class as training

[3]https://www.csie.ntu.edu.tw/ cjlin/libsvmtools/

TABLE III

COMPARISONS IN QUANTITATIVE METRICS AMONG DIFFERENT BAND SELECTION METHODS ON THE INDIAN PINES DATA SET WITH 30 SELECTED BANDS. WE REPORT THE MEAN AND STANDARD DEVIATION OF PERFORMANCE METRICS OF DIFFERENT APPROACHES OVER TEN INDIVIDUAL RUNS. THE BEST CLASSIFICATION PERFORMANCE IS HIGHLIGHTED IN **BOLD**

Models	k-NN			RF			MLP			SVM-RBF		
	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa
MVPCA [36]	59.45±0.95	48.86±2.22	53.58±1.11	64.50±0.52	55.15±1.43	59.04±0.58	61.85±2.57	56.38±2.18	56.15±2.67	70.10±1.14	62.33±2.83	65.65±1.32
ICA [37]	65.80±0.60	57.72±1.34	60.91±0.70	71.30±0.75	60.90±1.26	66.90±0.90	70.92±1.58	64.51±2.51	66.57±2.06	73.85±1.18	67.75±1.94	69.97±1.38
IE [38]	59.69±0.71	51.31±1.36	53.92±0.82	63.82±0.42	54.84±1.51	58.25±0.44	61.83±1.18	56.64±1.54	55.99±1.45	70.42±0.63	62.46±1.86	66.08±0.69
MEV-SFS [40]	60.71±0.78	52.91±1.99	55.15±0.86	69.21±0.76	57.96±1.18	64.35±0.88	67.11±1.55	60.65±2.32	62.18±1.93	69.72±0.94	63.40±2.08	65.23±1.14
OPBS [40]	63.17±0.58	54.86±1.17	57.90±0.67	71.08±0.70	59.93±1.21	66.61±0.81	69.30±0.89	62.58±1.89	64.75±1.07	71.08±0.80	63.88±1.78	66.86±0.94
WaLuDi [41]	63.27±0.68	52.77±1.98	57.97±0.77	73.18±0.76	58.97±1.70	69.01±0.92	73.36±1.21	65.45±2.19	69.54±1.42	77.52±0.61	69.91±1.73	74.25±0.68
E-FDPC [42]	61.83±0.86	48.52±1.18	56.21±0.96	64.09±0.50	47.72±1.17	58.22±0.60	62.67±1.21	46.75±1.54	56.44±1.34	67.24±0.69	56.86±3.22	62.06±0.74
OCF [43]	63.70±0.92	53.63±1.88	58.45±1.07	73.41±1.01	59.44±1.63	69.31±1.21	74.06±1.02	67.35±1.98	70.35±1.19	78.13±0.64	71.45±1.66	74.98±0.70
ASPS [44]	62.92±0.49	52.20±1.24	57.51±0.58	73.13±0.87	59.11±1.60	68.96±1.04	73.49±0.50	67.19±2.71	69.62±0.61	77.13±0.95	71.37±2.83	73.78±1.12
DARecNet [45]	64.21±0.82	57.72±1.71	59.10±0.93	71.18±0.60	60.65±1.52	66.85±0.70	70.24±0.70	65.08±2.14	65.96±0.78	74.90±0.74	69.46±2.56	71.25±0.85
DRL (proposed)	**67.85±0.63**	**59.85±1.17**	**63.22±0.73**	**74.16±0.79**	**61.47±1.42**	**70.18±0.94**	**75.80±0.76**	**71.32±1.65**	**72.34±0.90**	**78.70±0.86**	**74.39±2.14**	**75.62±1.00**

instances, while the remaining are exploited to test models. Regarding the Pavia University and MUUFL Gulfport data sets, 1% samples per class are chosen randomly to build the training set, and all the other samples are utilized for the purpose of testing. In order to know the stability of various band selection models, final results are achieved by averaging ten individual runs, and we report the mean and standard deviation of performance metrics of different approaches over the ten runs.

C. Information Entropy or Correlation: Whose Call Is It in Building the Reward Scheme?

Fig. 3 compares two instantiations of the reward scheme, namely, information entropy and correlation coefficient (see Section II-A), on the Pavia University data set. To quantitatively evaluate them, we make use of k-NN to perform classification using spectral bands selected by models using these two schemes. From Fig. 3, it can be observed that the former can achieve a higher OA, AA, and Kappa coefficient compared to the latter. Moreover, the computation cost of information entropy is lower than that of the correlation coefficient. Hence, we choose information entropy as the reward scheme in our model for the following experiments.

D. Results and Discussion

In this section, we assess the proposed approach by comparing it with several state-of-the-art band selection methods mentioned in Section III-B. For each data set, we plot OA curves showing OA variations with respect to the number of chosen bands K in Figs. 4–7. In our experiments, K varies from 5 to 60. Furthermore, in Tables III–VI, we also report OAs, AAs, and Kappa coefficients of different methods with a fixed K (following the setup in [52], it is set to 30).

Fig. 4 and Table III present results on the Indian Pines data set. As can be seen in Fig. 4, the proposed DRL is capable of achieving the highest OA using an SVM classifier with 5 to 60 selected bands. Although the OA of DRL is a little bit lower than that of WaLuDi when a k-NN is employed with five bands, our DRL model outperforms other competitors when more spectral bands are chosen. Moreover, we can see that, compared to other band selection models, the proposed approach can also provide gains when using an MLP (the only exception is when $K = 5$). With an RF classifier, OCF and WaLuDi outperform DRL when $K = 5$ and 20, but, in other cases, the proposed model is able to provide the best results. In Table III, we take an example of selecting 30 bands for classification and report numerical results. It can be observed that our DRL obtains the best results. Particularly, when using a k-NN classifier, our approach can gain an improvement of 2.05%, 2.13%, and 2.31% in OA, AA, and Kappa coefficient, respectively, compared with the second best model. In addition, it is noteworthy that, in comparison with original data with all bands, our method can offer almost the same or better results at some point, e.g., when over 20 bands are selected for k-NN.

Fig. 5 and Table IV exhibit classification results for the Pavia University data set. In Fig. 5(a)–(c) (with k-NN, RF, and MLP), when only a few bands are selected, e.g., 5, the OA of DRL is lower than that of WaLuDi and/or E-FDPC. However, when that number goes beyond 5, the proposed method outperforms other competitors. On the other hand, DRL performs well with an SVM classifier, and its OA exceeds the accuracies of all competitors [see Fig. 5(d)]. For instance, Table IV shows that, compared to the second best model, MEV-SFS, and OPBS, our DRL is able to obtain a gain of 1.54% and 2.07% in OA and Kappa coefficient, respectively. Besides, OAs produced by most methods grow when more bands are selected, and our band selection method can achieve higher accuracies than all bands at certain locations, e.g., $K \geq 20$ in Fig. 5(b) and $K = 30$ and ≥ 50 in Fig. 5(d).

Classification results on the Botswana data set are shown in Fig. 6 and Table V. As shown in Fig. 6, the proposed DRL model delivers the best and most stable results with k-NN, RF, and SVM, except for $K = 60$ in Fig. 6(a) and (d). For MLP, DRL performs best when $K = 30$, 50, and 60, and in other cases, it achieves the second best results. Besides, we notice that the performance of DRL and some competitors is very similar when selecting more than 50 spectral bands. However, overall, we can see significant gains in this data set.

In Fig. 7 and Table VI, we report results on the MUUFL Gulfport data set. It can be seen that several band selection models, e.g., WaLuDi, E-FDPC, OCF, ASPS, and DRL, behave very similarly. This may be because, compared to the

TABLE IV

COMPARISONS IN QUANTITATIVE METRICS AMONG DIFFERENT BAND SELECTION METHODS ON THE PAVIA UNIVERSITY DATA SET WITH 30 SELECTED BANDS. WE REPORT THE MEAN AND STANDARD DEVIATION OF PERFORMANCE METRICS OF DIFFERENT APPROACHES OVER TEN INDIVIDUAL RUNS. THE BEST CLASSIFICATION PERFORMANCE IS HIGHLIGHTED IN **BOLD**

Models	k-NN			RF			MLP			SVM-RBF		
	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa
MVPCA [36]	75.76±1.07	68.52±1.42	67.05±1.42	78.83±0.80	70.23±1.96	71.26±1.13	78.00±0.56	69.44±2.14	70.19±0.86	85.43±0.99	77.52±2.85	80.44±1.35
ICA [37]	72.41±0.88	66.39±1.29	62.53±1.16	75.19±0.95	67.83±1.11	65.85±1.05	75.99±1.02	69.93±1.28	67.68±1.37	80.25±1.19	74.19±2.53	73.23±1.87
IE [38]	76.29±1.08	68.95±1.61	67.75±1.44	79.08±0.72	69.98±1.99	71.49±1.06	78.47±0.94	70.75±2.33	70.93±1.20	85.74±1.10	77.75±3.51	80.90±1.54
MEV-SFS [40]	78.25±0.76	73.63±1.30	70.44±0.98	81.40±0.77	75.38±0.90	74.56±0.93	83.01±0.61	79.42±1.21	77.19±0.85	87.51±1.01	83.67±1.74	83.28±1.41
OPBS [40]	78.25±0.76	73.63±1.30	70.44±0.98	81.42±0.86	75.39±0.89	74.58±1.07	82.76±0.92	78.55±1.08	76.84±1.13	87.51±1.01	83.67±1.74	83.28±1.41
WaLuDi [41]	78.30±0.63	73.69±1.05	70.53±0.76	80.45±0.63	74.15±1.61	73.27±0.91	82.93±0.82	78.99±1.49	77.12±1.03	87.19±1.01	82.99±1.96	82.86±1.38
E-FDPC [42]	78.30±0.79	74.55±1.03	70.56±1.05	79.70±0.82	74.03±1.83	72.33±1.07	80.56±1.15	74.92±2.08	73.70±1.38	84.22±0.96	80.72±1.38	78.64±1.31
OCF [43]	78.07±0.82	73.34±1.44	70.21±1.12	80.58±0.77	74.25±1.38	73.47±1.02	82.00±0.86	77.77±1.32	75.77±1.03	86.61±0.82	83.58±0.95	82.09±1.06
ASPS [44]	78.23±0.59	73.80±1.04	70.42±0.79	80.93±0.54	74.87±1.33	73.96±0.72	82.70±1.17	79.19±1.19	76.79±1.50	87.07±0.92	83.53±1.83	82.67±1.27
DARecNet [45]	74.11±0.82	68.21±1.58	64.78±0.96	74.58±0.98	66.71±1.85	65.11±1.25	74.05±1.28	65.38±2.24	64.63±1.42	79.46±0.68	74.30±1.44	71.80±0.90
DRL (proposed)	**80.18±0.74**	**76.13±1.30**	**73.02±0.91**	**81.78±0.64**	**75.41±1.61**	**75.12±0.94**	**84.73±0.91**	**80.74±2.01**	**79.45±1.36**	**89.05±0.61**	**85.85±1.17**	**85.35±0.83**

TABLE V

COMPARISONS IN QUANTITATIVE METRICS AMONG DIFFERENT BAND SELECTION METHODS ON THE BOTSWANA DATA SET WITH 30 SELECTED BANDS. WE REPORT THE MEAN AND STANDARD DEVIATION OF PERFORMANCE METRICS OF DIFFERENT APPROACHES OVER TEN INDIVIDUAL RUNS. THE BEST CLASSIFICATION PERFORMANCE IS HIGHLIGHTED IN **BOLD**

Models	k-NN			RF			MLP			SVM-RBF		
	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa
MVPCA [36]	81.43±1.15	82.80±1.01	79.88±1.24	80.24±1.64	81.04±1.70	78.58±1.77	82.98±0.99	83.04±0.90	81.55±1.07	87.74±1.02	88.95±0.91	86.72±1.11
ICA [37]	82.89±0.85	83.45±1.04	81.46±0.93	81.96±1.10	82.77±1.09	80.45±1.19	87.21±0.77	87.51±0.79	86.14±0.83	88.58±1.05	89.36±1.21	87.62±1.14
IE [38]	79.44±0.86	80.01±0.92	77.72±0.93	78.55±1.29	78.48±1.31	76.74±1.39	85.59±1.05	86.15±1.03	84.39±1.13	88.51±1.29	89.50±1.25	87.55±1.40
MEV-SFS [40]	84.33±0.92	85.53±0.85	83.03±0.99	84.01±0.90	84.93±0.88	82.68±0.98	87.72±0.40	87.80±0.57	86.69±0.43	90.29±1.03	91.26±1.07	89.49±1.12
OPBS [40]	84.33±0.92	85.53±0.85	83.03±0.99	84.20±0.83	85.38±0.98	82.88±0.91	87.70±0.63	87.89±0.84	86.67±0.68	90.29±1.03	91.26±1.07	89.49±1.12
WaLuDi [41]	86.49±0.84	87.82±0.80	85.36±0.91	85.16±0.85	85.90±0.90	83.92±0.92	88.88±0.81	**89.20±1.03**	87.95±0.88	90.69±0.98	91.66±1.01	89.92±1.06
E-FDPC [42]	79.73±1.43	80.62±1.16	78.05±1.54	79.05±1.07	79.68±1.29	77.30±1.16	74.45±1.01	73.60±1.46	72.32±1.10	83.90±1.21	84.54±1.37	82.55±1.32
OCF [43]	84.76±1.12	86.02±0.93	83.50±1.21	83.66±0.73	84.72±0.89	82.31±0.79	86.64±1.02	86.83±1.23	85.53±1.10	89.64±1.25	90.66±1.20	88.77±1.35
ASPS [44]	85.60±0.75	86.96±0.63	84.41±0.81	84.41±0.95	85.20±0.88	83.11±1.02	87.48±0.84	87.17±1.34	86.43±0.92	90.45±0.90	91.42±0.74	89.65±0.98
DARecNet [45]	81.52±0.78	83.21±1.05	79.98±0.85	80.99±0.98	82.53±1.01	79.41±1.06	83.08±1.14	83.65±1.15	81.67±1.24	88.14±0.66	89.25±0.58	87.15±0.71
DRL (proposed)	**86.83±0.70**	**88.28±0.70**	**85.74±0.76**	**86.34±0.84**	**87.36±0.93**	**85.20±0.91**	**88.98±0.68**	88.87±0.86	**88.06±0.74**	**92.14±0.77**	**92.81±0.90**	**91.48±0.83**

TABLE VI

COMPARISONS IN QUANTITATIVE METRICS AMONG DIFFERENT BAND SELECTION METHODS ON THE MUUFL GULFPORT DATA SET WITH 30 SELECTED BANDS. WE REPORT THE MEAN AND STANDARD DEVIATION OF PERFORMANCE METRICS OF DIFFERENT APPROACHES OVER TEN INDIVIDUAL RUNS. THE BEST CLASSIFICATION PERFORMANCE IS HIGHLIGHTED IN **BOLD**

Models	k-NN			RF			MLP			SVM-RBF		
	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa
MVPCA [36]	69.66±0.79	46.47±3.65	59.14±1.07	73.28±1.01	50.62±2.68	63.95±1.49	76.66±0.87	52.43±2.19	68.96±1.15	77.32±1.24	54.43±3.62	69.83±1.65
ICA [37]	78.77±0.89	58.45±2.27	71.81±1.16	80.24±0.56	58.70±2.89	73.89±0.69	82.01±0.42	66.33±2.04	76.23±0.59	81.72±0.85	61.00±3.87	75.81±1.09
IE [38]	69.66±0.79	46.47±3.65	59.14±1.07	73.54±0.80	49.89±2.51	64.31±1.25	77.02±0.60	52.83±1.22	69.49±0.75	77.28±1.20	55.35±3.30	69.74±1.67
MEV-SFS [40]	77.61±0.85	58.25±3.42	70.17±1.17	79.92±0.54	60.39±1.81	73.43±0.69	81.55±0.77	64.30±1.20	75.52±0.88	81.24±0.98	62.40±3.36	75.14±1.24
OPBS [40]	77.61±0.85	58.25±3.42	70.17±1.17	80.08±0.62	59.17±2.73	73.66±0.81	81.88±0.74	64.94±1.77	76.01±1.01	81.24±0.98	62.40±3.36	75.14±1.24
WaLuDi [41]	79.26±0.70	59.38±2.55	72.47±0.96	80.52±0.56	59.85±2.12	74.26±0.69	83.10±0.56	67.69±1.37	77.60±0.79	82.62±0.88	65.48±4.38	77.01±1.16
E-FDPC [42]	78.91±0.78	58.28±1.67	71.96±1.11	80.30±0.55	60.66±1.93	73.94±0.76	82.76±0.73	**69.42±1.88**	77.17±0.98	83.10±0.75	67.58±2.66	77.66±1.01
OCF [43]	79.39±0.67	59.82±1.21	72.65±0.96	80.54±0.66	59.79±2.36	74.28±0.81	83.11±0.78	67.43±1.83	77.58±1.05	82.75±1.02	66.43±4.91	77.21±1.35
ASPS [44]	78.86±0.74	60.10±2.31	71.90±1.08	80.34±0.60	60.38±2.01	74.04±0.72	83.03±0.77	67.28±1.92	77.53±1.04	82.73±0.84	67.32±2.50	77.15±1.09
DARecNet [45]	76.34±0.86	56.82±1.79	68.51±1.20	79.44±0.93	59.94±2.17	72.78±1.22	80.89±1.20	60.25±3.10	74.57±1.68	80.60±0.87	60.35±4.45	74.27±1.15
DRL (proposed)	**79.68±0.63**	**60.85±2.03**	**73.08±0.84**	**80.67±0.52**	**61.08±1.65**	**74.47±0.67**	**83.24±0.75**	68.09±1.70	**77.81±0.96**	**83.33±0.75**	**68.10±2.47**	**77.97±1.02**

other three data sets, the MUUFL Gulfport data set has only 64 bands.

Overall, from the tables, we can see that, among all band selection models, the ranking-based methods perform relatively poorly, while the clustering-based approaches tend to achieve good results. The searching-based models, i.e., MEV-SFS and OPBS, can deliver good selected bands on some data sets, such as the Pavia University scene, but it is noteworthy that they are not robust against different data sets; for example, their performance on the Indian Pines scene is not satisfactory. By contrast, our method shows superior performance. This may be due to the fact that our approach is a data- and objective-driven learning-based model. Compared to other heuristic algorithms, it is able to explore more possible band subsets during the training phase.

In addition, we visualize bands selected by the proposed method on both four data sets in Figs. 8–11. From these figures, we see that DRL tends to select spectral bands with high information entropy. This is in line with our presumption and existing studies in hyperspectral band selection, in which information entropy is an important measurement. Classification maps using 30 bands selected by the proposed DRL model and an SVM-RBF classifier on the four data sets are shown in Fig. 12. Basically, these maps present satisfactory classification results although we see some salt-and-pepper noises that are inevitable in spectral classification.

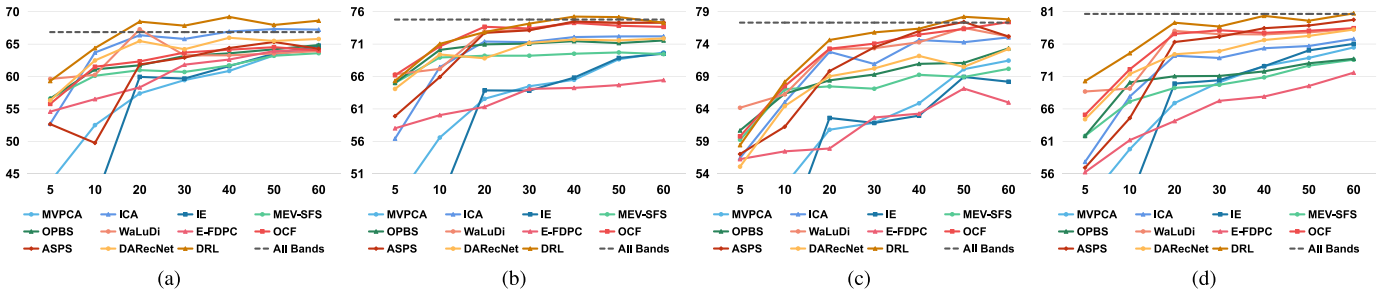


Fig. 4. OA curves of different band selection methods on the Indian Pines data set. The x-axis indicates OA (%), and the y-axis indicates the number of selected bands. (a) OA by k-NN. (b) OA by RF. (c) OA by MLP. (d) OA by SVM-RBF. All OAs are achieved by averaging ten individual runs.

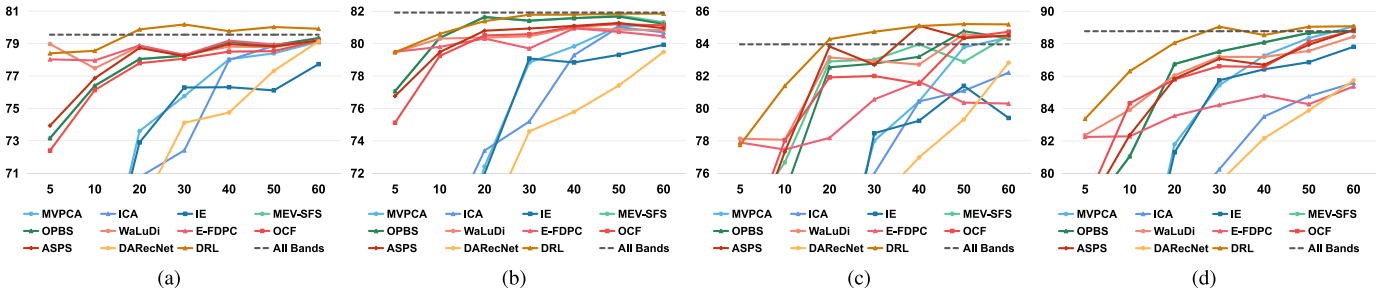


Fig. 5. OA curves of different band selection methods on the Pavia University data set. The x-axis indicates OA (%), and the y-axis indicates the number of selected bands. (a) OA by k-NN. (b) OA by RF. (c) OA by MLP. (d) OA by SVM-RBF. All OAs are achieved by averaging ten individual runs.

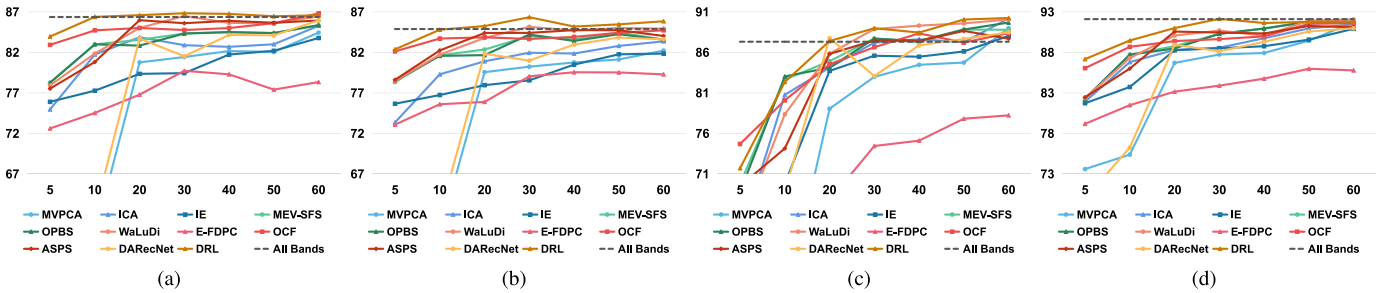


Fig. 6. OA curves of different band selection methods on the Botswana data set. The x-axis indicates OA (%), and the y-axis indicates the number of selected bands. (a) OA by k-NN. (b) OA by RF. (c) OA by MLP. (d) OA by SVM-RBF. All OAs are achieved by averaging ten individual runs.

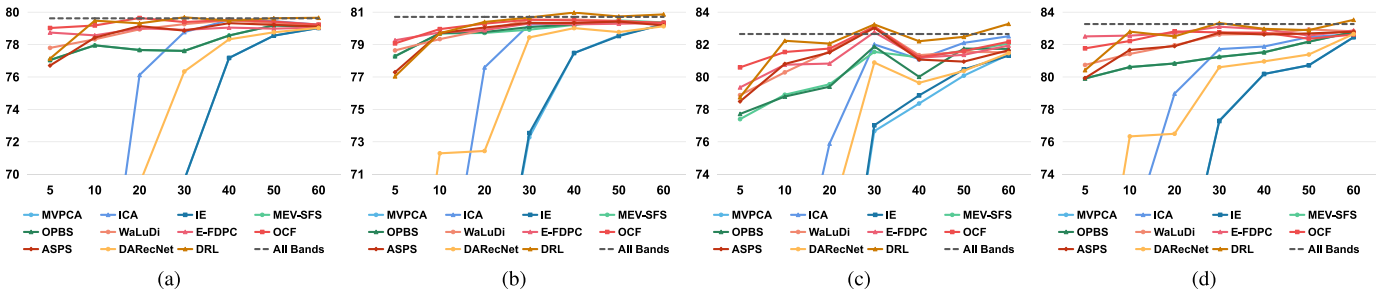


Fig. 7. OA curves of different band selection methods on the MUUFL Gulfport data set. The x-axis indicates OA (%), and the y-axis indicates the number of selected bands. (a) OA by k-NN. (b) OA by RF. (c) OA by MLP. (d) OA by SVM-RBF. All OAs are achieved by averaging ten individual runs.

E. Stability Against Classifiers and Robustness Against Data Sets

From experimental results, we observe that some competitors have unstable behaviors with different classifiers. For example, ASPS works quite well on the Indian Pines data set when RF, MLP, and SVM are employed but a little bit poor when using a k-NN. Similarly, E-FDPC can provide decent results on the Pavia University data set with k-NN, while, with an MLP or SVM classifier, it performs rather poorly compared to other band selection algorithms. This is probably because there exist noisy bands in selected bands, which leads

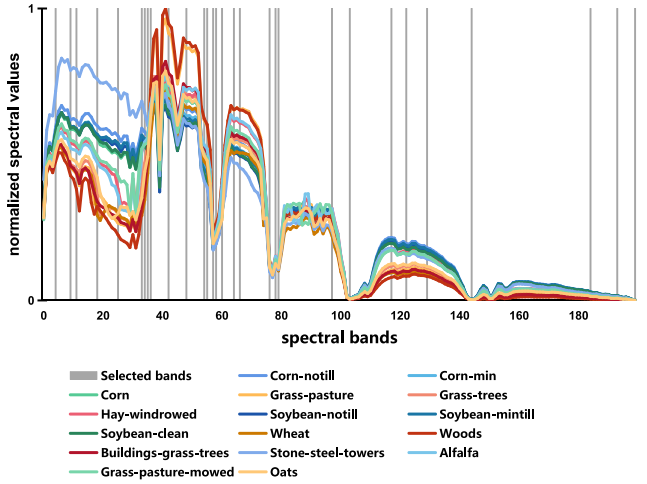


Fig. 8. Visualization of bands selected by the proposed method on the Indian Pines data set. We also show the average spectral signature of each class; 30 bands are selected here.

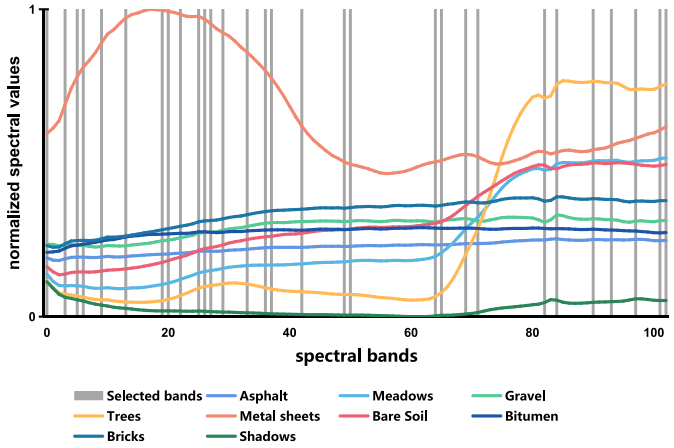


Fig. 9. Visualization of bands selected by the proposed method on the Pavia University data set. We also show the average spectral signature of each class; 30 bands are selected here.

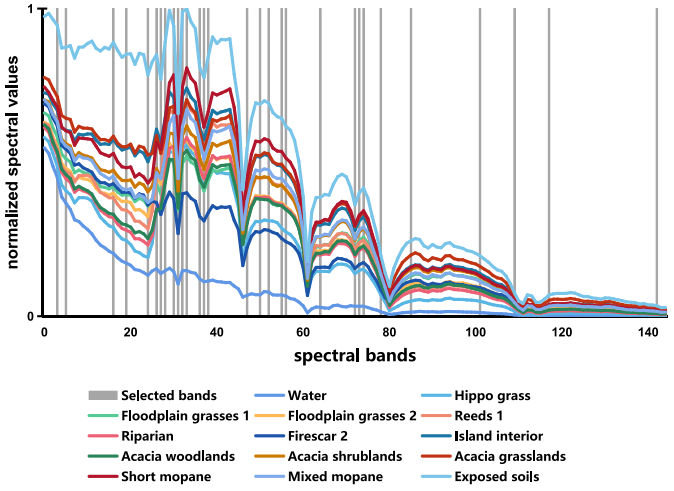


Fig. 10. Visualization of bands selected by the proposed method on the Botswana data set. We also show the average spectral signature of each class; 30 bands are selected here.

to an unsatisfactory performance on noise-sensitive classifiers. In contrast to most competitors, the proposed DRL model is more stable against classifiers.

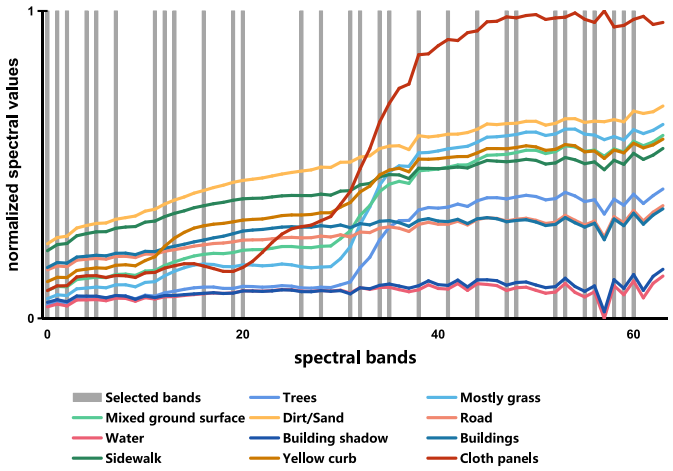


Fig. 11. Visualization of bands selected by the proposed method on the MUUFL Gulfport data set. We also show the average spectral signature of each class; 30 bands are selected here.

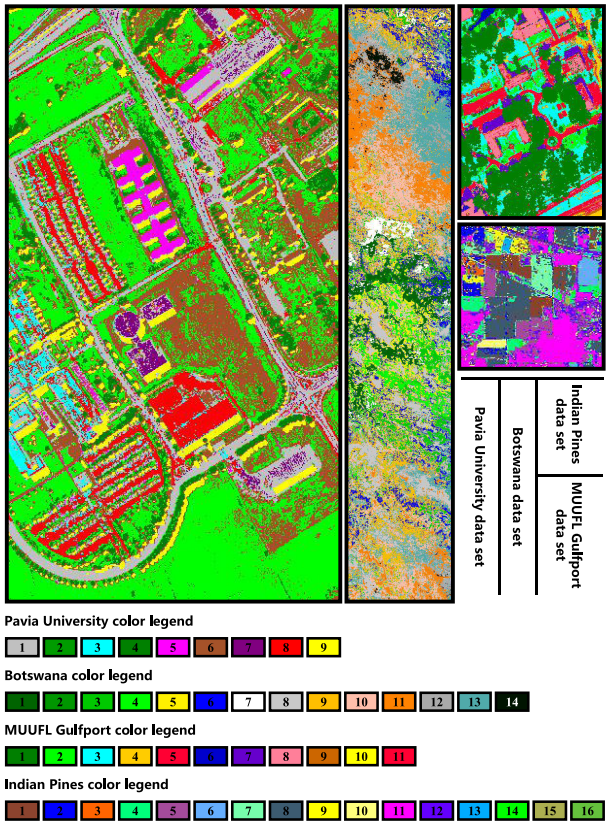


Fig. 12. Classification maps using 30 bands selected by the proposed DRL model and an SVM-RBF classifier on the four data sets.

Furthermore, we also notice that the robustness of several competitors against different data sets is not satisfactory. For example, when 30 bands are selected and making use of an SVM classifier, OCF is capable of achieving the second highest OA and Kappa coefficient on the Indian Pines data set (see Table III) but shows a lackluster performance on the Pavia University and Botswana data sets (see Tables IV and V). This may be because choosing an optimal combination of spectral

bands is a nontrivial task, and locally optimal solutions are not easy to always avoid. In this aspect, the proposed method is more robust against data sets.

F. Limitations

Furthermore, we would like to discuss the limitations of the proposed method. First, as to computational time, compared to other heuristic band selection methods, the proposed model needs more time, as it is a learning-based algorithm and takes some time to explore an effective band-selection policy during the training phase. Taking the Indian Pines data set and 30 selected bands as an example, most heuristic band selection approaches take a few seconds to several tens of seconds [50], and the proposed model needs around 350 s. However, we note that DARecNet [45], a CNN-based unsupervised band selection model, takes about 9000 s under recommended settings. Overall, the computational time of our model is acceptable. Second, since the objective function of our unsupervised DRL is structured such that the learning is aiming to maximize the reward rather than classification accuracy, we cannot intuitively assess the quality of the model in terms of classification accuracy during the training phase, which may lead to unstable model training and the inconvenience of monitoring model training.

IV. Conclusion and Outlook

This article proposes a deep reinforcement learning model for unsupervised hyperspectral band selection. In the training phase, the goal of the deep reinforcement learning agent (i.e., the Q-network) is to learn a band-selection policy that guides the sequential decision-making process of this agent. The policy is a function specifying the band to be chosen given the current state. Note that the training process does not need any labeled data. In the test phase, the agent acts sequentially according to the learned policy. We conduct extensive experiments, and results show the effectiveness of our approach. Moreover, two instantiations of the reward scheme in Section II-A are quantitatively compared, and we believe that more alternatives are possible and may improve results.

In the future, several studies intend to be carried out. For example, combining deep reinforcement learning and some heuristic band selection frameworks (e.g., the clustering-based method) is likely to offer better band selection solutions. Considering that different classes may have different optimal band subsets (with a variable number of bands), how to determine the best band combination for each category is an interesting but challenging problem. A supervised deep reinforcement learning model may be able to provide insights. Moreover, we believe that deep reinforcement learning can be applied to more remote sensing applications, such as multitemporal data analysis, visual reasoning in airborne or space-borne images, and other combinatorial optimization tasks in remote sensing.

Acknowledgment

The authors would like to thank F. Zhang for the helpful discussions on this article.

References

[1] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 45–54, Jan. 2014.

[2] Y. Gu, J. Chanussot, X. Jia, and J. A. Benediktsson, "Multiple kernel learning for hyperspectral image classification: A review," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6547–6565, Nov. 2017.

[3] L. He, J. Li, C. Liu, and S. Li, "Recent advances on spectral–spatial hyperspectral image classification: An overview and new guidelines," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1579–1597, Mar. 2018.

[4] N. Audebert, B. L. Saux, and S. Lefèvre, "Deep learning for classification of hyperspectral data: A comparative review," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 159–173, Jun. 2019.

[5] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Sep. 2019.

[6] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.

[7] Q. Shi, L. Zhang, and B. Du, "Semisupervised discriminative locally enhanced alignment for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4800–4815, Sep. 2013.

[8] J. Li, P. R. Marpu, A. Plaza, J. M. Bioucas-Dias, and J. A. Benediktsson, "Generalized composite kernel framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4816–4829, Sep. 2013.

[9] L. Mou, P. Ghamisi, and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3639–3655, Jul. 2017.

[10] W. Li, G. Wu, F. Zhang, and Q. Du, "Hyperspectral image classification using deep pixel-pair features," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 844–853, Feb. 2017.

[11] L. Mou, P. Ghamisi, and X. X. Zhu, "Unsupervised spectral–spatial feature learning via deep residual Conv–Deconv network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 391–406, Jan. 2018.

[12] J. M. Haut, M. E. Paoletti, J. Plaza, A. Plaza, and L. Plaza, "Hyperspectral image classification using random occlusion data augmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 11, pp. 1751–1755, Nov. 2019.

[13] Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral–spatial residual network for hyperspectral image classification: A 3-D deep learning framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 847–858, Feb. 2018.

[14] M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. J. Plaza, and F. Pla, "Deep pyramidal residual networks for spectral–spatial hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 740–754, Feb. 2019.

[15] Y. Xu, L. Zhang, B. Du, and F. Zhang, "Spectral–spatial unified networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5893–5909, Oct. 2018.

[16] W. Li and Q. Du, "Collaborative representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, Mar. 2015.

[17] Y. Zhang, B. Du, L. Zhang, and S. Wang, "A low-rank and sparse matrix decomposition-based mahalanobis distance method for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1376–1389, Mar. 2016.

[18] Y. Yuan, D. Ma, and Q. Wang, "Hyperspectral anomaly detection by graph pixel selection," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 3123–3134, Dec. 2016.

[19] X. Kang, X. Zhang, S. Li, K. Li, J. Li, and J. A. Benediktsson, "Hyperspectral anomaly detection with attribute and edge-preserving filters," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5600–5611, Oct. 2017.

[20] Z. Huang, L. Fang, and S. Li, "Subpixel-pixel-superpixel guided fusion for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 9, pp. 5998–6007, Sep. 2020, doi: 10.1109/TGRS.2019.2961703.

[21] L. Bruzzone and S. B. Serpico, "An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 4, pp. 858–867, Jul. 1997.

[22] F. Bovolo and L. Bruzzone, "A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 1, pp. 218–236, Jan. 2007.

[23] F. Bovolo, S. Marchesi, and L. Bruzzone, "A framework for automatic and unsupervised detection of multiple changes in multitemporal images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2196–2212, Jun. 2012.

[24] C. Wu, B. Du, and L. Zhang, "Slow feature analysis for change detection in multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2858–2874, May 2014.

[25] F. Bovolo and L. Bruzzone, "The time variable in data fusion: A change detection perspective," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 8–26, Sep. 2015.

[26] M. Zanetti, F. Bovolo, and L. Bruzzone, "Rayleigh-rice mixture parameter estimation via EM algorithm for change detection in multispectral images," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5004–5016, Dec. 2015.

[27] H. Lyu, H. Lu, and L. Mou, "Learning a transferable change rule from a recurrent neural network for land cover change detection," *Remote Sens.*, vol. 8, no. 6, p. 506, Jun. 2016.

[28] Q. Wang, Z. Yuan, Q. Du, and X. Li, "GETNET: A general end-to-end 2-D CNN framework for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 3–13, Jan. 2019.

[29] L. Mou, L. Bruzzone, and X. X. Zhu, "Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 924–935, Feb. 2019.

[30] B. Du, L. Ru, C. Wu, and L. Zhang, "Unsupervised deep slow feature analysis for change detection in multi-temporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9976–9992, Dec. 2019.

[31] W. Zhao, L. Mou, J. Chen, Y. Bo, and W. J. Emery, "Incorporating metric learning and adversarial network for seasonal invariant change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2720–2731, Apr. 2020.

[32] S. Saha, F. Bovolo, and L. Bruzzone, "Unsupervised deep change vector analysis for multiple-change detection in VHR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3677–3693, Jun. 2019.

[33] S. Saha, L. Mou, C. Qiu, X. X. Zhu, F. Bovolo, and L. Bruzzone, "Unsupervised deep joint segmentation of multitemporal high-resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8780–8792, Dec. 2020, doi: 10.1109/TGRS.2020.2990640.

[34] J. Wang and C.-I. Chang, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1586–1600, Jun. 2006.

[35] T. V. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of hyperspectral images with regularized linear discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 862–873, Mar. 2009.

[36] C.-I. Chang, Q. Du, T.-L. Sun, and M. L. G. Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 6, pp. 2631–2641, Nov. 1999.

[37] H. Du, H. Qi, X. Wang, R. Ramanath, and W. E. Snyder, "Band selection using independent component analysis for hyperspectral image processing," in *Proc. 32nd Appl. Imag. Pattern Recognit. Workshop*, Oct. 2003, pp. 93–98.

[38] L. Xie, G. Li, L. Peng, Q. Chen, Y. Tan, and M. Xiao, "Band selection algorithm based on information entropy for hyperspectral image classification," *J. Appl. Remote Sens.*, vol. 11, no. 2, May 2017, Art. no. 026018.

[39] C. Sheffield, "Selecting band combinations from multi spectral data," *Photogramm. Eng. Remote Sens.*, vol. 58, no. 6, pp. 681–687, 1985.

[40] W. Zhang, X. Li, Y. Dou, and L. Zhao, "A geometry-based band selection approach for hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4318–4333, Aug. 2018.

[41] A. MartÌnez-UsÓ, F. Pla, J. M. Sotoca, and P. GarcÌa-Sevilla, "Clustering-based hyperspectral band selection using information measures," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4158–4171, Dec. 2007.

[42] S. Jia, G. Tang, J. Zhu, and Q. Li, "A novel ranking-based clustering approach for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 88–102, Jan. 2016.

[43] Q. Wang, F. Zhang, and X. Li, "Optimal clustering framework for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5910–5922, Oct. 2018.

[44] Q. Wang, Q. Li, and X. Li, "Hyperspectral band selection via adaptive subspace partition strategy," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4940–4950, Dec. 2019.

[45] S. K. Roy, S. Das, T. Song, and B. Chanda, "DARecNet-BS: Unsupervised dual-attention reconstruction network for hyperspectral band selection," *IEEE Geosci. Remote Sens. Lett.*, early access, Aug. 11, 2020, doi: 10.1109/LGRS.2020.3013235.

[46] J. Yin, Y. Wang, and Z. Zhao, "Optimal band selection for hyperspectral image classification based on inter-class separability," in *Proc. Symp. Photon. Optoelectron.*, Jun. 2010, pp. 1–4.

[47] Y.-L. Chang, J.-N. Liu, Y.-L. Chen, W.-Y. Chang, T.-J. Hsieh, and B. Huang, "Hyperspectral band selection based on parallel particle swarm optimization and impurity function band prioritization schemes," *J. Appl. Remote Sens.*, vol. 8, no. 1, Aug. 2014, Art. no. 084798.

[48] Q. Wang, J. Lin, and Y. Yuan, "Salient band selection for hyperspectral image classification via manifold ranking," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 6, pp. 1279–1289, Jun. 2016.

[49] W. Sun, L. Zhang, B. Du, W. Li, and Y. M. Lai, "Band selection using improved sparse subspace clustering for hyperspectral imagery classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2784–2797, Jun. 2015.

[50] W. Sun, J. Peng, G. Yang, and Q. Du, "Fast and latent low-rank subspace clustering for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 3906–3915, Jun. 2020, doi: 10.1109/tgrs.2019.2959342.

[51] S. Patra, P. Modi, and L. Bruzzone, "Hyperspectral band selection based on rough set," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5495–5503, Oct. 2015.

[52] W. Sun and Q. Du, "Hyperspectral band selection: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 118–139, Jun. 2019.

[53] C. Yang, L. Bruzzone, H. Zhao, Y. Tan, and R. Guan, "Superpixel-based unsupervised band selection for classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 12, pp. 7230–7245, Dec. 2018.

[54] X. Zhang, Z. Gao, L. Jiao, and H. Zhou, "Multifeature hyperspectral image classification with local and nonlocal spatial information via Markov random field in semantic space," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1409–1424, Mar. 2018.

[55] J. Feng *et al.*, "Convolutional neural network based on bandwise-independent convolution and hard thresholding for hyperspectral band selection," *IEEE Trans. Cybern.*, early access, Jun. 29, 2020, doi: 10.1109/TCYB.2020.3000725.

[56] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, Jun. 2003.

[57] S. T. Roweis, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[58] J. B. Tenenbaum, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.

[59] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015.

[60] J. C. Caicedo and S. Lazebnik, "Active object localization with deep reinforcement learning," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 2488–2496.

[61] D. Silver *et al.*, "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, Jan. 2016.

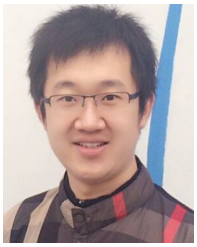
[62] A. Sallab, M. Abdou, E. Perot, and S. Yogamani, "Deep reinforcement learning framework for autonomous driving," *Electron. Imag.*, vol. 2017, no. 19, pp. 70–76, Jan. 2017.

[63] K. Shao, Z. Tang, Y. Zhu, N. Li, and D. Zhao, "A survey of deep reinforcement learning in video games," 2019, *arXiv:1912.10944*. [Online]. Available: https://arxiv.org/abs/1912.10944

[64] R. Bellman, "A Markovian decision process," *J. Math. Mech.*, vol. 6, no. 5, pp. 679–684, 1957.

[65] C. J. C. H. Watkins, "Learning from delayed rewards," Ph.D. dissertation, Dept. Psychol., Univ. Cambridge, Cambridge, U.K., 1989.

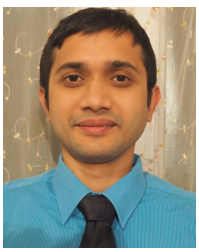
[66] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.
[67] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. Int. Conf. Artif. Intell. Statist. (AISTATS)*, Mar. 2010, pp. 249–256.
[68] T. Dozat. *Incorporating Nesterov Momentum Into Adam*. Accessed: Mar. 22, 2021. [Online]. Available: https://cs229.stanford.edu/proj2015/054_report.pdf
[69] P. Gader, A. Zare, R. Close, J. Aitken, and G. Tuell, "MUUFL Gulfport hyperspectral and LiDAR airborne data set," Univ. Florida, Gainesville, FL, USA, Tech. Rep. REP-2013-570, Oct. 2013.
[70] X. Du and A. Zare, "Technical report: Scene label ground truth map for MUUFL Gulfport data set," Univ. Florida, Gainesville, FL, USA, Tech. Rep. 20170417, Apr. 2017.
[71] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, Jun. 2014.



Lichao Mou received the bachelor's degree in automation from the Xi'an University of Posts and Telecommunications, Xi'an, China, in 2012, the master's degree in signal and information processing from the University of Chinese Academy of Sciences (UCAS), Beijing, China, in 2015, and the Dr.Ing. degree from the Technical University of Munich (TUM), Munich, Germany, in 2020.

In 2015, he spent six months at the Computer Vision Group, University of Freiburg, Freiburg im Breisgau, Germany. In 2019, he was a Visiting Researcher with the Cambridge Image Analysis Group (CIA), University of Cambridge, Cambridge, U.K. He is a Guest Professor with the Munich AI Future Lab AI4EO, TUM, and the Head of the Department "EO Data Science", Visual Learning and Reasoning Team, Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Weßling, Germany. Since 2019, he has been an AI Consultant for the Helmholtz AI. From 2019 to 2020, he was a Research Scientist with DLR-IMF.

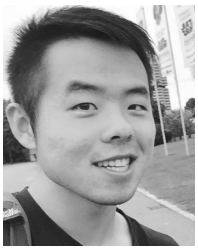
Dr. Mou was a recipient of the First Place in the 2016 IEEE GRSS Data Fusion Contest and a finalist for the Best Student Paper Award at the 2017 Joint Urban Remote Sensing Event and the 2019 Joint Urban Remote Sensing Event.



Sudipan Saha (Member, IEEE) received the B.Tech. degree in electronics and communication engineering from Institute of Engineering & Management, Kolkata, India, in 2011, the M.Tech. degree in electrical engineering from IIT Bombay, Mumbai, India, in 2014, and the Ph.D. degree in information and communication technologies from the University of Trento, Trento, Italy, in 2020.

During his Ph.D. degree, he was with the Fondazione Bruno Kessler (FBK), Trento. He was an Engineer with TSMC Ltd., Baoshan, Taiwan, from 2015 to 2016. He is a Post-Doctoral Researcher with the Technical University of Munich, Munich, Germany, where he was a Guest Researcher in 2019. His research interests are related to self-supervised learning, multitemporal remote sensing image analysis, domain adaptation, image segmentation, and uncertainty quantification.

Dr. Saha was a recipient of the FBK Best PhD Student Award 2020. He is also a reviewer for several international journals.



Yuansheng Hua (Student Member, IEEE) received the bachelor's degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2014, and the double master's degree in earth oriented space science and technology (ESPACE) and photogrammetry and remote sensing from the Technical University of Munich (TUM), Munich, Germany, and Wuhan University in 2018 and 2019, respectively. He is pursuing the Ph.D. degree with the German Aerospace Center (DLR), Weßling, Germany, and TUM.

In 2019, he was a Visiting Researcher with Wageningen University & Research, Wageningen, The Netherlands. His research interests include remote sensing, computer vision, and deep learning, especially their applications in remote sensing.

Francesca Bovolo (Senior Member, IEEE) received the Laurea (B.S.) degree, the Laurea Specialistica (M.S.) degree *(summa cum laude)* in telecommunication engineering, and the Ph.D. degree in communication and information technologies from the University of Trento, Trento, Italy, in 2001, 2003, and 2006, respectively.

She was a Research Fellow with the University of Trento until 2013. She is the Founder and the Head of Remote Sensing for Digital Earth Unit, Fondazione Bruno Kessler, Trento, and a member of the Remote Sensing Laboratory, Trento. She is one of the co-investigators of the Radar for Icy Moon Exploration Instrument of the European Space Agency Jupiter Icy Moons Explorer and a member of the Science Study Team of the EnVision Mission to Venus. Her research interests include remote-sensing image processing, multitemporal remote sensing image analysis, change detection in multispectral, hyperspectral, and synthetic aperture radar (SAR) images, very high-resolution images, time-series analysis, content-based time series retrieval, domain adaptation, and Light Detection and Ranging (LiDAR) and radar sounders. She conducts research on these topics within the context of several national and international projects.

Dr. Bovolo is a member of the program and scientific committee of several international conferences and workshops. She was a recipient of the First Place in the Student Prize Paper Competition of the 2006 IEEE International Geoscience and Remote Sensing Symposium (Denver, 2006). She was the Technical Chair of the Sixth International Workshop on the Analysis of Multitemporal Remote-Sensing Images (MultiTemp 2011 and 2019). She has been the Co-Chair of the SPIE International Conference on Signal and Image Processing for Remote Sensing since 2014. She is the Publication Chair of the International Geoscience and Remote Sensing Symposium in 2015. She has been an Associate Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING since 2011 and a Guest Editor of the Special Issue on Analysis of Multitemporal Remote Sensing Data of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. She is a referee for several international journals.

Lorenzo Bruzzone (Fellow, IEEE) received the Laurea (M.S.) degree *(summa cum laude)* in electronic engineering and the Ph.D. degree in telecommunications from the University of Genoa, Genoa, Italy, in 1993 and 1998, respectively.

He is a Full Professor of telecommunications with the University of Trento, Italy, where he teaches remote sensing, radar, and digital communications. He is also the Founder and the Director of the Remote Sensing Laboratory, Department of Information Engineering and Computer Science, University of Trento. He is the principal investigator of many research projects. Among the others, he is the Principal Investigator of the Radar for icy Moon exploration (RIME) Instrument in the framework of the JUpiter ICy moons Explorer (JUICE) Mission of the European Space Agency (ESA) and the Science Lead of the *High Resolution Land Cover* Project in the framework of the Climate Change Initiative of ESA. He is the author (or a coauthor) of 294 scientific publications in refereed international journals (221 in IEEE journals), more than 340 papers in conference proceedings, and 22 book chapters. He is an editor/a Coeditor of 18 books/conference proceedings and 1 scientific book. His articles are highly cited, as proven from the total number of citations (more than 37 000) and the value of the H-index (89) (source: Google Scholar). He was invited as a keynote speaker in more than 40 international conferences and workshops. His research interests are in the areas of remote sensing, radar, and synthetic aperture radar (SAR), signal processing, machine learning, and pattern recognition. He promotes and supervises research on these topics within the frameworks of many national and international projects.

Dr. Bruzzone has been a member of the Administrative Committee of the IEEE Geoscience and Remote Sensing Society (GRSS) since 2009, where he has been the Vice-President for Professional Activities since 2019. He ranked first place in the Student Prize Paper Competition of the 1998 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Seattle, in July 1998. Since then, he was a recipient of many international and national honors and awards, including the recent IEEE GRSS 2015 Outstanding Service Award, the 2017 and 2018 IEEE IGARSS Symposium Prize Paper Awards, and the 2019 WHISPER Outstanding Paper Award. Since 2003, he has been the Chair of the SPIE Conference on Image and Signal Processing for Remote Sensing. He was a guest coeditor of many special issues of international journals. He is the Co-Founder of the IEEE International Workshop on the Analysis of Multi-Temporal Remote-Sensing Images (MultiTemp) series and a member of the Permanent Steering Committee of this series of workshops. He has been the Founder of the *IEEE Geoscience and Remote Sensing Magazine* for which he has been Editor-in-Chief from 2013 to 2017. He is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He was a Distinguished Speaker of the IEEE Geoscience and Remote Sensing Society from 2012 to 2016.

Xiao Xiang Zhu (Fellow, IEEE) received the M.Sc., Dr.Ing., and Habilitation degrees in signal processing from the Technical University of Munich (TUM), Munich, Germany, in 2008, 2011, and 2013, respectively.

She is a Professor of data science in earth observation (former: signal processing in earth observation) with the Technical University of Munich (TUM) and the Head of the Department "EO Data Science," Remote Sensing Technology Institute, German Aerospace Center (DLR), Weßling, Germany. She was a Guest Scientist or a Visiting Professor with the Italian National Research Council (CNR-IREA), Naples, Italy, Fudan University, Shanghai, China, The University of Tokyo, Tokyo, Japan, and the University of California at Los Angeles, Los Angeles, CA, USA, in 2009, 2014, 2015, and 2016, respectively. Since 2019, she is a co-coordinator of the Munich Data Science Research School, TUM. Since 2019, she also heads the Helmholtz Artificial Intelligence—Research Field "Aeronautics, Space, and Transport." Since May 2020, she has been the Director of the International Future AI Lab "AI4EO—Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond," Munich. Since October 2020, she has been serving as a Co-Director of the Munich Data Science Institute (MDSI), TUM. She is a Visiting AI Professor with ESA's Phi-Lab. Her main research interests are remote sensing and Earth observation, signal processing, machine learning, and data science, with a special application focus on global urban mapping.

Dr. Zhu is also a member of the Young Academy (Junge Akademie/Junges Kolleg) at the Berlin-Brandenburg Academy of Sciences and Humanities and the German National Academy of Sciences Leopoldina and the Bavarian Academy of Sciences and Humanities. She is also an Associate Editor of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and serves as the Area Editor responsible for special issues of *IEEE Signal Processing Magazine*.