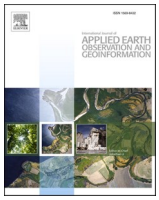
International Journal of Applied Earth Observation and Geoinformation 142 (2025) 104710

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



TEBS: Temperature–Emissivity–Driven band selection for thermal infrared hyperspectral image classification with structured State-Space model and gated attention

Enyu Zhao [a], Nianxin Qu [a], Yulei Wang [a,*], Caixia Gao [b,*], Jian Zeng [c]

[a] The Center for Hyperspectral Imaging in Remote Sensing (CHIRS) at Information Science and Technology College, Dalian Maritime University, Dalian 116026, China
[b] The Key Laboratory of Quantitative Remote Sensing Information Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China
[c] China Centre for Resources Satellite Data and Application, Beijing 100094, China

ARTICLE INFO

Keywords:
Thermal infrared
Unsupervised methods
Hyperspectral image classification
Band selection
Transformer network

ABSTRACT

Thermal infrared hyperspectral images (TIR-HSIs) provide unique spectral insights that are often unattainable with visible imagery, making them invaluable for applications such as land cover classification and geological mapping. However, the high spectral redundancy in TIR-HSIs often leads to increased computational complexity and potential performance degradation. To address this issue, this paper proposed an unsupervised temperature–emissivity–driven band selection method (TEBS) for TIR-HSIs classification, which integrated a structured state-space model (SSM) and a gated attention mechanism (GAM). Specifically, a feature extraction (FE) module is firstly designed to separate land surface temperature (LST) and land surface emissivity (LSE) information, incorporating superpixel segmentation to extract multi-scale LST features. Subsequently, a weight computation (WC) module, leveraging SSM and GAM, is developed to generate robust band weights by sequentially leveraging multi-scale LST features. Finally, a band evaluation (BE) module is employed to assess the band selection results and optimize the model parameters. Experimental comparisons conducted on two datasets using four classic classifiers show that TEBS framework outperforms state-of-the-art (SOTA) methods in classification accuracy. These results underscore the potential of TEBS to advance land cover classification in thermal infrared hyperspectral imaging. The data and code will be made publicly available at: https://github.com/Qu-NX/TEBS.

1. Introduction

Thermal infrared hyperspectral remote sensing technology facilitates the acquisition of radiative information of objects within the emissivity spectral range, which typically spans from 8 to 14 μm (Neinavaz et al., 2016). In contrast to visible and near-infrared hyperspectral images, thermal infrared hyperspectral images (TIR-HSIs) possess distinct spectral characteristics that reflect the thermal radiation emitted by objects, providing critical insights into both the temperature and material composition of the observed entities (Gao et al., 2021; Manolakis et al., 2019). Owing to their unique properties, TIR-HSIs have exhibited significant potential for application across a diverse array of fields, including military reconnaissance, medical diagnostics, land cover classification, and object detection (Aslett et al., 2018; Zhu et al., 2021).

The methods of land cover classification from TIR-HSIs exhibit significant differences compared to those employed in the visible and near-infrared imagery. In the thermal infrared spectrum, the key information captured by sensors includes land surface temperature (LST) and land surface emissivity (LSE) (Ren et al., 2018; Zhao et al., 2024a). LSE is defined as the ratio of the radiation energy emitted by an object at a specific wavelength to that emitted by a blackbody, thereby reflecting the surface characteristics of the object. By analyzing the LSE in TIR-HSIs, it becomes feasible to more accurately identify and distinguish different land covers, including various mineral types, vegetation species, or surface materials (Ermida et al., 2023; Rock et al., 2016; Baldridge et al., 2009). Furthermore, TIR-HSIs exhibit sensitivity to the LST, and the temperature information associated with different land covers is often closely correlated with their material composition, thereby offering an additional pathway for land cover classification (Taspinar, 2023; Batchuluun et al., 2022). Additionally, TIR-HSIs are capable of

* Corresponding authors.
E-mail addresses: wangyulei@dlmu.edu.cn (Y. Wang), gaocaixia@aoe.ac.cn (C. Gao).

https://doi.org/10.1016/j.jag.2025.104710
Received 6 February 2025; Received in revised form 5 June 2025; Accepted 29 June 2025
Available online 4 July 2025
1569-8432/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

functioning effectively in night-time conditions, as their operational principle relies on the inherent thermal radiation emitted by objects rather than relying on external light sources. This characteristic enables stable and reliable data acquisition even under low-light or night-time conditions (Cui et al., 2022; Weng, 2009). However, since TIR-HSIs are acquired by hundreds or even thousands of contiguous spectral bands (Zhao et al., 2024b), each spectral band is able to uncover and reveal subtle material substances. However, this is also traded for significant correlation among bands. In particular, many of spectral bands may convey similar information or contribute minimally to specific classification tasks. This results in many spectral bands which are actually redundant. So, finding an appropriate band subset not only can preserve information integrity but also reduce computational complexity, most importantly, can alleviate model training complexity (Xu et al., 2021; Sui et al., 2020). Consequently, it is imperative to develop effective band selection methods to eliminate redundant bands, while ultimately enhancing both the efficiency and accuracy of classification processes.

Currently, dimensionality reduction techniques for hyperspectral data can be broadly categorized into two main approaches: band feature extraction and band selection (Wang et al., 2022; Wang et al., 2020). Band feature extraction refers to a method that transforms the characteristics of the original bands into new features within a lower-dimensional space. Common methods for band feature extraction include Principal Component Analysis (PCA) (Uddin et al., 2021), Linear Discriminant Analysis (LDA) (Aved et al., 2017), and Independent Component Analysis (ICA) (Johnson et al., 2013), among others. These methods effectively reduce the dimensionality of data by mapping high-dimensional data to a new feature space through either linear or nonlinear transformations. A significant advantage of band feature extraction lies in its ability to substantially decrease the dimensionality of the dataset while preserving as much information as possible. However, a notable drawback is that the newly transformed features often lack physical meaning, which complicates the interpretation of the specific contributions of individual bands to the classification task (Sun et al., 2022a). On the other hand, band selection methods focus on retaining only the most representative or discriminative bands from hyperspectral data while eliminating redundant bands (Ou et al., 2023). A variety of band selection algorithms have been proposed, some of which employ clustering techniques to group similar bands and select representative bands from each cluster. These approaches effectively reduce redundancy and enhance efficiency. For instance, Sun et al. (Sun et al., 2022b) proposed a band selection method based on hyperbolic clustering that leveraged geometric information for unsupervised band selection in hyperspectral images. Another prevalent strategy involves scoring and ranking the bands according to their significance, subsequently selecting those with the highest scores. Such methods are straightforward, intuitive, and easy to interpret. For example, Chang et al. (Chang et al., 1999) developed a band selection algorithm aimed at image classification that prioritized bands through eigenvalue decomposition while incorporating divergence information to identify the most representative subset of bands. Additionally, certain methods employ optimization algorithms to identify the optimal combination of bands that maximizes classification performance. An example is provided by Wan et al. (Wan et al., 2023), who proposed an adaptive multi-strategy particle swarm optimization technique designed to address local optima stagnation issues in band selection by employing a multi-strategy evolutionary model and dynamically adjusting motion parameters, thereby effectively minimizing redundant information within hyperspectral images.

In recent years, deep learning methods have demonstrated significant potential and advantages in hyperspectral band selection. Traditional band selection approaches typically rely on feature engineering or criteria based on prior knowledge to evaluate and select bands, whereas deep learning methods can automatically learn and optimize band selection strategies. Deep learning approaches can be further categorized into supervised, unsupervised, and semi-supervised methods. Supervised methods leverage large amounts of labeled data to train models. For example, Zhan et al. (Zhan et al., 2017) proposed a hyperspectral band selection method based on deep convolutional neural networks and distance density, which enhanced computational efficiency by partitioning the distance density subspaces. However, obtaining data labels is often challenging in real-world scenarios, prompting the development of unsupervised methods. Unsupervised approaches can automatically uncover intrinsic relationships within the data without the need for labels, thereby exhibiting broad applicability. For instance, Goel and Majumdar (Goel and Majumdar, 2022) introduced a deep transformation learning method that integrated K-Means clustering to achieve hyperspectral band selection by combining feature extraction with clustering. Moreover, semi-supervised band selection algorithms have emerged as a research hotspot, as these methods require only a small amount of data for training. For example, Feng et al. (Feng et al., 2021) proposed a deep reinforcement learning-based band selection method that facilitated efficient band search and evaluation, thereby improving the classification performance of hyperspectral images.

Although numerous advanced band selection algorithms have been proposed, most of them are specifically designed for visible and near-infrared hyperspectral images. When these algorithms are directly applied to TIR-HSIs for classification tasks, performance degradation may occur due to the differing physical interpretations of the data involved. To address this issue, this paper introduces an unsupervised temperature–emissivity–driven band selection method (TEBS) for TIR-HSI classification. The model simultaneously computes spatial and spectral dimensional features and employs structured state-space models (SSMs) alongside gated attention mechanisms (GAMs) to perform band selection for thermal infrared hyperspectral data. Specifically, this study begins by decomposing TIR-HSIs into LSE and LST images, extracting LSE features for each band through depthwise separable convolutions. Subsequently, both LST and LSE features are fed into a network model composed of multiple computational modules that synergistically combine SSM and GAM. During this process, LST features derived from various superpixel segmentation scales contribute to the computation of band weights, thereby enhancing the spatial feature extraction related to land covers and enabling more accurate weight assignments for the bands. Finally, an unsupervised band evaluation network is employed to assess the quality of the selected bands and update the model parameters accordingly.

The main contributions of this paper can be summarized as follows:

1) An unsupervised band selection method is proposed for TIR-HSIs classification: To address the challenge of limited labelled data in remote sensing, this paper proposes a novel unsupervised band selection framework. This method, validated across multiple datasets and tested with various classifiers, effectively reduces spectral redundancy and computational complexity while preserving high classification accuracy, demonstrating its robustness and generalizability.
2) A multi-modal band selection network is introduced leveraging domain-specific features: This paper introduces a band selection framework that integrates LST and LSE features. By employing superpixel segmentation to extract multi-scale LST characteristics, this methodology significantly improves the performance of band selection algorithms. This highlights the value of incorporating domain-specific physical properties in hyperspectral image processing.
3) An innovative WC module is designed combining SSM and GAM: This work provides a novel perspective on the relationship between SSM and Transformer architectures, thereby providing valuable insights into the application of SSM in the field of remote sensing.

The remainder of this paper is organized as follows. Section 2 reviews the related researches on Transformers and SSMs. A

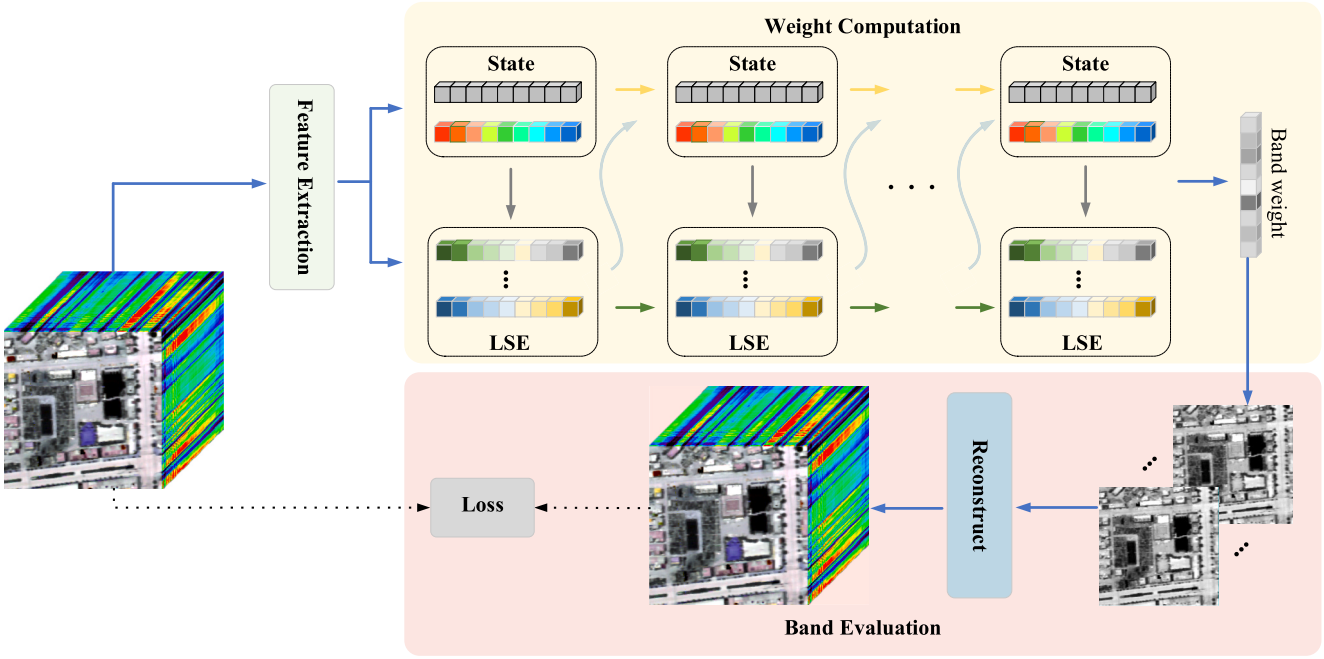


Fig. 1. Structure of the temperature–emissivity–driven band selection (TEBS) model.

comprehensive explanation of the proposed methodology is presented in Section 3. Section 4 provides a detailed account of the comparative experiments, ablation studies, sensitivity analysis, and hyperparameter analysis. Finally, Section 5 offers concluding remarks for the paper.

2. Related work

With the advancement of computational capabilities, deep learning algorithms have been widely applied in the remote sensing field, demonstrating superior performance. The method proposed in this paper leverages Transformer for processing sequential data and employs the SSM model for data control. An overview of Transformer and SSM is provided below.

2.1. Transformer models

The Transformer model was proposed by Vaswani et al. in 2017, originally designed for machine translation tasks within the realm of natural language processing (NLP) (Vaswani et al., 2017). This model replaces traditional recurrent neural networks with an attention mechanism, effectively capturing long-range dependencies among sequence elements, which significantly enhances the performance of Transformers across various NLP tasks. In the domain of computer vision, Dosovitskiy et al. (Dosovitskiy et al., 2021) introduced the Vision Transformer (ViT), marking a pioneering application of the Transformer architecture to image classification. ViT segments an image into a series of fixed-size patches and treats these patches as an input sequence that is processed by a Transformer encoder akin to those utilized in NLP tasks. By leveraging self-attention mechanisms, ViT captures global contextual information and demonstrates remarkable performance on large-scale datasets, thereby facilitating the widespread adoption of Transformers in computer vision. In recent years, Transformers have gradually been introduced into the remote sensing field, with applications spanning image classification (Wang et al., 2023; Ni et al., 2024), object detection (Wang et al., 2024a; Zhou et al., 2022), and super-resolution (Long et al., 2023; Yang et al., 2024). When addressing the challenge of data redundancy in hyperspectral images, Transformer architectures and attention mechanisms exhibit exceptional performance. For example, Li et al. (Li et al. 2021) proposed a non-local attention network to capture long-range dependencies among spectral bands, achieving outstanding results. Additionally, Cui et al. (Cui et al. 2024) developed a spatial-spectral cross-dimensional attention network that effectively reduced redundancy in hyperspectral images. In conclusion, Transformers demonstrate significant research potential in tackling the issue of hyperspectral data redundancy.

2.2. Structured State-Space models

SSM represents an emerging architecture for sequence modeling, drawing inspiration from the principles of classical state-space models, such as the Kalman filter (Gu et al., 2022). Dao and Gu (Dao and Gu, 2024) further refined the theoretical foundations of SSMs and introduced the Mamba2 model. Leveraging the exceptional performance and significant potential of Mamba2, SSMs have attracted increasing attention and have been applied in the field of remote sensing. For instance, Zhao et al. (Zhao et al., 2024c) proposed a model specifically designed for large-scale dense prediction tasks in remote sensing images. By incorporating an all-directional selection scanning module, this model achieves global modeling capabilities with linear complexity, thereby enhancing performance in semantic segmentation and change detection tasks. Zhang et al. (Zhang et al., 2024) developed a remote sensing image segmentation network based on link aggregation jump connections that integrated multi-scale feature aggregation with a hybrid self-attention mechanism to consolidate multi-dimensional information and facilitate cross-scale semantic exchange, thus improving accuracy and generalization in semantic segmentation tasks. Overall, SSMs are increasingly finding applications in remote sensing and are gradually demonstrating their research value.

3. Proposed method

To fully harness the LST information within TIR-HSIs, this study introduces a band selection methodology that integrates both LST and LSE, as illustrated in Fig. 1. In summary, the proposed method comprises three modules: Feature Extraction (FE), Weight Computation (WC), and Band Evaluation (BE). Initially, the LST and LSE features are extracted from TIR-HSIs. Subsequently, the weights of each band are computed, and a subset of bands to be retained is selected based on their respective

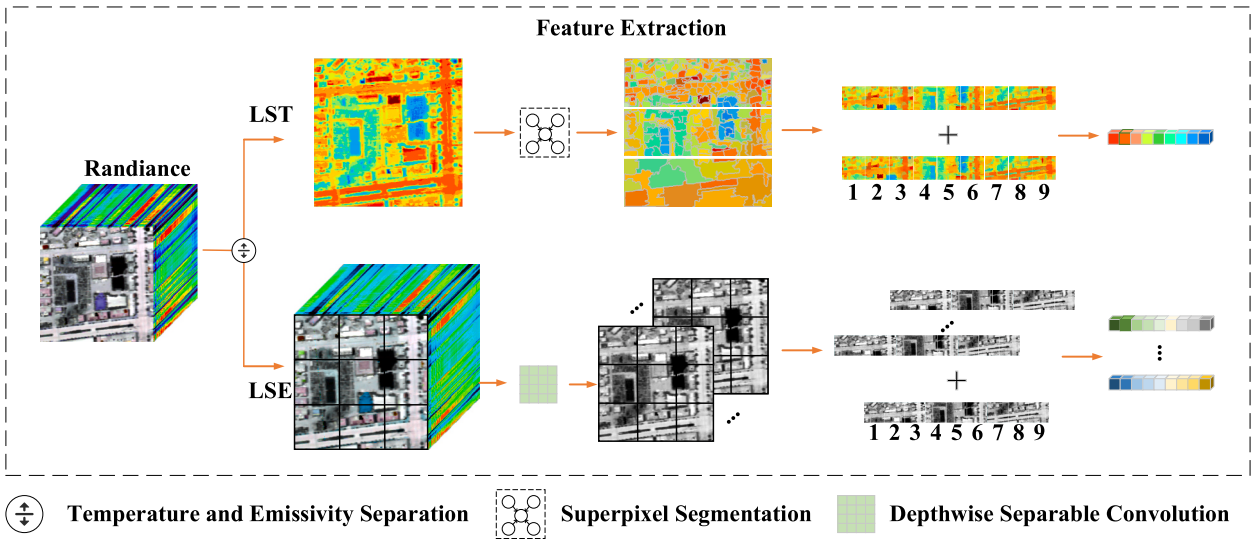


Fig. 2. Structure of the Feature Extraction (FE) module.

weights. Finally, an evaluation network is employed to assess the outcomes of the band selection process through band reconstruction.

3.1. Feature extraction (FE) module

Unlike hyperspectral images in the visible spectrum, TIR-HSIs not only capture the spectral characteristics of terrestrial targets but also their thermal properties. To decouple these two aspects, the Temperature and Emissivity Separation (TES) algorithm (Gillespie et al., 1998) is employed. This algorithm estimates LST and LSE from multi-band radiance data by utilizing the empirical relationship between spectral contrast and minimum emissivity.

Due to variations in physical properties such as specific heat capacity and thermal conductivity, different types of land cover typically exhibit distinct temperature variations. This temperature differential is usually closely associated with surface factors such as vegetation cover and soil moisture (He et al., 2023; Liu et al., 2023a), and manifests itself through texture features that are significant for research in target detection, object classification, and related fields.

However, due to the continuous nature of temperature variations between adjacent objects, quantifying or directly comparing these temperature differences poses significant challenges. For example, the water temperature at the center of a lake may differ from that at the edge–a phenomenon referred to as "the same material with different temperatures", which arises from factors such as energy exchange among substances. Additionally, under specific thermal equilibrium conditions, various land cover types within the same environment may demonstrate the occurrence of "different materials with the same temperature". Consequently, while the LST images contain substantial amounts of critical information, improper analysis can lead to diminished model accuracy.

To address this issue, the present study proposes a dedicated FE module, as illustrated in Fig. 2. Specifically, given the input radiance data $RD \in \mathbb{R}^{H\times W\times C}$, the TES algorithm is initially employed to obtain LST data $LTD \in \mathbb{R}^{H\times W}$ and LSE data $LED \in \mathbb{R}^{H\times W\times C}$. Subsequently, the Simple Linear Iterative Clustering (SLIC) algorithm (Achanta et al., 2012) is applied to the LST image for superpixel segmentation. This process divides the LST image into multiple superpixel regions composed of pixels with similar temperature characteristics. By integrating both temperature values and spatial positions of pixels, SLIC effectively captures the spatial variation patterns of LST. Following this step, a sliding window method is adopted to extract local features from both LST and LSE data. For each pixel within a fixed-size window measuring 10×10 pixels, this approach enables the capture of spatial variation and texture

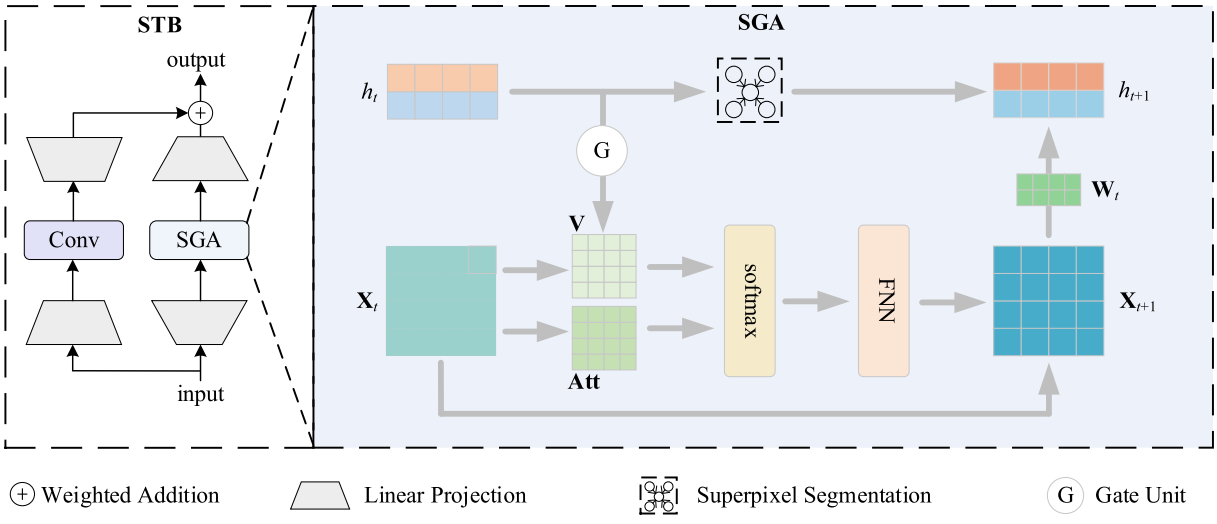


Fig. 3. Structure of the SSM-Transformer Block (STB).

information within its neighborhood. To fully leverage these spatial features, positional encoding is integrated into both LST and LSE data within the window context, allowing the model to perceive and utilize pixel locations in relation to one another within the image. Ultimately, one feature vector is generated per spectral channel resulting in C feature vectors–where C represents the number of spectral channels. Each feature vector not only retains the spectral information corresponding to its respective channel but also integrates local spatial features along with positional cues.

Overall, this methodology effectively addresses the challenges associated with "the same material with different temperatures" as well as "different materials with the same temperature". Through multi-scale clustering and superpixel segmentation techniques, it incrementally extracts temperature information from micro- to macro-scales, thereby revealing fundamental properties of surface materials in a more comprehensive and nuanced manner.

3.2. Weight computation (WC) module

The WC module assigns weights to each spectral band based on LST and LSE features. This module comprises several computational blocks, referred to as SSM-Transformer Blocks (STBs), which integrate the characteristics of SSM and Transformer. The operational mechanism involves calculating band correlations utilizing LSE features while introducing LST as an auxiliary feature, thereby progressively determining band weights across multiple scales. Within each STB, the fundamental computational unit is a State-Gated Attention (SGA) structure that synergizes SSM with a GAM, as shown in Fig. 3.

Specifically, the input data to the SGA comprises two components: the LSE information $\boldsymbol{X} \in \mathbb{R}^{C\times F}$ and the state information. The state information encompasses the LST feature $\boldsymbol{T} \in \mathbb{R}^{F}$ and the band weight feature $\boldsymbol{BW} \in \mathbb{R}^{C}$. Here, F denotes the dimensionality of the feature mapping. The computation process can be interpreted from two perspectives. Focusing on a specific LST state, the SGA can be viewed as a gated multi-head attention mechanism. For each attention head, the query matrix $\boldsymbol{Q}$, key matrix $\boldsymbol{K}$, and value matrix $\boldsymbol{V}$ can be defined as follows:

$$\boldsymbol{Q} = \boldsymbol{W^Q X}; \boldsymbol{K} = \boldsymbol{W^K X}; \boldsymbol{V} = \boldsymbol{W^V X} \tag{1}$$

where, $\boldsymbol{W^Q}$, $\boldsymbol{W^K}$, and $\boldsymbol{W^V}$ are learnable parameter matrices. The input data $\boldsymbol{X}$ encompasses the features across all bands. $\boldsymbol{Q}$ refers to the "query" generated by a specific band within the entire sequence of bands, aimed at assessing the correlation between that band and other bands. $\boldsymbol{K}$ represents the sequence data being queried. This computation can be represented as follows:

$$\boldsymbol{Score} = \boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}} \tag{2}$$

where, the $\boldsymbol{Score}$ represents the attention matrix, which illustrates the correlation between $\boldsymbol{Q}$ and $\boldsymbol{K}$ within the feature space. A larger value of an element in the $\boldsymbol{Score}$ indicates a stronger correlation between the corresponding bands. To mitigate excessively large values resulting from the inner product, these values are typically transformed into weights with the calculation performed as follows:

$$\boldsymbol{Att} = \mathrm{softmax}\left(\frac{\boldsymbol{Score}}{\sqrt{d_K}}\right) \tag{3}$$

where, $\boldsymbol{Att}$ represents the attention weights, d_{K} is the dimension of the keys for each attention head, and softmax(·) denotes a nonlinear activation function. Its primary purpose is to ensure that the sum of all output values equals 1, thereby allowing the output to be interpreted as a probability distribution and while enhancing numerical stability. For a traditional attention mechanism, the output of a given attention head is:

$$\boldsymbol{O} = \boldsymbol{Att} \times \boldsymbol{V} \tag{4}$$

where, $\boldsymbol{O}$ represents the output of the attention mechanism, which is the weighted sum of $\boldsymbol{V}$ using the attention weights, yielding an output that integrates the important band information.

Compared to the traditional attention mechanism, this approach integrates a gating mechanism utilizing $\boldsymbol{T}$ and $\boldsymbol{BW}$, effectively combining LST and band weight information with LSE. The computational process can be articulated as follows:

$$\boldsymbol{V_G} = (\boldsymbol{BW} \times \boldsymbol{W_G} \times \boldsymbol{T})\boldsymbol{V} \tag{5}$$

where, $\boldsymbol{V_G}$ denotes the values obtained following the application of the gating mechanism, and $\boldsymbol{W_G}$ is a trainable weight matrix designed to aggregate the information from both LST and band weights. Therefore, the output generated by the attention mechanism can be expressed as:

$$\boldsymbol{O_G} = \boldsymbol{Att} \times \boldsymbol{V_G} \tag{6}$$

By incorporating the gating mechanism, the features derived from LST are integrated with those from LSE. This approach effectively couples multiple data sources with physical significance, thereby enhancing the model's representational capacity. On the other hand, when considering the entire module as a sequence, it can also be considered a generalized SSM model. Specifically, within this module comprising multiple SGAs, each SGA receives input consisting of the state h and the data $\boldsymbol{X}$. Let t represent the time scale that varies according to the multi-scale superpixel segmentation of the LST image. Consequently, h_t={$\boldsymbol{BW}$, $\boldsymbol{T}$} represents the state information of an SGA at time t, while $\boldsymbol{X}_t$ signifies the LSE information fed to that SGA at time t. If we concentrate on examining the information flow between SGAs while disregarding residual connections and linear mappings, h_{t+1} and $\boldsymbol{X}_{t+1}$ can be articulated as:

$$h_{t+1} = Sp_{down}(h_t) + \boldsymbol{O_{Gt}}\boldsymbol{W}_t \tag{7}$$

$$\boldsymbol{X}_{t+1} = \boldsymbol{X}_t + \boldsymbol{O_{Gt}} \tag{8}$$

where, $\mathrm{Sp_{down}}(\cdot)$ represents the superpixel segmentation applied to the LST image in h for down-sampling. $\boldsymbol{O_{Gt}}$ denotes the output of the gated attention operation applied to the input $\boldsymbol{X}_t$ under the control of h_t; accordingly, $\boldsymbol{O_{Gt}}$ can be regarded as a function of h_t and $\boldsymbol{X}_t$. $\boldsymbol{W}_t$ signifies a trainable matrix that facilitates the mapping of LSE features to weight features. Based on (7) and (8), the multiple SGAs in the WC module can be interpreted as a generalized SSM.

Overall, the integration of multiple STBs within the TEBS forms a generalized SSM, which comprehensively extracts the spatial information and variation patterns from LST images across various segmentation scales. Meanwhile, the SGA embedded in each STB utilizes an LST image at a specific segmentation scale as auxiliary information to participate in the computation of the inter-band correlations. This structure facilitates a more thorough extraction of surface spatial features and enhances the calculation of correlations among thermal infrared hyperspectral bands. Under the simplified assumption that linear mappings and activation functions are ignored, the computation flow of the WC module proceeds as follows:

Algorithm 1 Weight Computation module

Input:
Feature $\boldsymbol{X}_0$ containing LSE information
State feature h_0 containing LST and band weight information
Output:
1 While Iterating over LST superpixel feature maps of all sizes:
Do
2 Advance to the next time step t and update the LST superpixel features.
3 Compute $\boldsymbol{Q}$, $\boldsymbol{K}$, $\boldsymbol{V}$ and $\boldsymbol{Score}$ at time step t using (1) and (2).
4 Calculate $\boldsymbol{Att}$ and $\boldsymbol{V_G}$ at time step t using (3) and (5).
5 Compute the output of the gated attention mechanism, $\boldsymbol{O_{Gt}}$, using (6).
6 Calculate $\boldsymbol{X}_{t+1}$ and h_{t+1} using (7) and (8).
Return the band weight matrix $\boldsymbol{BW}$ from h_{t+1}

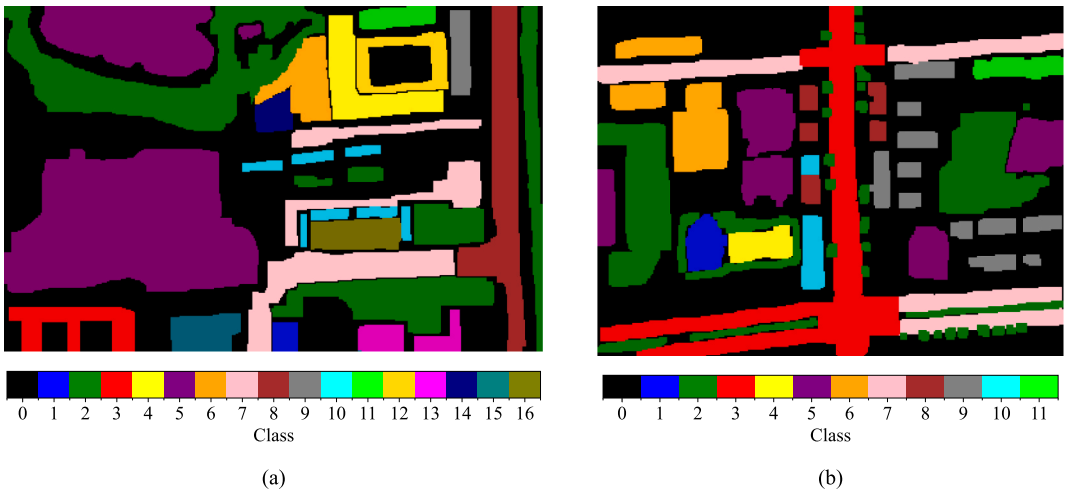


Fig. 4. Colorful ground-truth images used in this paper, with class 0 representing the background. (a) Ground-truth of Suburb dataset, divided into 16 classes. (b) Ground-truth of Urban dataset, divided into 11 classes.

3.3. Band evaluation (BE) module

Hyperspectral data typically encompass hundreds or even thousands of spectral bands; however, not all bands hold equal significance in conveying information. Consequently, selecting a representative subset of bands is crucial for reducing both data dimensionality and computational cost. This method introduces an unsupervised band evaluation module designed to assess the representativeness of the selected band subset by reconstructing spectral information. Specifically, the method begins with the selection of a band subset based on predetermined band weights, which serves as the input to the band evaluation module. Subsequently, an unsupervised band evaluation module comprising an encoder and a decoder is constructed. The encoder is responsible for mapping the input band subset into a low-dimensional latent space, while the decoder reconstructs the complete spectral data from this latent space. During training, the discrepancy between the reconstructed spectral data and the original full spectral data is utilized as the loss function, which can be expressed as:

$$Loss = \frac{1}{N}\sum_{i\in N} -\log\left(\frac{X_i^T Y_i}{|X_i||Y_i|}\right) \tag{9}$$

where, N represents the total number of pixels in the image, X_i denotes the predicted value of the pixel with index i, and Y_i represents the ground truth of the pixel with index i. By assessing the magnitude of the reconstruction error, the representativeness of the selected band subset can be determined. A minimal reconstruction error indicates that this band subset encompasses sufficient critical information to effectively represent the original high-dimensional data. Conversely, a larger error implies that the chosen bands may be deficient in essential information, thereby indicating inadequate representativeness.

4. Results and analysis

In this section, the two datasets utilized in the experiments are first introduced. Subsequently, the proposed method is compared against several SOTA methods to validate its effectiveness and computational performance. An ablation study follows, demonstrating the beneficial impact of incorporating LST information and the SGA structure. Next, sensitivity analysis experiments are performed to assess the robustness of the TEBS method against image noise. Finally, the impact of model hyperparameter settings on TEBS is evaluated through hyperparameter analysis.

4.1. Hyperspectral datasets

The datasets are acquired from two TIR-HSI images captured by an airborne system, as illustrated in Fig. 4. The images are obtained in Hengdian Town, Dongyang City, Zhejiang Province, China, on May 24, 2019. Both images have a spatial resolution of 1 m, cover a spectral range from 8.061 to 11.217 μm, and contain 110 spectral bands with a full width at half maximum of 38 nm. The original images represent radiance measurements expressed in unit of W/(m^2·sr·μm), and their high spectral resolution enables the extraction of abundant surface information. Sample annotation was conducted using ENVI software. Initially, labels were generated via the ENVI provided built-in classification tool. Subsequently, three experienced annotators were invited to join together to refine these initial annotations using two co-registered visible-light hyperspectral images with 0.25 m spatial resolution as reference images. Since the spatial resolution of the visible-light data is four times greater than that of the thermal-infrared hyperspectral imagery and also provides detailed spectral signatures, the quality check can be evaluated by visual inspection. Accordingly, this practice substantially mitigates subjectivity in the labeling process. The Suburb dataset comprises 16 land cover types, including water bodies, vegetation, and buildings, and is suitable for land cover classification and change detection. The Urban dataset is divided into 11 land cover types, primarily encompassing roads, buildings, and other artificial structures, with a focus on complex man-made surfaces.

4.2. Comparison experiments

To validate the effectiveness of TEBS, comparative experiments are conducted against other SOTA methods, encompassing different numbers of bands, with classification performance evaluated from four distinct classifiers: CNN, SVM, KNN, and Transformer. The CNN classifier is implemented as a three-layer one-dimensional convolutional network; the SVM classifier utilizes an RBF kernel and the regularization parameter is set to be 5; the KNN classifier is configured with 5 neighbors and uniform weighting; The Transformer processes 3 × 3-pixel patches as input and comprises 3 encoder layers, each containing 8 attention heads.

The classification overall accuracy (OA) of each method is recorded, and the performance of TEBS is comprehensively assessed by analyzing both the mean of OA. The comparison methods include: the deep reinforcement learning for band selection names as DRL (Mou et al., 2022), the GCN and CNN for band selection named as GCN-CNN-BS (Yu et al., 2022), the contrastive learning for band selection named as ContrastBS

Table 1
Comparison of different band selection methods on Suburb dataset based on CNN classifier.

Method / Band	DRL	GCN-CNN–BS	Contrast BS	BS Former	MLRLF MESC	SSANet –BS	GA	SPA	UBS	TEBS
10	83.34	83.93	84.08	84.02	84.29	83.79	80.84	75.50	**84.52**	84.47
20	85.56	86.04	85.95	85.44	85.52	85.64	84.79	76.30	85.74	**86.05**
30	85.54	86.24	86.23	86.03	85.95	86.21	85.62	78.38	86.20	**86.43**
40	86.56	86.44	86.63	86.53	86.48	86.63	86.44	82.92	86.58	**86.71**
50	86.70	86.55	87.00	86.27	86.18	86.65	86.70	82.72	86.68	**87.21**
60	86.72	87.03	87.00	86.89	86.94	87.07	87.15	84.13	87.16	**87.57**
70	87.30	87.04	86.68	87.42	86.94	87.21	87.49	86.82	87.34	**87.53**
80	87.48	87.55	87.38	87.48	86.89	87.59	87.51	87.77	87.31	**87.97**
90	87.34	87.39	87.26	87.62	87.57	87.50	87.68	87.65	87.54	**88.01**
100	87.71	88.02	87.72	87.86	87.21	87.74	87.86	87.85	87.65	**88.36**

The best values are highlighted in bold, and the second-best values are underlined.

Table 2
Comparison of different band selection methods on Suburb dataset based on SVM classifier.

Method / Band	DRL	GCN-CNN–BS	Contrast BS	BS Former	MLRLF MESC	SSANet –BS	GA	SPA	UBS	TEBS
10	83.88	85.07	84.55	85.24	84.70	84.55	83.36	75.17	85.18	**85.51**
20	86.20	86.34	86.46	86.33	86.59	86.40	86.06	77.17	86.70	**86.88**
30	86.96	87.01	87.28	87.28	87.27	87.12	87.26	79.02	87.37	**87.77**
40	**88.07**	88.01	87.80	87.92	87.49	87.81	87.77	82.69	87.84	87.96
50	88.44	88.47	88.26	88.29	87.91	88.20	88.30	84.54	88.53	**88.70**
60	88.63	88.77	88.77	88.68	88.66	88.70	88.58	86.31	88.90	**88.98**
70	89.08	89.19	89.09	88.98	88.67	89.04	89.03	88.13	89.16	**89.37**
80	89.38	89.39	89.42	89.29	88.90	89.39	89.11	88.86	89.32	**89.55**
90	89.67	89.66	89.55	89.63	89.22	89.62	89.47	89.31	89.59	**89.82**
100	89.85	89.89	89.78	89.79	89.25	89.81	89.66	89.59	89.91	**90.12**

The best values are highlighted in bold, and the second-best values are underlined.

Table 3
Comparison of different band selection methods on Suburb dataset based on KNN classifier.

Method / Band	DRL	GCN-CNN–BS	Contrast BS	BS Former	MLRLF MESC	SSANet –BS	GA	SPA	UBS	TEBS
10	85.39	85.85	85.39	85.77	85.82	85.51	84.42	76.21	86.19	**86.30**
20	86.68	86.65	86.90	86.64	87.00	86.80	86.42	78.93	87.06	**87.37**
30	87.06	87.06	87.26	87.22	87.45	87.24	87.39	80.61	87.38	**87.47**
40	87.61	87.63	87.49	87.57	87.40	87.55	87.44	83.09	87.59	**87.65**
50	87.65	87.75	87.59	87.65	87.29	87.53	87.60	84.16	87.65	**88.06**
60	87.53	87.76	87.74	87.72	87.86	87.71	87.62	85.37	87.68	**88.16**
70	87.75	87.90	87.84	87.77	87.74	87.81	87.78	86.76	87.76	**88.12**
80	88.01	87.86	87.81	87.79	87.75	87.95	87.80	87.36	87.88	**88.02**
90	87.97	87.94	87.92	87.87	87.61	87.84	87.80	87.58	87.89	**88.13**
100	88.08	87.91	88.05	87.94	87.87	87.78	87.88	87.63	87.94	**88.12**

The best values are highlighted in bold, and the second-best values are underlined.

Table 4
Comparison of different band selection methods on Suburb dataset based on Transformer classifier.

Method / Band	DRL	GCN-CNN–BS	Contrast BS	BS Former	MLRLF MESC	SSANet –BS	GA	SPA	UBS	TEBS
10	86.03	86.28	86.71	86.24	86.48	86.34	86.22	77.20	86.19	**86.82**
20	88.67	89.50	89.53	86.97	87.27	87.88	83.77	77.57	89.37	**89.64**
30	86.31	86.95	89.23	89.17	88.35	88.76	87.96	77.90	88.81	**89.97**
40	89.19	86.93	86.72	87.02	87.23	89.08	88.52	80.05	89.10	**89.74**
50	89.55	87.04	88.75	88.58	88.91	87.08	85.80	84.82	88.72	**89.92**
60	86.67	88.27	88.83	88.21	87.53	88.02	88.09	85.77	88.10	**90.28**
70	86.81	89.76	88.17	85.44	89.46	89.00	89.52	86.46	89.34	**89.91**
80	89.53	89.38	89.80	87.53	88.28	87.81	88.86	87.58	88.32	**90.24**
90	89.92	89.41	89.28	88.31	87.51	87.91	89.57	86.95	90.22	**90.79**
100	88.05	88.37	89.63	90.45	88.32	87.69	88.27	88.18	89.72	**90.56**

The best values are highlighted in bold, and the second-best values are underlined.

Fig. 5. Classification comparison from different band selection algorithms on the Suburb dataset when selecting 20 bands.

(Li et al., 2023), the transformer-based reconstruction network for band selection named as BSFormer (Liu et al., 2023b), the multi-level representation learning for band selection named as MLRLFMESC (Wang et al., 2024b), the spectral–spatial cross-dimensional attention network for band selection named as SSANet-BS (Cui et al., 2024), the genetic algorithm named as GA (Pałka et al., 2021), the successive projections algorithm named as SPA (Araújo et al., 2001), and the uniform band selection names as UBS. Among these methods, DRL is based on deep reinforcement learning; GCN-CNN-BS, ContrastBS, BSFormer, MLRLFMESC, and SSANet-BS are deep learning-based approaches; GA, SPA, and UBS are traditional methods. Additionally, during the training process, TEBS employs a WC module composed of 6 STBs. The optimizer

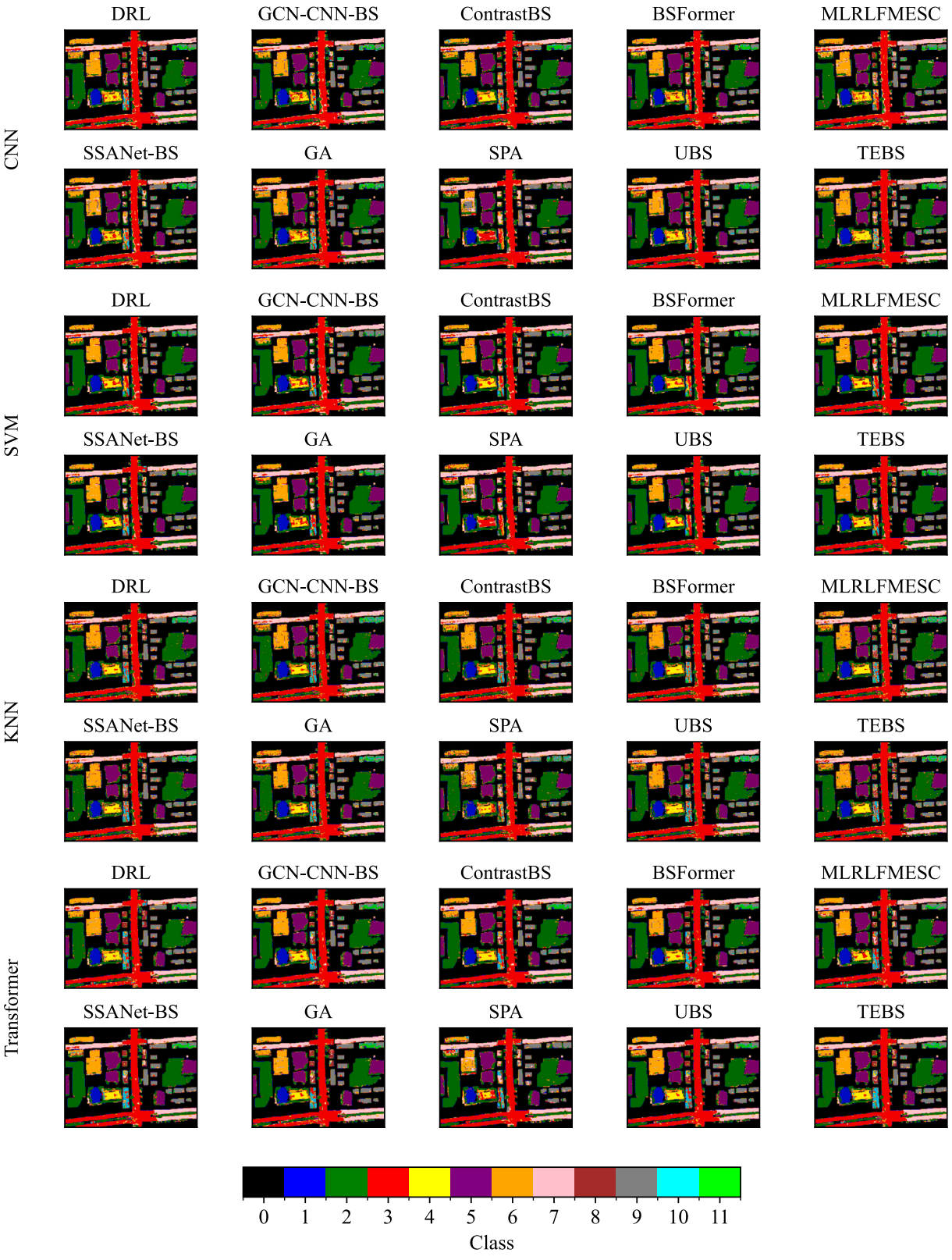


Fig. 6. Classification comparison from different band selection algorithms on the Urban dataset when selecting 20 bands.

employed is Adam, with a learning rate set to 1×10^{-3}, and the learning rate is dynamically reduced through a cosine annealing algorithm.

1) Suburb dataset: This study conducts comparative experiments on the Suburb dataset and performs a quantitative analysis, with the results presented in Tables 1-4 and Fig. 5. The findings indicate that, compared to other SOTA methods, the TEBS achieves the best performance in most of cases. Specifically, when classification is performed by CNN, SVM, KNN, and Transformer methods, using TEBS for band selection results in optimal classification with accuracies of 88.36 %, 90.12 %, 88.16 %, and 90.79 %, respectively. Additionally, it is noteworthy that the

Table 5
Comparison of different band selection methods on Urban dataset based on CNN classifier.

Method / Band	DRL	GCN-CNN–BS	Contrast BS	BS Former	MLRLF MESC	SSANet –BS	GA	SPA	UBS	TEBS
10	77.01	76.16	75.82	76.94	77.53	77.40	76.18	63.32	78.65	**79.15**
20	78.36	78.29	78.41	78.78	79.21	79.59	79.83	71.23	80.27	**80.94**
30	79.61	78.74	79.24	81.28	79.84	80.23	81.96	71.34	82.18	**82.38**
40	80.98	80.88	81.34	81.30	79.78	**81.88**	80.48	74.23	81.27	81.45
50	80.74	79.49	81.05	82.18	80.47	80.65	81.50	74.54	81.64	**82.74**
60	80.86	82.11	82.32	81.65	79.69	81.68	81.72	75.76	81.53	**82.58**
70	80.74	80.84	**83.02**	80.54	81.72	81.67	80.97	78.14	81.24	82.60
80	82.05	81.40	81.73	81.28	82.05	81.40	82.28	80.23	82.27	**83.10**
90	81.88	82.61	81.09	80.41	82.36	82.33	80.26	80.65	79.93	**82.98**
100	83.22	81.66	81.29	81.46	80.13	82.24	81.02	81.08	81.53	**83.33**

The best values are highlighted in bold, and the second-best values are underlined.

Table 6
Comparison of different band selection methods on Urban dataset based on SVM classifier.

Method / Band	DRL	GCN-CNN–BS	Contrast BS	BS Former	MLRLF MESC	SSANet –BS	GA	SPA	UBS	TEBS
10	74.52	73.96	74.32	74.36	74.07	74.93	75.15	66.29	75.93	**76.03**
20	77.51	76.78	77.73	77.42	78.43	78.42	78.54	71.73	78.95	**79.39**
30	79.43	78.85	80.10	79.89	79.90	79.91	80.13	72.62	80.04	**80.19**
40	81.35	81.03	80.89	**81.36**	79.80	81.24	81.02	75.59	81.25	81.14
50	82.09	81.64	81.68	82.06	81.44	82.02	81.90	77.18	82.01	**82.22**
60	81.75	82.58	82.48	82.68	81.85	82.66	82.48	77.94	82.60	**82.75**
70	83.19	83.50	83.07	83.30	82.27	83.33	83.17	80.02	83.14	**83.56**
80	83.65	**84.05**	83.79	83.55	82.56	83.75	83.54	81.89	83.74	83.58
90	84.15	84.06	84.02	83.86	82.99	83.97	83.90	82.71	84.01	**84.27**
100	84.38	84.16	84.49	84.23	83.47	84.36	84.39	83.42	84.46	**84.57**

The best values are highlighted in bold, and the second-best values are underlined.

Table 7
Comparison of different band selection methods on Urban dataset based on KNN classifier.

Method / Band	DRL	GCN-CNN–BS	Contrast BS	BS Former	MLRLF MESC	SSANet –BS	GA	SPA	UBS	TEBS
10	77.33	77.94	77.16	77.86	77.66	77.99	78.04	69.16	79.39	**79.40**
20	79.09	78.67	79.38	78.98	79.90	79.69	79.76	73.56	80.13	**80.22**
30	79.78	79.71	80.33	80.30	80.50	80.06	80.47	74.83	80.63	**80.67**
40	80.43	80.59	80.52	80.52	80.14	80.64	80.38	76.72	80.67	**80.71**
50	80.86	80.74	80.56	80.50	80.62	80.58	80.69	77.29	80.59	**81.23**
60	80.49	**81.01**	80.87	80.80	80.56	80.67	80.53	77.64	80.94	**81.01**
70	80.82	81.11	80.99	80.78	80.69	80.85	80.69	78.82	80.78	**81.26**
80	**81.18**	81.13	81.00	80.85	80.76	81.00	80.89	79.90	80.88	80.92
90	81.13	81.12	81.03	80.98	80.60	80.91	80.96	79.99	80.76	**81.42**
100	81.09	81.12	81.05	81.00	80.86	80.94	80.70	80.42	80.98	**81.46**

The best values are highlighted in bold, and the second-best values are underlined.

Table 8
Comparison of different band selection methods on Urban dataset based on Transformer classifier.

Method / Band	DRL	GCN-CNN–BS	Contrast BS	BS Former	MLRLF MESC	SSANet –BS	GA	SPA	UBS	TEBS
10	83.61	82.59	82.42	82.72	83.06	83.28	83.13	73.93	83.30	**85.02**
20	84.32	83.05	83.32	83.75	82.97	84.21	84.62	75.86	83.76	**86.25**
30	83.75	84.72	84.15	84.24	83.57	85.58	85.12	76.77	86.01	**86.04**
40	85.30	86.21	86.06	85.81	82.61	85.82	85.30	82.20	85.81	**86.62**
50	84.63	85.85	85.04	85.94	85.45	85.55	85.43	81.12	85.57	**87.76**
60	85.67	85.59	85.40	85.66	84.71	86.56	85.03	81.72	86.44	**86.77**
70	86.08	85.80	86.52	85.48	85.15	85.48	85.32	84.16	86.80	**87.82**
80	**87.44**	86.89	86.41	86.86	86.47	86.40	87.06	86.31	86.25	87.34
90	85.32	85.25	86.15	87.24	85.34	86.83	86.23	85.55	87.17	**87.93**
100	86.63	86.14	86.97	86.67	86.16	86.77	86.22	86.60	86.53	**87.99**

The best values are highlighted in bold, and the second-best values are underlined.

Table 9
The runtime (s) of selecting 20 bands by different band selection methods on the Suburb dataset.

Method Band	GCN-CNN-BS	ContrastBS	BSFormer	MLRLFMESC	SSANet-BS	TEBS
20	73.2	41.2	19.6	<u>8.9</u>	3961.7	**8.7**

The best values are highlighted in bold, and the second-best values are underlined.

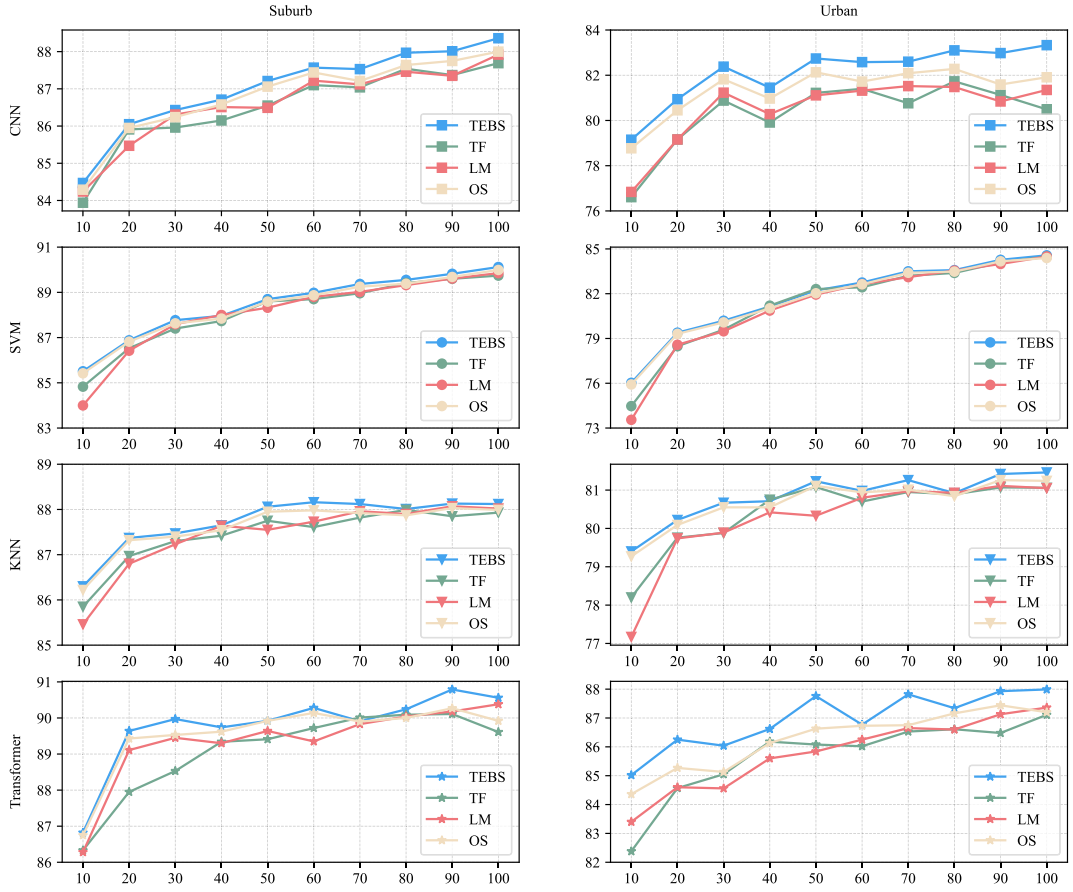


Fig. 7. Classification OA of TEBS and other ablation methods on the Suburb and Urban datasets using four classifiers.

classification accuracy does not exhibit a linear increase with the number of bands utilized. The experimental results indicate that, while increasing the number of bands significantly enhances classification accuracy during the initial stages, this improvement gradually diminishes and eventually stabilizes once the number of bands reaches a certain threshold (e.g., 20 bands). For instance, when employing the CNN classifier, the accuracy of TEBS at 20 bands is 86.05 %, whereas at 100 bands, it only marginally increases to 88.36 %. This indicates that as the number of bands increasing, the classification performance of the algorithm approaches its upper limit, and further increments in band count may not enhance the accuracy and could potentially lead to a decline in performance. This phenomenon is closely associated with the redundancy inherent in hyperspectral data and underscores the critical importance of judicious band selection for improving classification accuracy. Merely increasing the number of bands is not an effective strategy for enhancing classification performance.

2) Urban dataset: To further validate the advantages of the TEBS, comparative experiments are conducted using the Urban dataset, with the results illustrated in Fig. 6 and Tables 5-8. The results indicate that TEBS exhibits superior classification performance to other SOTA methods. In particular, TEBS achieves the best classification accuracies of 83.33 %, 84.57 %, 81.46 %, and 87.99 % for CNN, SVM, KNN, and Transformer classifiers, respectively, which demonstrates high generalizability. Consistent with findings from the Suburb dataset, these experiments indicate that classification accuracy does not exhibit a linear increase relative to the number of bands utilized. In the initial phases of increasing the band count, there is an improvement in classification accuracy; however, upon reaching a certain threshold, this growth begins to decelerate gradually and even displays minor fluctuations. This suggests inherent redundancy in hyperspectral data and shows that indiscriminately increasing more bands does not improve performance.

3) Computation Time: To evaluate the computational performance of the TEBS, a comparative experiment on processing time is proposed. Specifically, the computational time of TEBS is compared with that of other deep learning-based methods when selecting 20 bands from the Suburb dataset. The experiments utilize an NVIDIA RTX 4080 GPU and an Intel Core i9-14900KF CPU. The training time per epoch for each deep model is provided in Table 9. The results indicate that TEBS incurs lower computational overhead, demonstrating its high practical applicability.

4.3. Ablation experiments

To further verify the efficacy of TEBS, ablation experiments focusing on LST information and the model structure are conducted. Specifically,

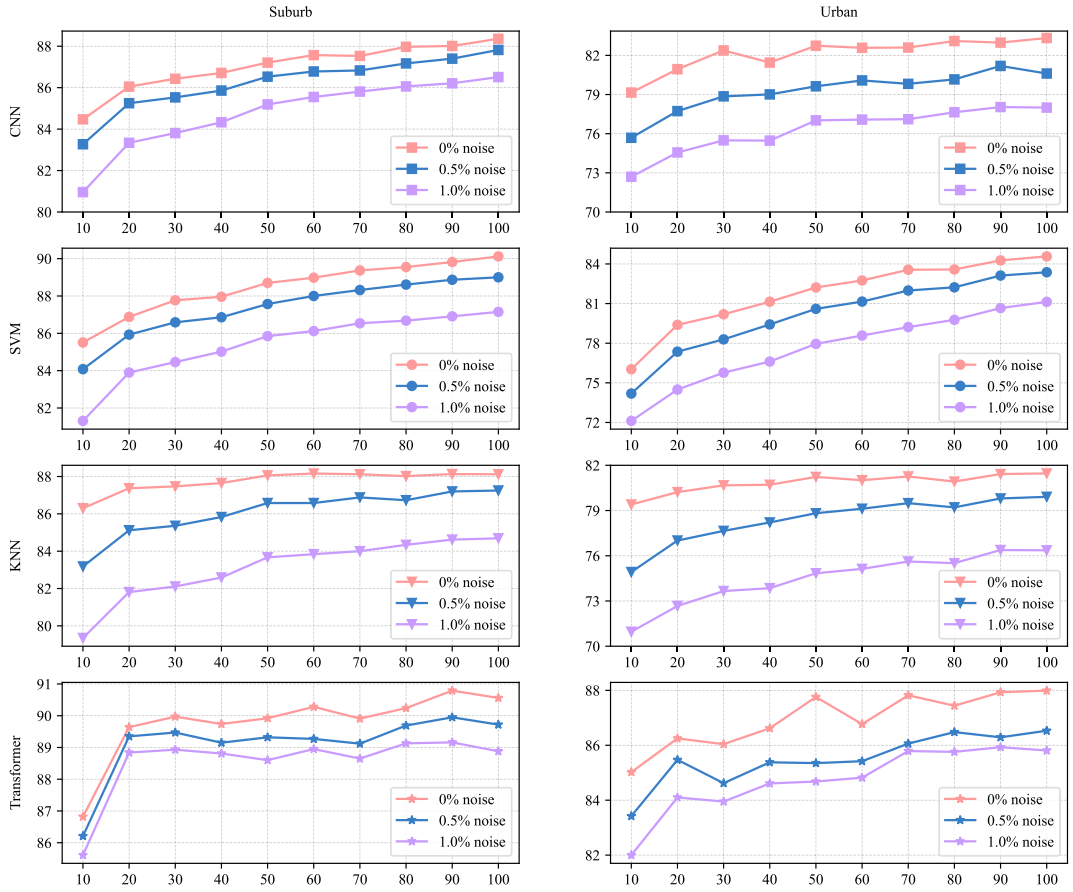


Fig. 8. Classification OA on the Suburb and Urban datasets with varying levels of added noise using four classifiers.

three comparative approaches are designed. Firstly, all LST-related information and corresponding network structures in TEBS are entirely removed, resulting in a simplified model that can be considered as a standard Transformer model, abbreviated as TF. Secondly, while preserving the LST information, the SGA module is eliminated. Alternatively, the LST and LSE information is integrated through a straightforward linear mapping, yielding a model designated as LM. Lastly, the residual connections within the STB module are removed, while only the SGA module is retained, and this variant is named as OS. According to Fig. 7, the classification performance of both the TEBS and OS are superior to that of the TF, thereby validating the beneficial impact of incorporating LST information into the band selection process. The land cover distribution features embedded in the LST images provide models with more comprehensive spatial semantic information, which is crucial for enhancing classification accuracy. Conversely, despite the integration of LST information into the LM, its classification performance does not align with expectations. This indicates that not all methods are capable of effectively extracting the information from LST images; only through the application of suitable feature extraction techniques can surface distribution features be more efficiently captured and utilized. Additionally, a performance comparison between the TEBS and OS reveals that the SGA module, as the core component of the approach, plays a pivotal role in the overall feature extraction process, ultimately enhancing classification performance. Finally, it is noteworthy that when employing CNN or Transformer as the classifier, the advantages of the TEBS become even more significant. This may be attributed to the high data dependency inherent in deep learning methods; compared with SVM and KNN, deep learning approaches rely more heavily on an effective band selection strategy to fully exploit the model's potential.

4.4. Sensitivity analysis experiment

Compared to the visible spectrum, the thermal infrared spectrum generally exhibits lower spectral contrast and signal-to-noise ratio (Zhu et al., 2021). This characteristic amplifies the impact of noise during classification tasks. To investigate the robustness of the bands selected by TEBS against noise interference, a sensitivity analysis experiment is conducted. Specifically, Gaussian noise levels of 0.5 % and 1 % are added to the emissivity data from both datasets, and the classification accuracy of the bands selected by TEBS is evaluated using four classifiers, with the results presented in Fig. 8. The experimental findings indicate that for both datasets, there is a slight decrease in the classification accuracy when the noise level is set at 0.5 %. However, a more pronounced decline in accuracy occurs as the noise level increases to 1 %. Additionally, it is observed that the effect of noise at equivalent level differs between the two datasets. Overall, adding noise has a relatively minor effect on the classification performance for Suburb dataset, while Urban dataset experiences a more significant reduction in classification accuracy. This phenomenon may be attributed to Suburb dataset containing land cover types with higher spectral contrast, whereas Urban dataset comprises land cover types that are inherently similar. Consequently, this similarity leads to further blurring of spectral features due to noise interference, ultimately diminishing classification accuracy. In summary, although the band subset selected by the TEBS demonstrates a certain level of robustness against noise disturbances, its resistance appears relatively weak when classifying similar land cover types.

4.5. Hyperparameter analysis experiment

As hyperparameter settings can substantially affect deep learning

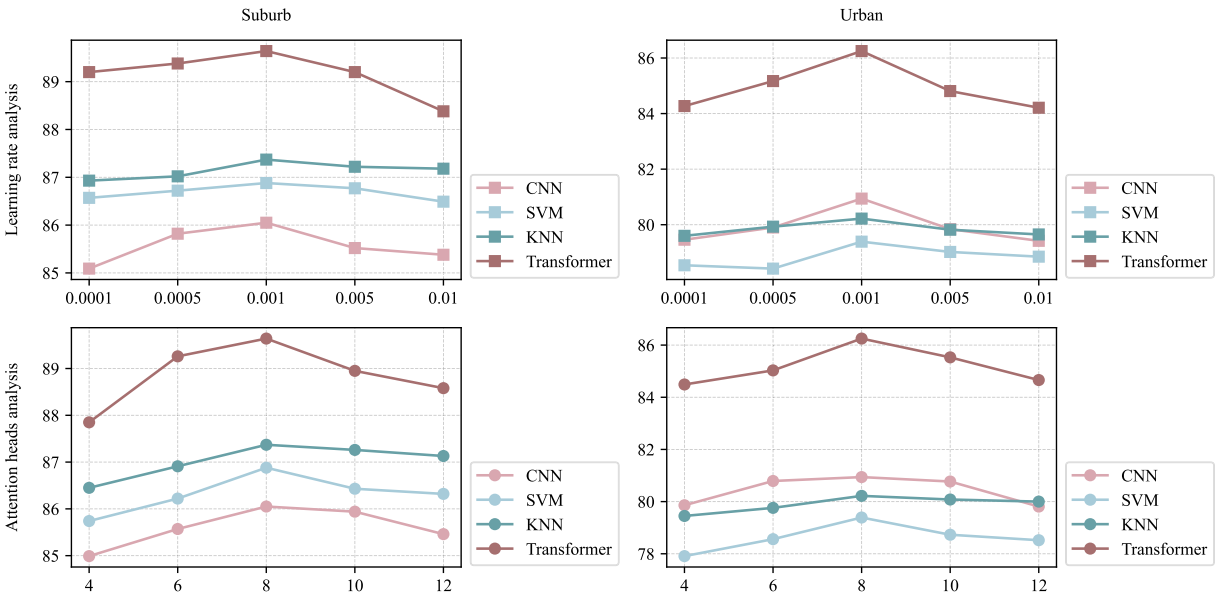


Fig. 9. Classification OA on the Suburb and Urban datasets with different hyperparameter using four classifiers.

models, experiments are conducted to assess the influence of learning rate and the number of attention heads. The results are presented in Fig. 9. For instance, when TEBS is applied to select twenty spectral bands, the learning rate is set to 0.0001, 0.0005, 0.001, 0.005, and 0.01, band selection is carried out, and the overall accuracy of four classifiers (CNN, SVM, KNN, and Transformer) is measured on the Suburb and Urban datasets. The number of attention heads in TEBS's multi-head attention mechanism is also varied (4, 6, 8, 10, and 12), and the OA of the same classifiers on both datasets is again evaluated. Fig. 9 shows that as the learning rate increases, band-selection performance first improves and then declines. A learning rate that is too low leads to sluggish training and under-convergence, whereas a rate that is too high makes the model prone to becoming trapped in local optima and prevents full exploitation of its capacity. Similarly, increasing the number of attention heads enhances TEBS's band-selection effectiveness up to eight heads, beyond which no further appreciable improvements can be observed. This is likely because more attention heads may improve feature extraction, but its excessive number of heads does increase training complexity and may eventually degrade performance.

5. Conclusions and discussions

This paper presents a novel band selection method, named TEBS, designed specifically for TIR-HSI classification. A key innovation of TEBS lies in its ability to simultaneously extract spatial information from LST and spatial-spectral information from LSE. The ablation study demonstrates the superiority of this multimodal feature extraction approach over state-of-the-art methods that rely exclusively on LSE features. Within the proposed TEBS framework, a structured module (named STB) is designed, integrating SSM and GAM, wherein the weight information for each band is explicitly calculated in terms of states. Comparative experiments performed on two datasets, utilizing four classical classifiers, demonstrate that TEBS significantly reduces spectral redundancy while preserving classification accuracy. Furthermore, the proposed TEBS outperforms the state-of-the-art methods, highlighting its effectiveness and robustness. Despite the promising results and conclusions, several issues remain that warrant further exploration. For instance, due to labeling limitations in the datasets, this paper does not investigate the performance of TEBS in other applications, such as land surface temperature retrieval. In addition, the interpretability of deep learning models is also worth further exploration. Addressing these limitations will be an important focus of our future work to further improve the generalizability and versatility of the TEBS algorithm.

Funding

This work is supported in part by National Nature Science Foundation of China (42271355).

CRediT authorship contribution statement

Enyu Zhao: Writing – review & editing, Funding acquisition, Methodology. **Nianxin Qu:** Validation, Writing – original draft, Methodology. **Yulei Wang:** Writing – review & editing, Validation. **Caixia Gao:** Validation, Writing – review & editing. **Jian Zeng:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: [Enyu ZHAO reports financial support was provided by National Nature Science Foundation of China. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.].

Data availability

Data will be made available on request.

References

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S., 2012. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. IEEE Trans. Geosci. Remote Sens. 34 (11), 2274–2282. https://doi.org/10.1109/TPAMI.2012.120.

Araújo, M.C.U., Saldanha, T.C.B., Galvão, R.K.H., Yoneyama, T., Chame, H.C., Visani, V., 2001. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. Chemometric Intell. Lab. Syst. 57 (2), 65–73. https://doi.org/10.1016/S0169-7439(01)00119-8.

Aslett, Z., Taranik, J., Riley, D., 2018. Mapping rock forming minerals at Boundary Canyon, Death Valley National Park, California, using aerial SEBASS thermal infrared hyperspectral image data. Int. J. Appl. Earth Obs. Geoinf. 64, 326–339. https://doi.org/10.1016/j.jag.2017.08.001.

Aved, A., Blasch, E., Peng, J., 2017. Regularized Difference Criterion for Computing Discriminants for Dimensionality Reduction. IEEE Trans. Aerosp. Electron. Syst. 53 (5), 2372–2384. https://doi.org/10.1109/TAES.2017.2696236.

Baldridge, A., Hook, S., Grove, C., Rivera, G., 2009. The ASTER spectral library version 2.0. Remote Sens. Environ. 113 (4), 711–715. https://doi.org/10.1016/j.rse.2008.11.007.

Batchuluun, G., Nam, S., Park, K., 2022. "Deep learning-based plant classification and crop disease classification by thermal camera," *J. King Saud Univ.-Comput.* Inf. Sci. 34 (10), 10474–10486. https://doi.org/10.1016/j.jksuci.2022.11.003.

Chang, C.-I., Du, Q., Sun, T.-L., Althouse, M., 1999. A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 37 (6), 2631–2641. https://doi.org/10.1109/36.803411.

Cui, C., Sun, X., Fu, B., Shang, X., 2024. SSANet-BS: Spectral–Spatial Cross-Dimensional attention Network for Hyperspectral Band selection. Remote Sens. 16 (15), 1–20. https://doi.org/10.3390/rs16152848.

Cui, S., Ma, A., Wan, Y., Zhong, Y., Luo, B., Xu, M., 2022. Cross-Modality image Matching Network with Modality-Invariant Feature Representation for Airborne-Ground thermal Infrared and Visible Datasets. IEEE Trans. Geosci. Remote Sens. 60, 1–14. https://doi.org/10.1109/TGRS.2021.3099506.

Dao, T. and Gu, A., 2024. "Transformers are SSMs: Generalized models and efficient algorithms through structured state space duality," in *Proc. Int. Conf. Mach. Learn.*, Vienna, AT, doi.: 10.48550/arXiv.2405.21060.

Dosovitskiy, A. et al., 2021. "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Representations*, Online, doi: 10.48550/arXiv.2010.11929.

Ermida, S., Hulley, G., Goettsche, F., Trigo, I., 2023. "A combined Vegetation Cover and Temperature-Emissivity Separation (V-TES) Method to Estimate Land Surface Emissivity", Aug. IEEE Trans. Geosci. Remote Sens. 61, 1–18.

Feng, J., et al., 2021. Deep Reinforcement Learning for Semisupervised Hyperspectral Band selection. IEEE Trans. Geosci. Remote Sens. 60, 1–19. https://doi.org/10.1109/TGRS.2021.3049372.

Gao, L., Cao, L., Zhong, Y., Jia, Z., 2021. Field-based High-Quality Emissivity Spectra Measurement using a Fourier Transform thermal Infrared Hyperspectral Imager. Remote Sens. 13 (21), 1–20. https://doi.org/10.3390/rs13214453.

Gillespie, A., Rokugawa, S., Matsunaga, T., Cothern, J., Hook, S., Kahle, A., 1998. A temperature and emissivity separation algorithm for Advanced Spaceborne thermal Emission and Reflection Radiometer (ASTER) images. IEEE Trans. Geosci. Remote Sens. 36 (4), 1113–1126. https://doi.org/10.1109/36.700995.

Goel, A., Majumdar, A., 2022. K-Means embedded Deep Transform Learning for Hyperspectral Band selection. IEEE Geosci. Remote Sens. Lett. 19, 1–5. https://doi.org/10.1109/LGRS.2022.3165313.

Gu, A., Goel, K., and Ré, C., 2022. "Efficiently modeling long sequences with structured state spaces," in *Proc. Int. Conf. Learn. Representations*, Online, doi: 10.48550/arXiv.2111.00396.

He, L., et al., 2023. Non-symmetric responses of leaf onset date to natural warming and cooling in northern ecosystems. PNAS Nexus 2 (9), 1–9. https://doi.org/10.1093/pnasnexus/pgad308.

Johnson, R., Williams, J., Bauer, K., 2013. AutoGAD: an improved ICA-Based Hyperspectral Anomaly Detection Algorithm. IEEE Trans. Geosci. Remote Sens. 51 (6), 3492–3503. https://doi.org/10.1109/TGRS.2012.2222418.

Li, T., Cai, Y., Liu, X., Hu, Q., 2021. "Nonlocal Band attention Network for Hyperspectral image Band selection," *IEEE J.* Sel. Top. Appl. Earth Observ. Remote Sens. 14, 3462–3474. https://doi.org/10.1109/JSTARS.2021.3065687.

Li, X., Liu, Y., Hua, Z., Chen, S., 2023. An Unsupervised Band selection Method via Contrastive Learning for Hyperspectral Images. Remote Sens. 15 (23), 1–17. https://doi.org/10.3390/rs15235495.

Liu, X., et al., 2023a. Local temperature responses to actual land cover changes present significant latitudinal variability and asymmetry. Sci. Bull. 68 (22), 2849–2861. https://doi.org/10.1016/j.scib.2023.09.046.

Liu, Y., Li, X., Xu, Z., Hua, Z., 2023b. BSFormer: Transformer-based Reconstruction Network for Hyperspectral Band selection. IEEE Geosci. Remote Sens. Lett. 20, 1–5. https://doi.org/10.1109/LGRS.2023.3297746.

Long, Y., Wang, X., Xu, M., Zhang, S., Jiang, S., Jia, S., 2023. Dual self-attention Swin Transformer for hyperspectral image super-resolution. IEEE Trans. Geosci. Remote Sens. 61, 1–12. https://doi.org/10.1109/TGRS.2023.3275146.

Manolakis, D., et al., 2019. Longwave Infrared Hyperspectral Imaging: Principles, Progress, and challenges. IEEE Geosci. Remote Sens. Mag. 7 (2), 72–100. https://doi.org/10.1109/MGRS.2018.2889610.

Mou, L., Saha, S., Hua, Y., Bovolo, F., Bruzzone, L., Zhu, X., 2022. Deep Reinforcement Learning for Band selection in Hyperspectral image Classification. IEEE Trans. Geosci. Remote Sens. 60, 1–14. https://doi.org/10.1109/TGRS.2021.3067096.

Neinavaz, E., Darvishzadeh, R., Skidmore, A., Groen, T., 2016. Measuring the response of canopy emissivity spectra to leaf area index variation using thermal hyperspectral data. Int. J. Appl. Earth Obs. Geoinf. 53, 40–47. https://doi.org/10.1016/j.jag.2016.08.002.

Ni, K., Wu, Q., Li, S., Zheng, Z., Wang, P., 2024. Remote Sensing Scene Classification via Second-Order Differentiable Token Transformer Network. IEEE Trans. Geosci. Remote Sens. 62, 1–15. https://doi.org/10.1109/TGRS.2024.3407879.

Ou, X., Wu, M., Tu, B., Zhang, G., Li, W., 2023. Multi-Objective Unsupervised Band selection Method for Hyperspectral Images Classification. IEEE Trans. Image Process. 32, 1952–1965. https://doi.org/10.1109/TIP.2023.3258739.

Pałka, F., Książek, W., Pławiak, P., Romaszewski, M., Książek, K., 2021. Hyperspectral Classification of Blood-like Substances using Machine Learning Methods combined with Genetic Algorithms in Transductive and Inductive Scenarios. Sensors 21 (7), 1–18. https://doi.org/10.3390/s21072293.

Ren, H., Ye, X., Liu, R., Dong, J., Qin, Q., 2018. Improving Land Surface Temperature and Emissivity Retrieval from the Chinese Gaofen-5 Satellite using a Hybrid Algorithm. IEEE Trans. Geosci. Remote Sens. 56 (2), 1080–1090. https://doi.org/10.1109/TGRS.2017.2758804.

Rock, G., Gerhards, M., Schlerf, M., Hecker, C., Udelhoven, T., 2016. Plant species discrimination using emissive thermal infrared imaging spectroscopy. Int. J. Appl. Earth Obs. Geoinf. 53, 16–26. https://doi.org/10.1016/j.jag.2016.08.005.

Sui, C., Li, C., Feng, J., Mei, X., 2020. Unsupervised Manifold-Preserving and Weakly Redundant Band selection Method for Hyperspectral Imagery. IEEE Trans. Geosci. Remote Sens. 58 (2), 1156–1170. https://doi.org/10.1109/TGRS.2019.2944189.

Sun, H., et al., 2022a. Adaptive Distance-based Band Hierarchy (ADBH) for Effective Hyperspectral Band selection. IEEE t. Cybern. 52 (1), 215–227. https://doi.org/10.1109/TCYB.2020.2977750.

Sun, H., Zhang, L., Ren, J., Huang, H., 2022b. Novel hyperbolic clustering-based band hierarchy (HCBH) for effective unsupervised band selection of hyperspectral images. Pattern Recognit. 130, 1–12. https://doi.org/10.1016/j.patcog.2022.108788.

Taspinar, Y., 2023. "Light weight convolutional neural network and low-dimensional images transformation approach for classification of thermal images," *Case Stud.* Therm. Eng. 41, 1–16. https://doi.org/10.1016/j.csite.2022.102670.

Uddin, M., Mamun, M., Afjal, M., Hossain, M., 2021. Information-theoretic feature selection with segmentation-based folded principal component analysis (PCA) for hyperspectral image classification. Int. J. Remote Sens. 42 (1), 286–321. https://doi.org/10.1080/01431161.2020.1807650.

Vaswani, A. et al., 2017. "Attention is All You Need," in *Proc. Adv. Neural Inf. Process. Syst.*, Long Beach, CA, USA, doi: 10.48550/arXiv.1706.03762.

Wan, Y., Chen, C., Ma, A., Zhang, L., Gong, X., Zhong, Y., 2023. Adaptive Multistrategy Particle Swarm Optimization for Hyperspectral Remote Sensing image Band selection. IEEE Trans. Geosci. Remote Sens. 61, 1–15. https://doi.org/10.1109/TGRS.2023.3305545.

Wang, F., Ji, J., Wang, Y., 2023. "DSViT: Dynamically Scalable Vision Transformer for Remote Sensing image Segmentation and Classification," *IEEE J.* Sel. Top. Appl. Earth Observ. Remote Sens. 16, 5441–5452. https://doi.org/10.1109/JSTARS.2023.3285259.

Wang, Q., Zhang, F., Li, X., 2020. Hyperspectral Band selection via Optimal Neighborhood Reconstruction. IEEE Trans. Geosci. Remote Sens. 58 (12), 8465–8476. https://doi.org/10.1109/TGRS.2020.2987955.

Wang, Y., Chen, X., Zhao, E., Zhao, C., Song, M., Yu, C., 2024a. "An Unsupervised Momentum Contrastive Learning based Transformer Network for Hyperspectral Target Detection," *IEEE J.* Sel. Top. Appl. Earth Observ. Remote Sens. 17, 9053–9068. https://doi.org/10.1109/JSTARS.2024.3387985.

Wang, Y., Ma, H., Yang, Y., Zhao, E., Song, M., Yu, C., 2024b. Self-Supervised Deep Multi-Level Representation Learning Fusion-based Maximum Entropy Subspace Clustering for Hyperspectral Band selection. Remote Sens. 16 (2), 1–18. https://doi.org/10.3390/rs16020224.

Wang, Y., Zhu, Q., Ma, H., Yu, H., 2022. A Hybrid Gray Wolf Optimizer for Hyperspectral image Band selection. IEEE Trans. Geosci. Remote Sens. 60, 1–13. https://doi.org/10.1109/TGRS.2022.3167888.

Weng, Q., 2009. Thermal infrared remote sensing for urban climate and environmental studies: Methods, applications, and trends. ISPRS-J. Photogramm. Remote Sens. 64 (4), 335–344. https://doi.org/10.1016/j.isprsjprs.2009.03.007.

Xu, B., Li, X., Hou, Y.W., Wei, Y., 2021. A Similarity-based Ranking Method for Hyperspectral Band selection. IEEE Trans. Geosci. Remote Sens. 59 (11), 9585–9599. https://doi.org/10.1109/TGRS.2020.3048138.

Yang, Y., Wang, Y., Wang, H., Zhang, L., Zhao, E., Song, M., 2024. "Spectral-Enhanced Sparse Transformer Network for Hyperspectral Super-Resolution Reconstruction," *IEEE J.* Sel. Top. Appl. Earth Observ. Remote Sens. 17, 17278–17291. https://doi.org/10.1109/JSTARS.2024.3457814.

Yu, C., Zhou, S., Song, M., Gong, B., Zhao, E., Chang, C.-I., 2022. Unsupervised Hyperspectral Band selection via Hybrid Graph Convolutional Network. IEEE Trans. Geosci. Remote Sens. 60, 1–15. https://doi.org/10.1109/TGRS.2022.3179513.

Zhan, Y., Hu, D., Xing, H., Yu, X., 2017. Hyperspectral Band selection based on Deep Convolutional Neural Network and Distance Density. IEEE Geosci. Remote Sens. Lett. 14 (12), 2365–2369. https://doi.org/10.1109/LGRS.2017.2765339.

Zhang, Q., Geng, G., Zhou, P., Liu, Q., Wang, Y., Li, K., 2024. Link Aggregation for Skip Connection–Mamba: Remote Sensing image Segmentation Network based on link Aggregation Mamba. Remote Sens. 16 (19), 1–22. https://doi.org/10.3390/rs16193622.

Zhao, E., Qu, N., Wang, Y., Gao, C., 2024a. Spectral Reconstruction from thermal Infrared Multispectral image using Convolutional Neural Network and Transformer Joint Network. Remote Sens. 16 (7), 1–17. https://doi.org/10.3390/rs16071284.

Zhao, E., Qu, N., Wang, Y., Gao, C., Duan, S.-B., Zeng, J., 2024b. Thermal Infrared Hyperspectral Band selection via Graph Neural Network for Land Surface Temperature Retrieval. IEEE Trans. Geosci. Remote Sens. 62, 1–14. https://doi.org/10.1109/TGRS.2024.3391008.

Zhao, S., Chen, H., Zhang, X., Xiao, P., Bai, L., Ouyang, W., 2024c. RS-Mamba for Large Remote Sensing image Dense Prediction. IEEE Trans. Geosci. Remote Sens. 62, 1–14. https://doi.org/10.1109/TGRS.2024.3425540.

Zhou, Y., Chen, S., Zhao, J., Yao, R., Xue, Y., Saddik, A., 2022. CLT-Det: Correlation Learning based on Transformer for Detecting Dense Objects in Remote Sensing Images. IEEE Trans. Geosci. Remote Sens. 60, 1–15. https://doi.org/10.1109/TGRS.2022.3204770.

Zhu, X., Cao, L., Wang, S., Gao, L., Zhong, Y., 2021. Anomaly Detection in Airborne Fourier Transform thermal Infrared Spectrometer Images based on Emissivity and a Segmented Low-Rank prior. Remote Sens. 13 (4), 1–20. https://doi.org/10.3390/rs13040754.