

Thermal Infrared Hyperspectral Band Selection via Graph Neural Network for Land Surface Temperature Retrieval

Enyu Zhao¹, Member, IEEE, Nianxin Qu, Yulei Wang², Member, IEEE, Caixia Gao³, Member, IEEE, Si-Bo Duan⁴, Senior Member, IEEE, Jian Zeng, and Qiang Zhang⁵, Member, IEEE

Abstract—Thermal infrared hyperspectral imagery presents a superior capability for capturing intricate spectral details of atmospheres and ground objects compared to multispectral images, thus offering a more nuanced dataset for land surface temperature (LST) retrieval. However, extensive inter-band correlations pose computational challenges and undesirable “dimension disaster” problems. To address this issue, this article proposes a purpose-built framework of thermal infrared hyperspectral band selection (BS) using a graph neural network for LST retrieval. Specifically, the thermal infrared hyperspectral data is first mapped onto a graph topology, followed by feeding it into a graph attention module with brightness temperature constraints to extract band features. Following this, the extracted band features undergo a comprehensive analysis through a multiscale convolution module consisting of convolution kernels with multiple sizes, which have more variety and larger receptive fields for calculating the correlation between different band features, assigning different weights to each band. Finally, a weight selection module is designed to filter the bands based on their assigned weights, creating a subset of bands with greater significance for LST retrieval. Training the designed model, 65 100 observations are simulated utilizing MODTRAN, with 80% allocated for training and 20% for testing. The experimental results validate the effectiveness of the proposed model, with a root mean square error (RMSE) of 1.85 K in practical applications on IASI imagery. This accomplishment substantiates the model’s capacity to reliably employ a judiciously selected subset of thermal infrared hyperspectral bands for LST retrieval applications, thus offering a promising contribution to the advancement of thermal infrared hyperspectral image processing methodologies.

Index Terms—Band selection (BS), graph attention network (GAT), hyperspectral image, land surface temperature (LST), thermal infrared.

I. INTRODUCTION

LAND surface temperature (LST) is one of the most important climate system variables. Because of its ability to characterize longwave radiation and turbulent heat fluxes at the land-atmosphere interface, it has played a crucial role in estimating soil moisture [1], [2], [3], [4], monitoring drought [5], [6], [7], estimating evapotranspiration [8], [9], [10], monitoring the effects of global warming [11], [12], [13], [14], quantifying the urban heat island effect [15], [16], and predicting crop maturity and pests [17], [18]. Advances in aerospace technologies have facilitated the acquisition of LST data on a global scale through thermal infrared remote sensing satellites. Numerous algorithms have been developed to extract LST from thermal infrared remote sensing data [19], delineated into broad categories that include traditional statistical modeling methods and contemporary deep learning-based approaches.

Traditional statistical modeling methods for LST retrieval encompass various approaches, notably single-channel algorithms, multichannel algorithms, and other techniques. The single-channel algorithm, a prominent method, utilizes data from a singular thermal infrared channel, integrating it with surface emissivity and atmospheric water vapor content to retrieve LST. Qin et al. [20] undertook foundational work in this domain, which has been richly extended by other scholars in subsequent studies [21], [22], [23]. To mitigate reliance on atmospheric parameters, McMillin proposed a split-window algorithm, leveraging the differential absorption of atmospheric water vapor between two adjacent thermal infrared channels to ascertain temperature [24], [25], [26], [27]. Over the past four decades, additional algorithms have emerged, including the temperature emissivity separation (TES) algorithm [28], multiangle [29], and multitemporal phase [30] approaches. Furthermore, thermal infrared hyperspectral remote sensing imagery, characterized by its rich surface feature information and near-continuous spectral resolution, provides a new way for LST retrieval, such as the algorithm proposed by Broel [31], which utilizes spectral

Manuscript received 12 January 2024; revised 21 March 2024; accepted 10 April 2024. Date of publication 18 April 2024; date of current version 30 April 2024. This work was supported by the National Nature Science Foundation of China under Grant 42271355 and Grant 61801075. (Corresponding authors: Yulei Wang; Caixia Gao.)

Enyu Zhao, Nianxin Qu, Yulei Wang, and Qiang Zhang are with the Center for Hyperspectral Imaging in Remote Sensing (CHIRS), Information Science and Technology College, Dalian Maritime University, Dalian, Liaoning 116026, China (e-mail: zhaoenyu@dlmu.edu.cn; qvnx@dlmu.edu.cn; wangyulei@dlmu.edu.cn; qzhang95@dlmu.edu.cn).

Caixia Gao is with the Key Laboratory of Quantitative Remote Sensing Information Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (e-mail: gaocaixia@aoe.ac.cn).

Si-Bo Duan is with the Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China (e-mail: duansibo@caas.cn).

Jian Zeng is with China Center for Resources Satellite Data and Application, Beijing 100094, China (e-mail: zengjian@cresda.com).

Digital Object Identifier 10.1109/TGRS.2024.3391008

smoothness as a constraint for LST retrieval. However, these methodologies often require either precise atmospheric correction or exact surface emissivity data, posing significant challenges in practical scenarios.

In recent years, deep learning methods have been widely applied and studied in the field of LST retrieval from hyperspectral thermal infrared data [32], [33], [34], [35]. Deep learning's capacity for resolving nonlinear issues and handling high-dimensional data, owing to its potent learning and adaptive abilities, presents a substantial advancement. Lan et al. [36] proposed a deep mixing model and retrieved LST from infrared atmospheric sounding interferometer (IASI) data. Given that deep learning is inherently data-driven, the performance of these methods is proportional to the spectral resolution of remote sensing images and the richness of the spectral data. While, these algorithms often overlook the criticality of band selection (BS) in hyperspectral remotely sensed data, which typically comprises hundreds to thousands of bands, encapsulating a broad spectrum of electromagnetic wave information. Despite their information richness, hyperspectral images are prone to data redundancy due to their continuity and correlation, potentially leading to increased computational demands and the risk of dimensional catastrophe. Consequently, dimensionality reduction has become a mainstream and necessary preprocessing technique for hyperspectral data [37].

Dimensionality reduction of hyperspectral data, a critical step in data preprocessing, is primarily achieved through two approaches: feature extraction (FE) and BS. FE aims to enhance separability and representativeness by mapping original data into a new feature space. Common methods in this category include principal component analysis (PCA) [38] and independent component analysis (ICA) [39]. Conversely, BS focuses on choosing the most representative bands from the original dataset, aiming to reduce redundancy while preserving the inherent physical properties of the data. This approach is notable for its superior data interpretability compared to FE.

BS methods are broadly classified into supervised, semi-supervised, and unsupervised categories. Supervised methods, requiring data labeling, are esteemed for their impressive outcomes despite the labeling challenges. For instance, Cao et al. [40] introduced a supervised machine-learning algorithm for BS, leveraging the local spatial information of hyperspectral images combined with a packing method. This approach incorporates local spatial smoothing to enhance algorithmic performance. Additionally, Yang et al. [41] developed a dual deep reinforcement learning network (DDQN) for BS, integrating a novel reward mechanism to circumvent local optima. The challenges in labeling samples for supervised algorithms have spurred the development of numerous semi-supervised techniques. A noteworthy example is the model by Feng et al. [42], which employs a semi-supervised deep reinforcement learning framework. This model constructs an evaluation network that reinforces the intraclass compactness of samples, trained through random sampling. Unsupervised algorithms, a significant subset, are further divided into sorting-based and clustering-based methods. Two classic examples include the information scatter-based sorting

algorithm by Chang and Wang [43], along with a clustering algorithm by Martinez-Usó et al. [44] that utilizes a hierarchical clustering structure to optimize intracluster variance and intercluster variance. The field of deep learning has seen groundbreaking advances in BS methodologies for hyperspectral data analysis. A significant contribution in this domain is the dual global-local attention network by He et al. [45]. This network adeptly incorporates spatial and spectral information across both global and local dimensions, enabling a nuanced reweighting of hyperspectral data, which facilitates the identification and selection of the most pertinent bands. Li et al. [46] proposed a nonlocal band attention network, by extracting the remote attention to get the remote relationship between bands, and then realizing stable and effective BS. Moreover, Zhang et al. [47] proposed a dual graph self-representation method that integrates superpixel segmentation and the $l_{2,1}$ -norm to achieve robust unsupervised BS. Such innovations highlight the dynamic evolution of hyperspectral data processing, wherein deep learning techniques are increasingly crucial.

Despite these advancements, a notable gap remains: most existing BS methods predominantly cater to visible-near-infrared hyperspectral data, with limited applicability to thermal infrared hyperspectral data. Addressing this gap, the objective of this article is to advance a graph neural network-based approach for BS in thermal infrared hyperspectral data specifically for LST retrieval. Our proposed method is a supervised algorithm, utilizing LST as a label to select the most representative band subset. It involves mapping hyperspectral data onto a graph structure, followed by a FE module that extracts radiance and brightness temperature features across the spectrum. A subsequent band weighting (BW) module assigns differential weights to each band, culminating in the selection of an optimal subset based on these weights for effective LST retrieval.

This article is organized as follows. Section I introduces the background of the BS study. The proposed model is delineated in Section II. In Section III, the utilized experimental data are described. Section IV conducts tests on the proposed model, presenting experimental results and facilitating a discussion thereof. Section V applies the algorithm to IASI data. The conclusion is shown in the final section.

II. METHODOLOGY

The proposed method consists of four parts: the FE module based on graph neural networks, the BW module based on multiscale convolutions, the BS module based on the weighting factors obtained above, and the temperature regression (TR) module based on feedforward neural network, as shown in Fig. 1. Specifically, the input thermal infrared hyperspectral data is first processed by the FE module, which performs operations on the input thermal infrared hyperspectral data in the spectral domain and extracts the key features. Then, all of the thermal infrared hyperspectral bands are weighted by the BW module, and the weighted bands are selected via the BS module. Finally, the selected thermal infrared hyperspectral bands are used to retrieve LST, and the discrepancy between

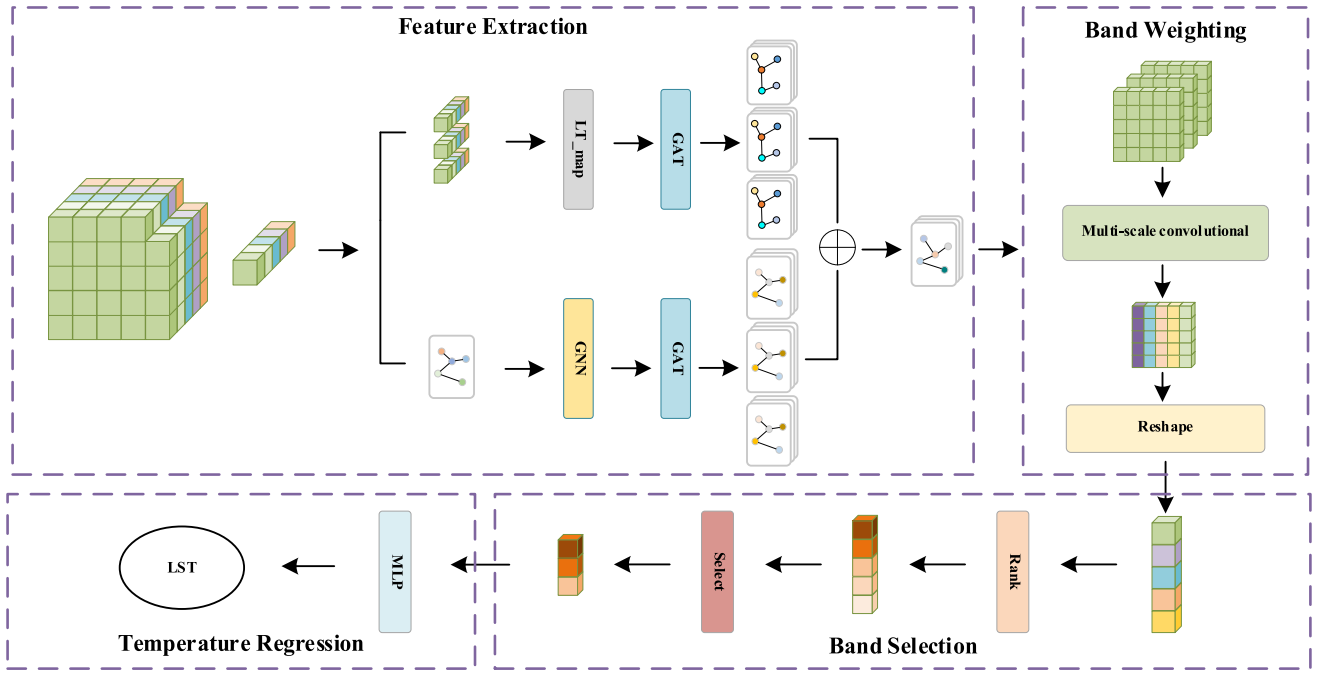


Fig. 1. Diagram of the proposed neural network architecture.

the temperature retrieval results and the actual values is incorporated as part of the loss to train the entire network. On the whole, the FE module is responsible for the extraction of band features, then the BW module weights each input thermal infrared hyperspectral band, the BS module is in charge of selecting the subset of bands based on the weights, and the TR module takes charge of calculating the LST through the selected bands.

A. Graph Data Structure of Hyperspectral Data

The graph is a special type of topological structure composed of nodes and edges, which can be represented as $G = (V, \mathbf{A}, E)$, where V is the set of nodes $V = \{v_1, v_2, \dots, v_n\}$ and E is the set of edges of the graph. In graph data structures, if two nodes satisfy a certain condition, it can be considered that the two nodes are connected by an edge, i.e., there exists an edge between two nodes $v_i \in V, v_j \in V$ only if $(v_i, v_j) \in E$. Additionally, the adjacency matrix \mathbf{A} of the graph is a square matrix of size $|V| \times |V|$, where the value of $A_{i,j}$ indicates whether there is an edge between nodes i and j . It can be expressed as

$$A_{i,j} = \begin{cases} 0, & (v_i, v_j) \notin E, \quad v_i, v_j \in V \\ 1, & (v_i, v_j) \in E, \quad v_i, v_j \in V. \end{cases} \quad (1)$$

In the absence of dependable measured datasets, this study employed simulated thermal infrared hyperspectral data to train the model. This data type encompasses spectral dimension information, characterized by substantial similarity and redundancy across the spectral bands. These traits present notable challenges when transforming hyperspectral data into a graph-based structure.

Graph attention network (GAT), a specialized variant of graph neural networks, demonstrates unique strengths in adaptive feature aggregation, flexible structure modeling, and

efficient information integration. By dynamically weighting each node through attention mechanisms, GAT can concentrate on pivotal features, thus enhancing the detection of interband dependencies and the flow of information. A critical aspect of employing GAT involves determining node connectivity and constructing an adjacency matrix in a judicious manner. However, the process of formulating this matrix encounters several constraints. Primarily, the variations between bands in thermal infrared hyperspectral data are often subtle, complicating the distinction process. Moreover, as the quantity of spectral bands escalates, the computational complexity associated with the adjacency matrix increases quadratically. Notably, the architecture of the adjacency matrix exerts a direct influence on node aggregation within the GAT, subsequently impacting model performance. These factors collectively introduce significant challenges in establishing an appropriate threshold for adjacency in this research context.

In this investigation, a novel approach is devised to transform thermal infrared hyperspectral data into a graph topology, taking into account both the radiance and wavelength attributes of the spectral bands. Conceptually, this method treats each spectral band as a node within the graph structure. The presence of an edge between any two nodes is contingent on a specific relationship between their corresponding bands.

The adjacency matrix, a key component in graph-based analysis, is constructed using a formula that integrates both radiance and wavelength differences between bands. This formula can be expressed as

$$A_{i,j} = \begin{cases} 1, & |v_i - v_j|/v_i < 0.1\% \cap |s_i - s_j| < 0.5 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where $|v_i - v_j|$ represents the radiance difference between bands i and j , and $|s_i - s_j|$ represents the wavelength difference between bands i and j . Specifically, if the radiance relative

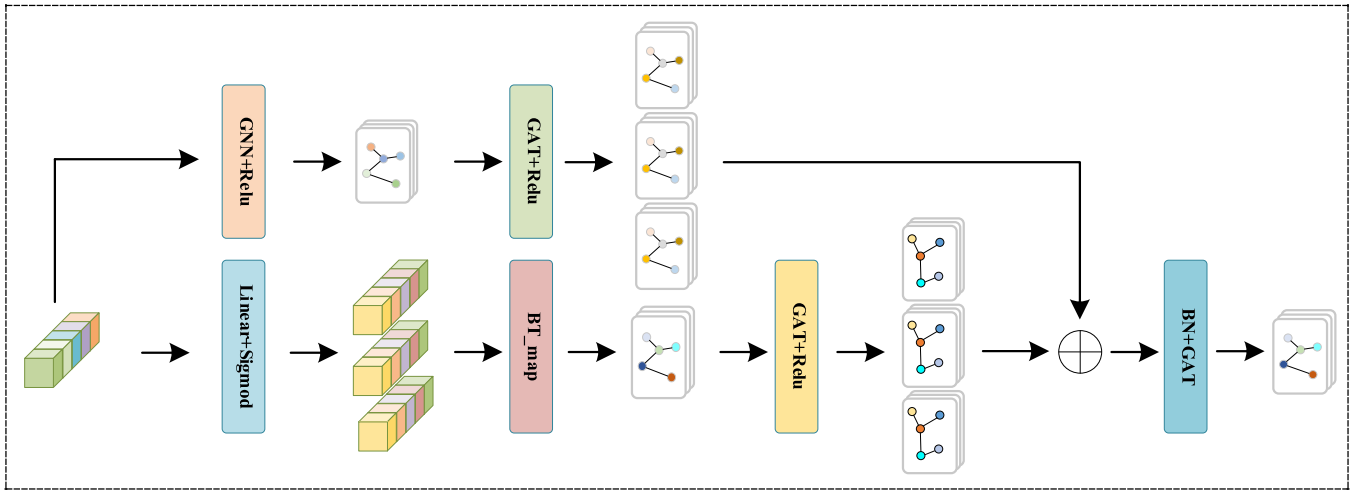


Fig. 2. Structure of the FE module.

difference between two nodes is less than 0.1% and the wavelength difference between the two bands represented by the nodes is less than $0.5 \mu\text{m}$, there is an edge between the corresponding nodes.

This methodology facilitates enhanced aggregation of information across nodes, leveraging the nuances of both radiance and wavelength disparities to form a more informative and representative graph structure of thermal infrared hyperspectral data. The threshold values for radiance and wavelength differences are chosen to ensure meaningful connectivity within the graph, which is instrumental in capturing the intrinsic relationships and characteristics in thermal infrared hyperspectral bands.

B. FE Module

To address the challenge of redundancy among spectral bands in thermal infrared hyperspectral data and to efficiently extract temperature-related features, this section introduces a custom-designed FE module, as illustrated in Fig. 2. This module is architected to process and refine thermal infrared hyperspectral data, ultimately enhancing the extraction of meaningful temperature features.

The FE module initiates its process by receiving a set of thermal infrared hyperspectral data, represented as a tensor of dimensions $(b, 1760)$, where b denotes the batch size and 1760 represents the total spectral bands of the IASI sensor within the wavelength range of 760 to 1200 cm^{-1} . This input undergoes sequential processing through two critical branches: graph attention brightness temperature mapping and radiance graph attention. These branches are interconnected via residual connections, which are instrumental in preserving information and facilitating learning. The output from this module is a feature tensor with dimensions $(b, 3, 1760)$.

Delving into specifics, the FE module executes two pivotal operations on the input radiance features. The first operation employs a multihead graph attention layer. This layer is adept at calculating the similarity between spectral bands, subsequently performing an aggregation of features. This approach allows for an understanding and representation of

the relationships between different spectral bands. The synchronous operation involves mapping the radiance features and converting these mapped multidimensional features into brightness temperature. This conversion leverages the Planck blackbody formula, a fundamental concept in physics describing the relationship between an object's temperature and the radiation it emits. By applying this formula, the module transforms the radiance features into temperature information, representing them in a higher-dimensional space.

This two-stream approach, combining graph-based attention mechanisms with physical principles like the Planck formula, enables the model to not only compute temperature features in a more nuanced, higher-dimensional space but also to extract features that have a direct physical significance. This methodology is devised to ensure that the extracted features are not only relevant for the specific task, but are also grounded in the fundamental physical properties of the data, enhancing the robustness and applicability of the model in practical scenarios.

The multihead graph attention layer can be considered as mapping the input node features into multiple dimensions and then calculating self-attention for each dimension. Features of all nodes within the input are

$$\vec{\mathbf{h}} = \{\vec{h}_1, \vec{h}_2, \dots, \vec{h}_N\}, \quad \vec{h}_i \in \mathbb{R}^F \quad (3)$$

where N is the number of nodes, \vec{h}_i is the feature of the node i , and \mathbb{R}^F signifies an F -dimension real number space.

After the calculation of the graph attention layer, the output features are

$$\vec{\mathbf{h}}' = \{\vec{h}'_1, \vec{h}'_2, \dots, \vec{h}'_N\}, \quad \vec{h}'_i \in \mathbb{R}^{F'}. \quad (4)$$

Specifically, the attention layer begins by calculating the relevance between the node i and its neighboring nodes, then normalizing it according to the significance of different neighboring nodes, and finally weights and sums them to calculate the relevance of the node. The process can be represented as

$$e_{ij} = a(\mathbf{W}\vec{h}_i, \mathbf{W}\vec{h}_j) \quad (5)$$

where, e_{ij} represents the relevance between the node i and its neighboring node j ; \mathbf{W} is the weight matrix, which is a

shared parameter that maps the features to a multidimensional space for a more comprehensive FE; a is a feedforward neural network that maps the high-dimensional features to a real number in the final calculation.

In order to make it easier to compare the attention coefficients of each neighboring node of node i , the attention coefficients need to be normalized, represented as

$$\alpha_{ij} = \text{softmax}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}_i} \exp(e_{ik})} \quad (6)$$

where α_{ij} represents the weight of node j among all the neighbor nodes of node i .

After obtaining the weights of neighbor nodes, the information of node i can be aggregated by means of weighted summation, which can be expressed as

$$\vec{h}'_i = \sigma \left(\sum_{j \in \mathcal{N}_i} \alpha_{ij} \mathbf{W} \vec{h}_j \right). \quad (7)$$

Finally, the self-attention results for multiple dimensions are aggregated using the following formula:

$$\vec{h}'_i(K) = \prod_{k=1}^K \sigma \left(\sum_{j \in \mathcal{N}_i} \alpha_{ij}^k \mathbf{W}^k \vec{h}_j \right) \quad (8)$$

where K represents the number of attention heads in the multihead attention mechanism.

Within the FE module, a distinctive aspect of the processing routine includes the implementation of a specialized attention mechanism, particularly geared toward temperature-related features. This is in addition to the standard multihead graph attention computation. The uniqueness of this approach lies in its focus on the intrinsic physical properties of the hyperspectral data, namely radiance and wavelength, which are inherent attributes of each node in the graph structure.

For each node, the features comprise the radiance and the corresponding wavelength that the node represents. Utilizing these features, the module calculates the brightness temperature of each node. This calculation is grounded in the Planck blackbody radiation formula, a fundamental principle in physics that describes the relationship between the temperature of an object and the spectrum of radiation it emits.

The Planck formula is expressed as follows:

$$M_\lambda(T) = \frac{2\pi hc}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda kT}} - 1} \quad (9)$$

where h represents the Planck constant and has a value of approximately 6.626×10^{-34} (J·s); k represents the Boltzmann's constant and has a value of approximately 1.3806×10^{-23} (J/K); c represents the speed of light and has a value of approximately 2.998×10^8 (m/s); λ represents the wavelength in meters; and T represents the thermodynamic temperature in Kelvin.

By applying the Planck formula, the FE module is capable of transforming the radiance and wavelength data into a meaningful measure of brightness temperature. This process not only enhances the module's ability to focus on temperature-related features but also embeds a layer of

physical significance into the computational framework. The attention constraint, tailored for temperature-related features, ensures that the module effectively processes the most relevant information for temperature estimation, thus enhancing the accuracy and robustness of the FE process in the context of hyperspectral data analysis. In summary, the FE module will calculate the node's radiance and brightness temperature features respectively, thus realizing the complete FE, and the FE module is shown in Fig. 2.

C. BW Module

The process commences with the linear mapping of graph data structure nodes into a 4-D tensor format, yielding a tensor of size $(b, 6, 42, 42)$, where b indicates the batch size. Subsequently, three distinct convolutional operations are applied to this tensor data

- 1) Utilization of a large convolutional kernel, sized 10×10 , to process the data. This results in an output tensor with dimensions $(b, 6, 32, 32)$.
- 2) Implementation of a large-scale dilated convolution, characterized by a kernel size of 4 and a dilation rate of 3. The output from this convolution also maintains the shape $(b, 6, 32, 32)$.
- 3) Execution of four consecutive convolutions using a smaller kernel, sized 4×4 . During these convolutions, the tanh activation function is employed. The final output from this series of convolutions preserves the shape $(b, 6, 32, 32)$.

Post convolutional encoding with these three kernel sizes, the data undergoes a convergence and subsequent redecoding phase. The decoding segment is composed of a 1-D convolution followed by a twofold up-sampling process.

Overall, the BW module facilitates an encoding-decoding process, as shown in Fig. 3. Throughout the encoding phase, the module extracts varying scales of band features. This extraction is crucial, as it enables the module to integrate information from both adjacent and distant bands, ensuring a comprehensive representation of the spectral data. By leveraging diverse convolutional scales and techniques, the BW module plays a pivotal role in enhancing the fidelity and utility of the processed thermal infrared hyperspectral data for subsequent analysis stages.

D. BS Module

Following the comprehensive extraction and analysis of spectral band features by the FE and BW modules, the next critical phase is the BS module. This module's primary function is to select spectral bands based on the weights assigned by the BW module. This selection is pivotal for optimizing the LST retrieval process with a reduced number of bands, thereby enhancing efficiency.

The BW module's output is a matrix \mathbf{w} with dimensions $(b, 1760)$, where 1760 corresponds to the weighted values of all spectral bands, as determined by the encoding and decoding processes of the BW module. The BS module's task is to select a subset of these spectral bands, ensuring that the most relevant and informative bands are utilized for LST retrieval.

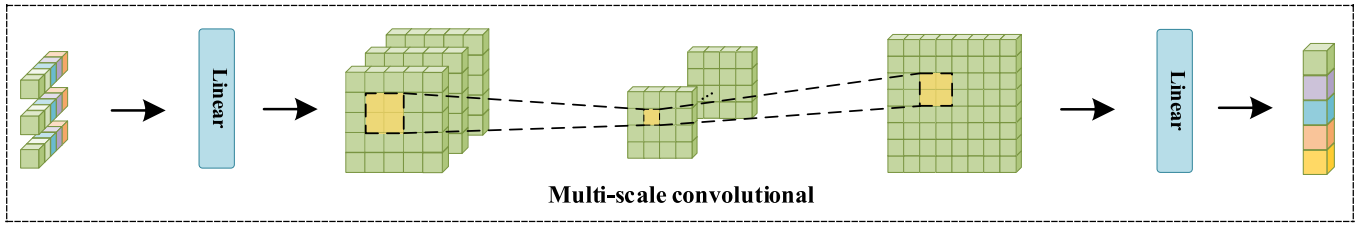


Fig. 3. Structure of the BW module.

The selection process is governed by the following formula:

$$\mathbf{x} = S_n(\mathbf{X} + \mathbf{w} \times \partial) - S_n(\mathbf{w} \times \partial), \quad \mathbf{w} \times \partial \gg \mathbf{X} \quad (10)$$

where \mathbf{X} represents the input raw data, S_n denotes the operation of sorting and selecting the top n weights, and ∂ is a sufficiently large coefficient such that $\mathbf{w} \times \partial \gg \mathbf{X}$. The purpose of this is to maintain the completeness of the computation graph by sorting and selecting the original data \mathbf{X} based on the weights in a basic matrix operation.

Overall, the BS module aims to choose n spectral bands from the entirety of available data, and the final output format of the BS module is (b, n) .

E. TR Module

The TR module in the system serves the function of computing LST based on the radiances of the selected spectral bands. This module employs the calculated LST to generate a loss value by comparing it with the true temperature values. This loss is then utilized to facilitate the backpropagation of gradients within the neural network, which is essential for the iterative refinement and optimization of the model.

Specifically, the TR module consists of a four-layer feedforward neural network, with input data in the format of (b, n) and output in the format of $(b, 1)$.

In addition, the loss function assesses the error between the model's predicted outputs and the actual values. Utilizing the backpropagation algorithm, it updates the model parameters to enhance model performance. To make the selected combination of spectral bands more representative while also targeting the LST retrieval problem, a loss function is designed as follows:

$$\text{Loss} = \text{MSE}(y', y) + 1/\text{Var}(\mathbf{x}) \quad (11)$$

where $\text{MSE}(y', y)$ represents the mean squared error between the output value of the feedforward neural network and the true value, and $\text{Var}(\mathbf{x})$ represents the variance of the input of the feedforward neural network. The purpose of this is to add a constraint on the information entropy to increase the representativeness of the selected spectral bands based on the LST retrieval. During training, the focus is mainly on the value of $\text{MSE}(y', y)$, with $\text{Var}(\mathbf{x})$ being a secondary component of the loss.

In summary, the loss function of the model is designed to minimize the difference between the predicted and actual LST, while also incorporating an information entropy constraint. This approach ensures that the selected spectral bands are highly representative and closely relevant to the LST retrieval

Algorithm 1 BT-GAT Method

Input:

Thermal Infrared Hyperspectral Data: \mathbf{X}

The number of selected bands n

Initialize:

Construct an adjacency matrix \mathbf{A} according to (2) and represent \mathbf{X} as a graph data structure G .

Output:

 Band index

1 While Model not converging

Do

2 The feature vector \vec{h} of each node v is mapped and converted into brightness temperature according to (9).

3 Use (3)-(8) to aggregate the features of each node v to obtain the vector \vec{h}' .

4 Calculate the weight matrix \mathbf{w} of all bands through multi-scale convolution.

5 Select n bands with greater weights according to (10).

6 LST is calculated via a feedforward neural network using a selected subset of bands.

7 Calculate Loss according to (11).

8 Update model parameters.

End

9 Select a subset of bands with greater weight.

problem. This method enhances prediction accuracy and guarantees that the model focuses on the most crucial features for LST prediction, leading to precise and efficient performance. The overall algorithmic workflow is shown in Algorithm.

III. DATA

A. Dataset Creating

On the basis of radiative transfer theory, under atmospheric conditions without clouds, the radiance obtained through the thermal infrared sensor onboard the satellite can be expressed as shown below

$$B(T) = \varepsilon B(T_s)\tau + R_{\text{atm}}^{\uparrow} + (1 - \varepsilon)R_{\text{atm}}^{\downarrow}\tau \quad (12)$$

where B is the Planck function, T is the brightness temperature at the top of the atmosphere observed by the sensor, $B(T_s)$ represents the Planck radiance of an object at a temperature T_s , ε is the surface emissivity, τ represents the atmospheric transmittance, and $R_{\text{atm}}^{\uparrow}$ and $R_{\text{atm}}^{\downarrow}$ represent atmospheric upwelling and downwelling radiation, respectively.

In the context of the atmospheric thermal radiation transfer equation's theoretical framework, the radiance for a given

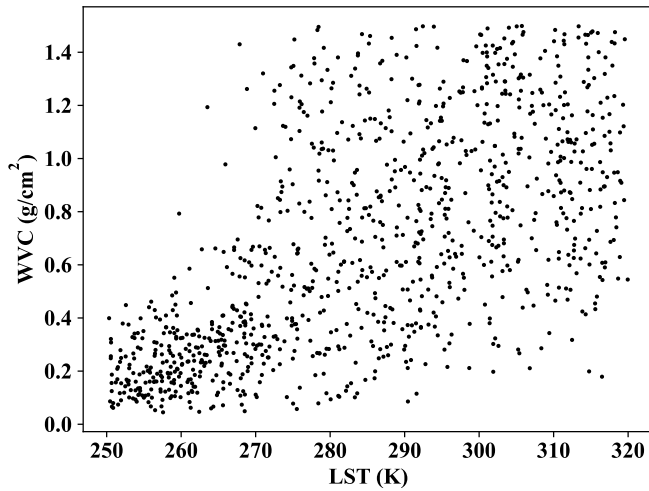


Fig. 4. Relationship between WVC and LST of atmospheric profiles.

spectral band can be deduced through knowledge of several key parameters: LST, surface emissivity, atmospheric transmittance, atmospheric upwelling radiation, and atmospheric downwelling radiation, which can be simulated by MODerate resolution atmospheric TRANsmiSSion (MODTRAN). In this part, a comprehensive dataset comprising 65 100 vertical observation conditions was generated, utilizing 62 representative land cover spectra and 1050 clear-sky atmospheric profiles. Each of these conditions encapsulates radiance information for 1760 spectral bands, covering a wavelength range of 760 to 1200 cm^{-1} . The dataset was subjected to a random shuffle and subsequently partitioned into a training set (80%) and a test set (20%). The subsequent sections will provide more detailed insights into the land cover spectra and atmospheric profiles utilized in this part.

B. ECMWF Reanalysis v5 (ERA5)

ERA5, the fifth-generation atmospheric reanalysis product developed by the European Centre for Medium-Range Weather Forecasts (ECMWF), provides a comprehensive depiction of global climate conditions, spanning from January 1940 to the present. This reanalysis is a product of the Copernicus Climate Change Service (C3S) initiative at ECMWF.

For this part, data were specifically sourced from the Asian region, confined within the geographical bounds of 75°–120°E longitudes and 24°–48°N latitudes, during the year 2021, with a variety of surface types, including urban areas, forests, lakes, deserts, and grasslands. The LST within this region was segmented into seven distinct intervals, ranging from 250 to 320 K. These intervals were established with increments of 10 K. From each temperature interval, 150 atmospheric profiles were randomly selected, culminating in a total of 1050 atmospheric profiles. Fig. 4 illustrates the relationship between WVC and LST of the selected atmospheric profile, where the x -axis represents the LST and the y -axis represents the WVC, respectively.

C. JHU Land Cover Spectral Database

This research utilized spectral data from the JHU Spectral Library, an extensive database established by Johns Hopkins

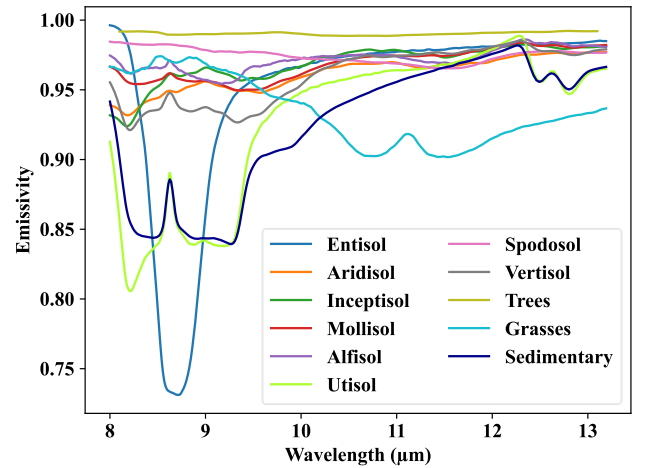


Fig. 5. Emissivity profiles for 11 common feature types.

University in the United States, with a wide array of spectra, including rocks, minerals, vegetation, and soils. A total of 62 characteristic spectra from 11 classical features were selected for analysis, with emissivity curves plotted from randomly chosen samples of these 11 feature classes, as depicted in Fig. 5. This approach allows for a comprehensive understanding of the emissive properties of various objects, which is crucial for accurate remote sensing and environmental analysis.

The emissivity of different feature types can be very similar in one band and very different in another, as shown in Fig. 5, underscoring the criticality of BS, with the goal of discerning and selecting a band subset that capture the most comprehensive emissivity information, thereby maximizing the efficacy and accuracy of spectral analysis.

D. Data Analysis

To illustrate the redundancy of the thermal infrared hyperspectral data, 20 bands were selected from the database, where Pearson coefficients were calculated utilizing average and random sampling methods, as shown in Fig. 6.

It can be seen that there is a high correlation and redundancy between the thermal infrared hyperspectral bands, and increasing the number of bands does not necessarily bring a better result for LST retrieval. The reason for this phenomenon may be that the radiance received by the sensor is coupled by a variety of factors, such as LST, emissivity, atmospheric absorption, etc., and the variations of these factors are not significant or even highly correlated in different thermal infrared channels. Under ideal conditions, the most representative subset of bands should exhibit the lowest possible correlation to reduce the coupling of information, so as to enhance the efficiency of the target parameter of LST, which is also the ultimate goal of BS.

IV. RESULTS AND ANALYSIS

A. Comparison Experiments

To prove the effectiveness of this method, several sets of comparison experiments are utilized for validation. LST retrieval result is influenced by three factors: the number

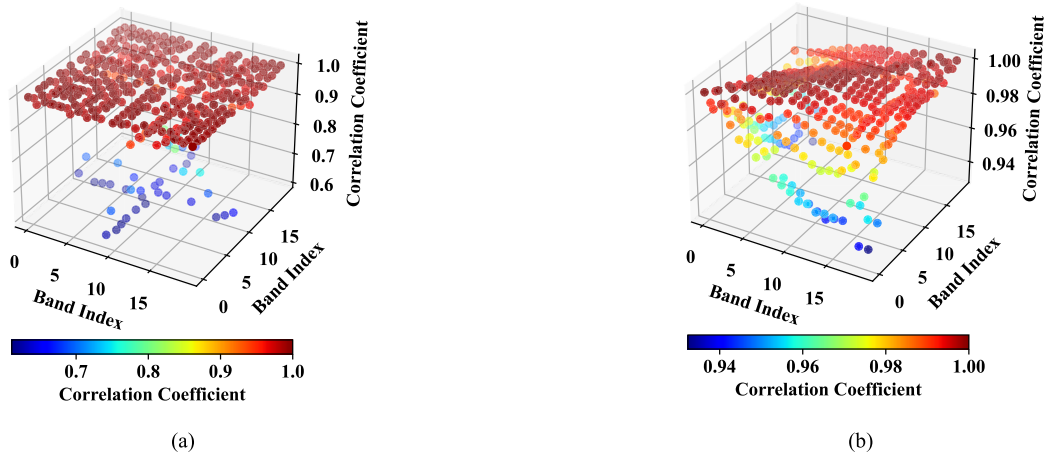


Fig. 6. Schematic of the Pearson coefficients for each band. (a) Correlation coefficients for a random sample of 20 selected bands and (b) correlation coefficients for an average sample of 20 selected bands.

TABLE I
COMPARISON OF RMSE (K) FOR DIFFERENT BS METHODS IN GRU NETWORK-BASED LST ESTIMATION

Band \ Method	UBS	GA	SPA	BS_Net_FC	DARecnet	BT-GAT
10	0.86±0.12	1.62±0.07	1.52±0.19	1.69±0.08	1.31±0.09	0.74±0.04
20	0.98±0.24	1.38±0.09	1.38±0.40	2.35±0.42	1.90±0.36	0.91±0.16
30	0.88±0.03	1.81±0.08	1.86±0.24	2.13±0.08	1.62±0.04	0.89±0.11
40	0.98±0.06	1.64±0.24	1.40±0.16	1.94±0.31	1.80±0.41	0.86±0.15
50	1.05±0.05	1.47±0.18	1.52±0.34	2.11±1.05	1.79±0.12	0.95±0.10
60	1.02±0.11	1.46±0.10	1.69±0.35	1.92±0.47	1.53±0.28	0.83±0.03
70	0.96±0.08	1.48±0.25	1.76±0.14	1.87±0.14	1.50±0.11	0.94±0.15
80	1.05±0.14	1.71±0.20	1.55±0.27	1.71±0.11	1.86±0.09	0.98±0.18
90	1.00±0.17	1.58±0.24	1.95±0.28	1.74±0.12	1.72±0.17	1.00±0.05
100	1.03±0.19	1.68±0.11	1.84±0.14	1.59±0.12	2.11±0.23	0.90±0.05

of the chosen bands, the index of the chosen bands, and the model used for LST retrieval. The three sets of comparative experiments correspond to LST retrieval using three models: GRU, LSTM, and DNN. This eliminates the uncertainty caused by different LST retrieval models. Where GRU and LSTM are branches of recurrent neural networks, with Ye et al. [48] having previously employed LSTM for LST retrieval. The deep neural network (DNN), evolving from multilayer perceptron, represents a relatively simple network structure, and Wang et al. [35] have utilized DNN for temperature retrieval and conducted a detailed analysis thereof. Moreover, we compared the effects of different BS methods and different selected band numbers on LST retrieval in each experiment. The compared BS methods are BS_Net_FC [49], DARecnet [50], Genetic Algorithm (GA) [51], Successive Projection Algorithm (SPA) [52], and Uniform Selection Algorithm (UBS) [53]. Here, BS_Net_FC and DARecnet are unsupervised BS algorithms based on deep learning, whereas GA and SPA are traditional machine learning algorithms.

As demonstrated in Tables I–III, employing the BT-GAT method proposed in this study has an obvious advantage in the case of LST retrieval using the GRU network. When

utilizing the LSTM network for LST retrieval, optimal outcomes are obtained via the implementation of the BT-GAT method for BS, and the RMSE of LST retrieval can reach 0.42 K. When utilizing the DNN network for LST retrieval, the advantage of the BT-GAT method is still large, but not as obvious as in Tables I and II, which may be due to the relatively poor stability of the DNN network. Overall, the method proposed in this study is superior to other methods in most cases. Furthermore, it is noteworthy that utilizing the DNN model for extracting LST results in a noticeable increase in the RMSE of the LST obtained as the number of bands rises. However, when the LSTs are retrieved using the LSTM model, the augmentation of bands exhibits a nominal effect on the retrieval results. It is imperative to note that this observation does not diminish the inherent utility of the BS process. Beyond its impact on computational efficiency, the specifically selected bands significantly impact the results. Consequently, under the condition of selecting an identical number of bands, different BS methods persist in exerting discernible effects on the resultant calculation results. For this result, our conjecture is that for the problem of retrieving LST from thermal infrared hyperspectral data, there is indeed

TABLE II
COMPARISON OF RMSE (K) FOR DIFFERENT BS METHODS IN LSTM NETWORK-BASED LST ESTIMATION

Band \ Method	UBS	GA	SPA	BS_Net_FC	DARecnet	BT-GAT
10	0.50±0.06	0.58±0.09	0.76±0.13	0.82±0.37	0.71±0.06	0.55±0.06
20	0.50±0.16	0.56±0.19	0.62±0.21	0.63±0.31	0.80±0.10	0.43±0.05
30	0.46±0.05	0.43±0.09	0.62±0.10	0.46±0.05	0.70±0.12	0.46±0.04
40	0.47±0.05	0.58±0.14	0.60±0.10	0.47±0.09	0.58±0.16	0.42±0.06
50	0.53±0.13	0.51±0.09	0.58±0.14	0.52±0.10	0.54±0.06	0.47±0.03
60	0.55±0.08	0.56±0.08	0.57±0.14	0.55±0.05	0.71±0.13	0.54±0.18
70	0.55±0.08	0.52±0.10	0.69±0.48	0.63±0.17	0.57±0.13	0.43±0.10
80	0.54±0.14	0.52±0.06	0.53±0.05	0.64±0.32	0.78±0.42	0.53±0.04
90	0.57±0.10	0.52±0.07	0.76±0.51	0.51±0.09	0.52±0.10	0.46±0.07
100	0.69±0.17	0.54±0.12	0.60±0.12	0.63±0.20	0.58±0.18	0.52±0.15

TABLE III
COMPARISON OF RMSE (K) FOR DIFFERENT BS METHODS IN DNN NETWORK-BASED LST ESTIMATION

Band \ Method	UBS	GA	SPA	BS_Net_FC	DARecnet	BT-GAT
10	1.61±0.13	1.59±0.33	1.59±0.41	1.61±0.47	1.88±0.28	1.55±0.39
20	1.74±0.55	1.58±0.59	1.73±0.38	1.78±0.36	1.81±0.25	1.42±0.29
30	1.82±0.30	2.02±0.64	3.20±0.79	2.13±0.60	2.45±0.56	1.75±0.36
40	2.49±0.90	2.14±0.59	2.96±0.66	2.20±0.56	2.20±0.61	2.06±0.44
50	2.39±0.69	2.44±0.46	2.34±0.53	2.94±0.83	3.85±1.43	2.10±0.36
60	2.42±0.48	2.00±0.24	2.79±0.43	2.72±0.58	3.16±0.97	1.98±0.39
70	3.44±1.17	2.56±0.62	2.40±0.34	2.98±1.10	2.60±0.66	2.55±0.61
80	3.09±0.72	2.98±0.69	2.21±0.34	3.35±0.89	3.52±1.04	2.18±0.14
90	3.33±0.44	3.35±0.74	2.76±0.48	3.16±0.89	4.10±1.47	2.45±0.34
100	3.33±1.51	2.88±0.54	3.40±0.56	2.91±0.79	4.00±1.44	2.46±0.20

a Hughes phenomenon. However, different neural networks perform differently in combating problems such as overfitting or gradient explosion, and the Hughes phenomenon is not observed to the same extent. For example, the LSTM neural network, which has memory units and gating mechanisms to better handle sequential data, exhibits more robustness in the problem of LST retrieval compared to the DNN neural network, so the Hughes phenomenon is not notably significant in this scenario.

Furthermore, to more intuitively demonstrate the differences in the selection methods across various spectral bands, Fig. 7 depicts the bands chosen by different algorithms when the number of selected bands is set to 40.

Additionally, the computational efficiency of the BT-GAT model was assessed. The network training was conducted on a system with an Intel Xeon¹ Platinum 8255C CPU and RTX-3090 GPU. During the training phase, the entire model was segmented into three distinct sections. The first section encompasses the FE and BW modules, employing the Adam optimizer with a learning rate of 10^{-6} . The second section comprises the BS module without trainable parameters. The

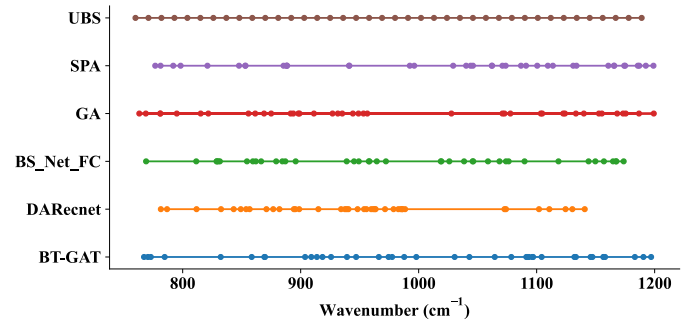


Fig. 7. Band locations obtained by different methods.

third section, incorporating the TR module, also utilizes the Adam optimizer, with a learning rate of 10^{-3} . Notably, the network parameters of both the first and third sections are updated concurrently following the computation of loss. In the training process, with a batch size of 32, computing one epoch took approximately 75 s. The complete training process, spanning 20 epochs, required about 25 min in total.

B. Ablation Experiments

1) *Performance Comparison of Different Ablation Methods:*
To assess the efficacy of the brightness temperature physical

¹Registered trademark.

TABLE IV
DIFFERENT ABLATION EXPERIMENT METHODS COMPARISON OF RMSE (K) IN GRU NETWORK-BASED LST ESTIMATION

Method	BT_Attention	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$
	BigCNN			
Band	10	1.46±0.05	1.60±0.14	0.74±0.04
	20	1.58±0.08	1.74±0.10	0.91±0.16
	30	1.61±0.28	1.74±0.25	0.89±0.11
	40	1.38±0.11	1.48±0.12	0.86±0.15
	50	1.47±0.10	1.43±0.11	0.95±0.10
	60	1.76±0.40	1.80±0.21	0.83±0.03
	70	1.61±0.37	1.58±0.15	0.94±0.15
	80	1.45±0.13	1.72±0.19	0.98±0.18
	90	1.73±0.17	1.29±0.06	1.00±0.05
	100	1.54±0.14	1.68±0.31	0.90±0.05

TABLE V
DIFFERENT ABLATION EXPERIMENT METHODS COMPARISON OF RMSE (K) IN LSTM NETWORK-BASED LST ESTIMATION

Method	BT_Attention	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$
	BigCNN			
Band	10	0.62±0.17	0.65±0.06	0.55±0.06
	20	0.50±0.09	0.56±0.11	0.43±0.05
	30	0.48±0.04	0.53±0.10	0.46±0.04
	40	0.48±0.11	0.48±0.13	0.42±0.06
	50	0.50±0.10	0.58±0.20	0.47±0.03
	60	0.58±0.21	0.62±0.18	0.54±0.18
	70	0.55±0.12	0.58±0.11	0.43±0.10
	80	0.59±0.08	0.66±0.15	0.53±0.04
	90	0.58±0.08	0.56±0.10	0.46±0.07
	100	0.64±0.12	0.53±0.07	0.52±0.15

constraints and the multiscale convolution with large kernels in BS proposed in this study, ablation experiments were conducted, as illustrated in Tables IV and V. In this experiment, the following methods were tested: using a standard GAT instead of brightness temperature-constrained GAT, substituting standard convolution kernels for multiscale convolution with large kernels, and the BT-GAT algorithm proposed in this study.

In the ablation study, the BT-GAT algorithm achieved the best results, as shown in Tables IV and V. Specifically, during the ablation process under the physical constraint of brightness temperature, all experiments exhibited varying degrees of degradation. This phenomenon validates the efficacy of the brightness temperature physical constraint, which can better represent the relevant information of LST, thereby enhancing algorithm performance. Furthermore, the experiments indicated that replacing multiscale convolution with large kernels by ordinary convolution led to an increase in the RMSE of LST retrieval, suggesting that incorporating large-scale convolution

is beneficial for the model performance. Its unique large-scale receptive field can aggregate band information from greater distances.

Overall, the proposed method BT-GAT outperforms its ablated versions; each component of the model plays a crucial role in enhancing the model's ability to accurately retrieve LST. The results affirm the benefit of integrating domain-specific knowledge through physical constraints.

2) *Performance Comparison of Different Graph Neural Networks*: An experiment was designed to demonstrate that the performance of using GAT surpasses that of other graph neural networks. In the experiment, all GAT layers in the BT-GAT algorithm were individually replaced with GNN layers and GCN layers for BS. The LST retrieval results using the selected bands are presented in Fig. 8.

In this experiment, the performance of GAT was found to be almost comprehensively superior to that of GNN and GCN. The reason for this outcome may be due to the ability of GAT to assign different weights to each node through the attention

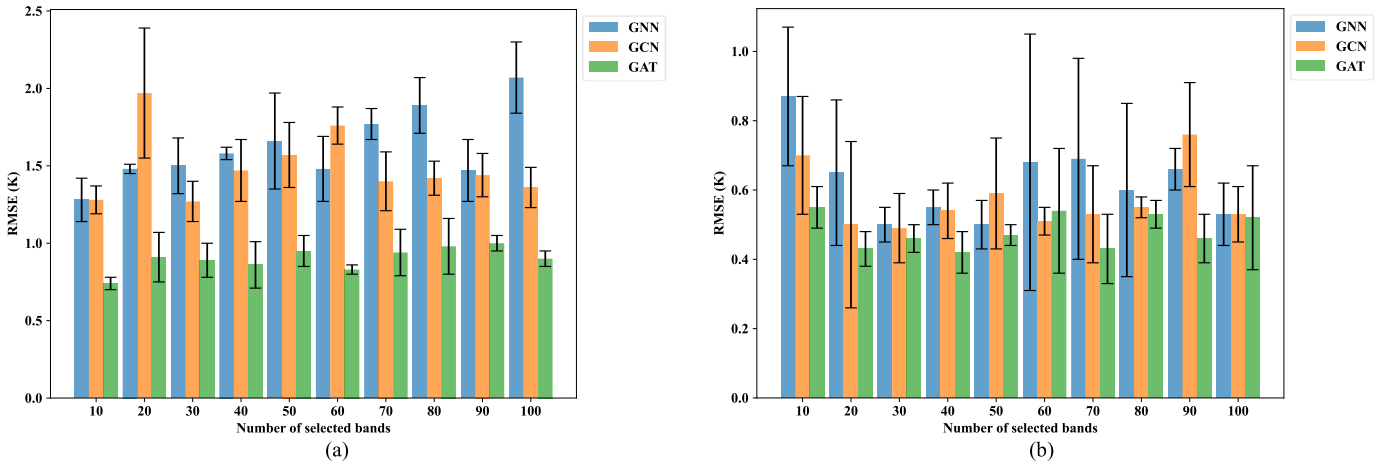


Fig. 8. Comparison of different graph neural networks for BS. (a) RMSE (K) for LST retrieval using GRU network. (b) RMSE (K) for LST retrieval using LSTM network.

TABLE VI
COMPARISON OF RMSE (K) FOR DIFFERENT BS METHODS IN PRACTICAL APPLICATIONS

Method	UBS	GA	SPA	BS_Net_FC	DARecnet	BT-GAT
Error						
RMSE (K)	2.27	1.99	2.07	2.18	2.20	1.85

mechanism. This adaptive weight allocation makes GAT more flexible and effective in capturing important relationships between nodes, which, for the BS, means that GAT can better understand and extract band information most relevant to the specific task.

C. Sensitivity Analysis

Instrument noise, arising from factors like body vibration, calibration error, and dark current during sensor operation, can introduce minor inaccuracies in sensor imaging [54]. To evaluate the effects of the instrumental noise on the BS method proposed in our study, a Gaussian noise with a standard deviation of $0.02 \text{ W} \cdot \text{m}^{-2} \cdot \mu\text{m}^{-1} \cdot \text{sr}^{-1}$ was introduced to all the data. This noise-altered data was then used for BS, and subsequently, for obtaining the LST using a subset of the selected bands. The results of these experiments are presented in Fig. 9.

As observed in Fig. 9, incorporating Gaussian noise at a level of $0.02 \text{ W} \cdot \text{m}^{-2} \cdot \mu\text{m}^{-1} \cdot \text{sr}^{-1}$ leads to a marginal increase in the RMSE of LST retrieval when using the GRU or DNN algorithms. However, the RMSE of LST retrieval remains relatively stable with the LSTM algorithm. This suggests that while noisy data can impact the stability of LST retrieval results, the LSTM algorithm, in particular, demonstrates resilience and continues to perform effectively. In light of these findings, the LSTM algorithm will be the preferred choice for LST retrieval in subsequent experiments based on the demonstrated robustness in handling noisy data, the crucial attribute for accurate LST retrieval.

V. APPLICATION

To enhance the model's practical applicability, IASI thermal infrared hyperspectral data is utilized for LST retrieving and

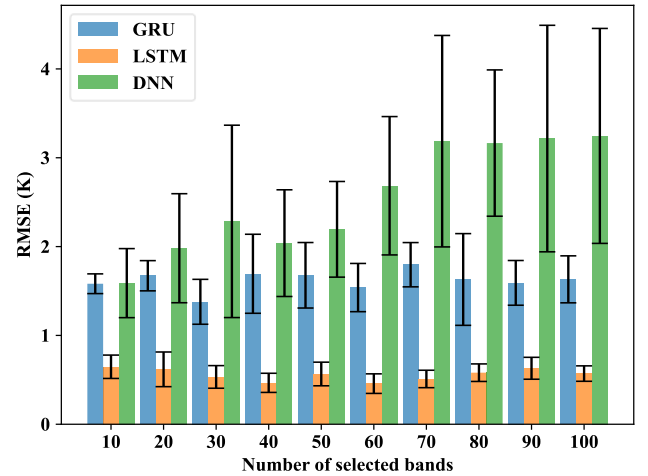


Fig. 9. RMSE for LST retrieval using additive noise data.

validating. The experiment utilized the IASI L1C radiance and IASI L02 temperature products specifically focusing on data collected in January 2022 from image elements observed at zenith angles less than 20° . The chosen experimental region covered an area spanning from 24° to 48°N and 75° to 120°E in Southeast Asia. Employing the BT-GAT and other BS models, 40 bands were selected to execute LST retrieval using the LSTM network. The retrieved LST values were then compared with the IASI L02 product.

To assess the influence of different BS methods on LST retrieval, various BS strategies were applied in practical settings, with the findings detailed in Table VI. The implementation of BT-GAT for data dimensionality reduction followed by LST retrieval resulted in a best RMSE of 1.85 K, whereas alternative approaches yielded RMSEs ranging from 1.99 to

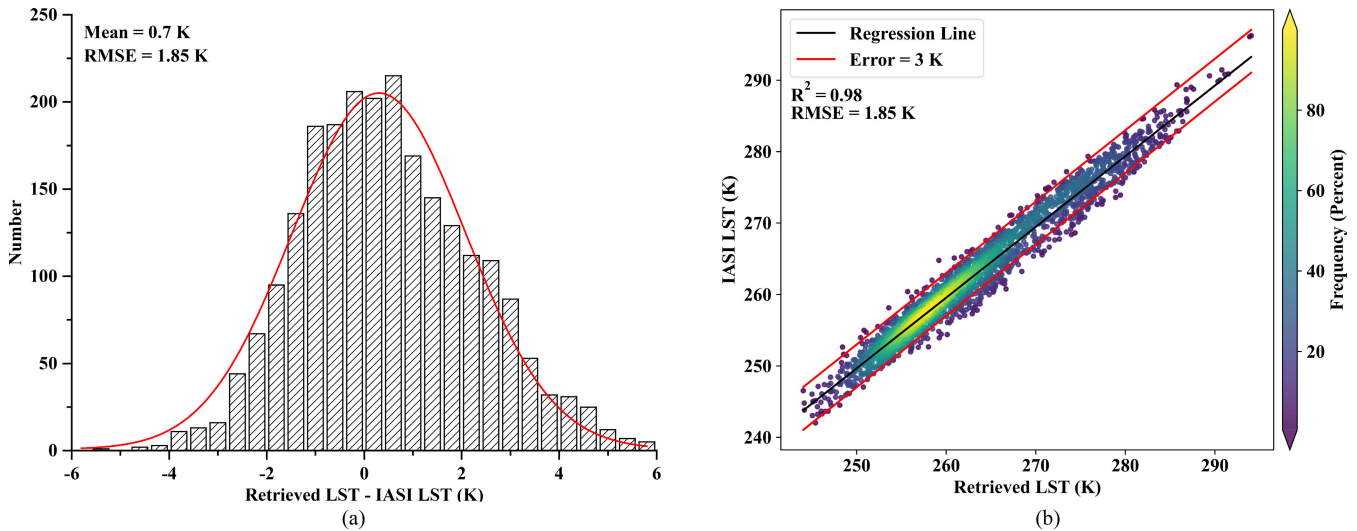


Fig. 10. Comparison of LST retrieval results using the IASI L1C product with the IASI L02 product: (a) shows the absolute error of the LST retrieval results using the IASI L1C product versus the IASI L02 product and (b) shows the scatter plot of the LST retrieval results using the IASI L1C product versus the IASI L02 product.

2.27 K. A further visualization analysis of the BT-GAT outcomes, as shown in Fig. 10, demonstrates that the absolute error distribution of the LST retrieval closely resembles a normal distribution, characterized by a mean of 0.7 K and a standard deviation of 1.85 K. The experimental data suggest a close proximity between the retrieved LST values and the actual LST measurements, which substantiates the efficacy of the BT-GAT BS approach. This validation offers valuable insights for the development of hyperspectral-based LST inversion algorithms.

VI. CONCLUSION AND DISCUSSION

A methodological framework utilizing a specialized graph data structure for efficient hyperspectral data mapping is proposed in this article. This structure enables a detailed analysis of band correlations by examining features of adjacent nodes, thereby providing insights into spectral relationships. A noteworthy aspect of our approach is the integration of object brightness temperature, a significant physical parameter crucial for extracting relevant hyperspectral data features. Additionally, the proposed model incorporates a series of multisize convolutional kernels, which are designed to assign weights to different spectral bands, facilitating a refined approach to BS. This aspect of the model is pivotal in addressing challenges associated with LST retrieval. Compared to existing BS methods, the proposed approach shows improved performance in LST retrieval. This is demonstrated by achieving the RMSE of 1.85 K against the IASI temperature product. This achievement underscores the accuracy and practical applicability of the proposed method in environmental and climatic studies. The contributions of the proposed framework could be summarized as follows:

- 1) This algorithm addresses the gap in methodologies for the thermal infrared hyperspectral BS using deep learning models, providing a reference for subsequent research in thermal infrared hyperspectral remote sensing.

- 2) The manifestation of the Hughes phenomenon in the experimental results emphasizes that indiscriminate augmentation of spectral bands does not necessarily yield improved performance in the field of LST retrieval, thereby fortifying the substantive significance of the BS methodology proposed in this article.
- 3) During the implementation of the graph attention neural network, physical constraints related to blackbody radiation are integrated, alongside the utilization of multiscale convolution with large kernels, effectively enhancing the model's performance compared to other network structures. Additionally, the algorithm has been cross-validated with IASI-related products, confirming its effectiveness and practical value.

The proposed method, characterized by the integration of physical quantities pertinent to LST retrieval, exhibits significant utility in the domain of LST retrieval. However, this specialized focus may inadvertently result in diminished performance when applied to other areas outside of this specific field. Recognizing this limitation, future work will be directed toward developing more generalized algorithms. These algorithms will aim to maintain the efficacy in LST retrieval while broadening the scope of applicability, thereby enhancing versatility and performance across a wider range of applications. This endeavor will involve balancing the specificity required for accurate LST retrieval with the flexibility needed for broader utility.

REFERENCES

- [1] P. Leng et al., "Enhanced surface soil moisture retrieval at high spatial resolution from the integration of satellite observations and soil pedo-transfer functions," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4513711.
- [2] H. Yan, G. Zhou, and X. Lu, "Surface soil moisture retrieval through combining LST-VI feature space with soil porosity," *Int. J. Remote Sens.*, vol. 44, no. 2, pp. 517–541, Jan. 2023.
- [3] J. Tian et al., "Improving surface soil moisture estimation through assimilating satellite Land Surface Temperature with a linear SM-LST relationship," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 7777–7790, 2023.

- [4] S. Ahmadi, H. Alizadeh, and B. Mojaradi, "Land Surface Temperature assimilation into a soil moisture-temperature model for retrieving farm-scale root zone soil moisture," *Geoderma*, vol. 421, Sep. 2022, Art. no. 115923.
- [5] Z. Zhang, W. Xu, Q. Qin, and Z. Long, "Downscaling solar-induced chlorophyll fluorescence based on convolutional neural network method to monitor agricultural drought," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1012–1028, Feb. 2021.
- [6] Y.-A. Liou and M.-T. Thai, "Surface water availability and temperature (SWAT): An innovative index for remote sensing of drought observation," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4507312.
- [7] T. Hu et al., "On agricultural drought monitoring in Australia using Himawari-8 geostationary thermal infrared observations," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 91, Sep. 2020, Art. no. 102153.
- [8] Z. Li et al., "Satellite remote sensing of global Land Surface Temperature: Definition, methods, products, and applications," *Rev. Geophys.*, vol. 61, no. 1, pp. 2–77, Mar. 2023.
- [9] K. Liu, H. Su, X. Li, and S. Chen, "Development of a 250-m downscaled Land Surface Temperature data set and its application to improving remotely sensed evapotranspiration over large landscapes in northern China," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5000112.
- [10] D. Sun et al., "Land Surface Temperature derivation under all sky conditions through integrating AMSR-E/AMSR-2 and MODIS/GOES observations," *Remote Sens.*, vol. 11, no. 14, p. 1704, Jul. 2019.
- [11] P. Wu et al., "A two-step deep learning framework for mapping gap-less all-weather Land Surface Temperature using thermal infrared and passive microwave data," *Remote Sens. Environ.*, vol. 277, Aug. 2022, Art. no. 113070.
- [12] Y. Li et al., "Biophysical impacts of Earth greening can substantially mitigate regional Land Surface Temperature warming," *Nature Commun.*, vol. 14, no. 1, pp. 1–12, Jan. 2023.
- [13] X. Liu et al., "Local temperature responses to actual land cover changes present significant latitudinal variability and asymmetry," *Sci. Bull.*, vol. 68, no. 22, pp. 2849–2861, Nov. 2023.
- [14] X. Shen et al., "Grassland greening impacts on global Land Surface Temperature," *Sci. Total Environ.*, vol. 838, Sep. 2022, Art. no. 155851.
- [15] C. Ru et al., "Land surface temperature retrieval from Landsat 8 thermal infrared data over urban areas considering geometry effect: Method and application," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5000716.
- [16] T. Pede, G. Mountrakis, and S. B. Shaw, "Improving corn yield prediction across the U.S. corn belt by replacing air temperature with daily MODIS Land Surface Temperature," *Agric. Forest Meteorol.*, vols. 276–277, Oct. 2019, Art. no. 107615.
- [17] B. Wu et al., "Challenges and opportunities in remote sensing-based crop monitoring: A review," *Nat. Sci. Rev.*, vol. 10, no. 4, pp. 1–17, Mar. 2023.
- [18] M. Blum, I. M. Lensky, and D. Nestel, "Estimation of olive grove canopy temperature from MODIS thermal imagery is more accurate than interpolation from meteorological stations," *Agric. Forest Meteorol.*, vol. 176, pp. 90–93, Jul. 2013.
- [19] E. Zhao et al., "An operational Land Surface Temperature retrieval methodology for Chinese second-generation huanjing disaster monitoring satellite data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1283–1292, 2022.
- [20] Z. Qin, A. Karnieli, and P. Berliner, "A mono-window algorithm for retrieving Land Surface Temperature from Landsat TM data and its application to the Israel–Egypt border region," *Int. J. Remote Sens.*, vol. 22, no. 18, pp. 3719–3746, Jan. 2001.
- [21] Z.-W. He and B.-H. Tang, "Retrieval of rugged mountainous areas Land Surface Temperature from high-spatial-resolution thermal infrared remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4508216.
- [22] X. Zhu et al., "Retrieval of Land Surface Temperature with topographic effect correction from Landsat 8 thermal infrared data in mountainous areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6674–6687, Aug. 2021.
- [23] Z. Qiao, L. Liu, Y. Qin, X. Xu, B. Wang, and Z. Liu, "The impact of urban renewal on Land Surface Temperature changes: A case study in the main city of Guangzhou, China," *Remote Sens.*, vol. 12, no. 5, p. 794, Mar. 2020.
- [24] L. M. McMillin, "Estimation of sea surface temperatures from two infrared window measurements with different absorption," *J. Geophys. Res.*, vol. 80, no. 36, pp. 5113–5117, Dec. 1975.
- [25] Y. Gui et al., "A physical-based method for pixel-by-pixel quantifying uncertainty of Land Surface Temperature retrieval from satellite thermal infrared data using the generalized split-window algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5001015.
- [26] H. Li et al., "An operational split-window algorithm for generating long-term Land Surface Temperature products from Chinese Fengyun-3 series satellite data," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5004514.
- [27] R. Li et al., "Land Surface Temperature retrieval from sentinel-3A SLSTR data: Comparison among split-window, dual-window, three-channel, and dual-angle algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5003114.
- [28] C. Ru, S.-B. Duan, X.-G. Jiang, Z.-L. Li, C. Huang, and M. Liu, "An extended SW-TES algorithm for Land Surface Temperature and emissivity retrieval from ECOSTRESS thermal infrared data over urban areas," *Remote Sens. Environ.*, vol. 290, May 2023, Art. no. 113544.
- [29] Z. Bian et al., "Retrieval of leaf, sunlit soil, and shaded soil component temperatures using airborne thermal infrared multiangle observations," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4660–4671, Aug. 2016.
- [30] Z. Wan, "New refinements and validation of the MODIS Land-Surface Temperature/emissivity products," *Remote Sens. Environ.*, vol. 112, no. 1, pp. 59–74, Jan. 2008.
- [31] C. Borel, "Error analysis for a temperature and emissivity retrieval algorithm for hyperspectral imaging data," *Int. J. Remote Sens.*, vol. 29, nos. 17–18, pp. 5029–5045, Sep. 2008.
- [32] X. Ye, J. Hui, P. Wang, J. Zhu, and B. Yang, "A modified transfer-learning-based approach for retrieving Land Surface Temperature from Landsat-8 TIRS data," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4411511.
- [33] Y. Chen, Y. Yang, X. Pan, X. Meng, and J. Hu, "Spatiotemporal fusion network for Land Surface Temperature based on a conditional variational autoencoder," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5002813.
- [34] Z. Yin et al., "Spatiotemporal fusion of Land Surface Temperature based on a convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1808–1822, Feb. 2021.
- [35] H. Wang et al., "A method for Land Surface Temperature retrieval based on model-data-knowledge-driven and deep learning," *Remote Sens. Environ.*, vol. 265, Nov. 2021, Art. no. 112665.
- [36] X. Lan, E. Zhao, Z.-L. Li, J. Labed, and F. Nerry, "Deep mixture model-based Land Surface Temperature retrieval for hyperspectral thermal IASI sensor," *IEEE Access*, vol. 8, pp. 218122–218130, 2020.
- [37] C. Yu, S. Zhou, M. Song, B. Gong, E. Zhao, and C.-I. Chang, "Unsupervised hyperspectral band selection via hybrid graph convolutional network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5530515.
- [38] M. P. Uddin, M. A. Mamun, M. I. Afjal, and M. A. Hossain, "Information-theoretic feature selection with segmentation-based folded principal component analysis (PCA) for hyperspectral image classification," *Int. J. Remote Sens.*, vol. 42, no. 1, pp. 286–321, Nov. 2020.
- [39] R. J. Johnson, J. P. Williams, and K. W. Bauer, "AutoGAD: An improved ICA-based hyperspectral anomaly detection algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 6, pp. 3492–3503, Jun. 2013.
- [40] X. Cao, T. Xiong, and L. Jiao, "Supervised band selection using local spatial information for hyperspectral image," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 329–333, Mar. 2016.
- [41] H. Yang, M. Chen, G. Wu, J. Wang, Y. Wang, and Z. Hong, "Double deep Q-network for hyperspectral image band selection in land cover classification applications," *Remote Sens.*, vol. 15, no. 3, p. 682, Jan. 2023.
- [42] J. Feng et al., "Deep reinforcement learning for semisupervised hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5501719.
- [43] C.-I. Chang and S. Wang, "Constrained band selection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1575–1585, Jun. 2006.
- [44] A. Martínez-Usó, F. Pla, J. M. Sotoca, and P. García-Sevilla, "Clustering-based hyperspectral band selection using information measures," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4158–4171, Dec. 2007.
- [45] K. He et al., "A dual global-local attention network for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5527613.

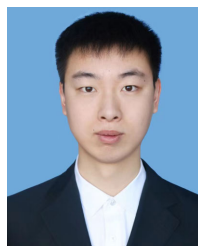
- [46] T. Li, Y. Cai, Z. Cai, X. Liu, and Q. Hu, "Nonlocal band attention network for hyperspectral image band selection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3462–3474, 2021.
- [47] Y. Zhang, X. Wang, X. Jiang, and Y. Zhou, "Robust dual graph self-representation for unsupervised hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5538513.
- [48] X. Ye et al., "Simultaneous estimation of land surface and atmospheric parameters from thermal hyperspectral data using a LSTM-CNN combined deep neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [49] Y. Cai, X. Liu, and Z. Cai, "BS-Nets: An end-to-end framework for band selection of hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1969–1984, Mar. 2020.
- [50] 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.*, vol. 18, no. 12, pp. 2152–2156, Dec. 2021.
- [51] Y. Wang, Q. Zhu, H. Ma, and H. Yu, "A hybrid gray wolf optimizer for hyperspectral image band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5527713.
- [52] M. C. U. Araújo, T. C. B. Saldanha, R. K. H. Galvão, T. Yoneyama, H. C. Chame, and V. Visani, "The successive projections algorithm for variable selection in spectroscopic multicomponent analysis," *Chemometric Intell. Lab. Syst.*, vol. 57, no. 2, pp. 65–73, Jul. 2001.
- [53] Y. Wang et al., "Constrained-target band selection for multiple-target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 6079–6103, Aug. 2019.
- [54] J. Andrey-Andrés et al., "A simulated observation database to assess the impact of the IASI-NG hyperspectral infrared sounder," *Atmos. Meas. Techn.*, vol. 11, no. 2, pp. 803–818, Feb. 2018.



Enyu Zhao (Member, IEEE) was born in Dalian, Liaoning, China, in 1987. He received the Ph.D. degree in cartography and geographic information system from the College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, China, in 2017.

He was a joint Ph.D. Student at the Engineering Science, Computer Science and Imaging Laboratory, University of Strasbourg, Strasbourg, France, from 2014 to 2016. He is currently an Associate Professor with the College of Information Science and

Technology, Dalian Maritime University, Dalian, China. His research interests include quantitative remote sensing and hyperspectral image processing.



Nianxin Qu was born in Liaoyang, Liaoning, China in 1999. He received the B.S. degree in computer science and technology from Liaoning Technical University, Huludao, China, in 2021. He is currently pursuing the M.S. degree in computer science and technology with Dalian Maritime University, Dalian, Liaoning.

His research interests include hyperspectral image processing and deep learning.



Yulei Wang (Member, IEEE) was born in Yantai, Shandong, China, in 1986. She received the B.S. and Ph.D. degrees in signal and information processing from Harbin Engineering University, Harbin, China, in 2009 and 2015, respectively.

She was a joint Ph.D. student at the Remote Sensing Signal and Image Processing Laboratory, University of Maryland, Baltimore County, Baltimore, MD, USA, from 2011 to 2013. From 2011 to 2013, she was a Research Assistant at the Shock, Trauma and Anesthesiology Research organized research center

(STAR-ORC), School of Medicine, University of Maryland, College Park, MD, USA. She is currently an Associate Professor and a Doctoral Supervisor at the Center of Hyperspectral Imaging in Remote Sensing (CHIRS), Information Science and Technology College, Dalian Maritime University, Dalian, China. Her current research interests include hyperspectral image processing, multisource remote sensing fusion, and vital signs signal processing. More details can be found at <https://YuleiWang1.github.io/>.



Caixia Gao (Member, IEEE) received the B.S. degree in electronic and information engineering from Xi'an University of Posts and Telecommunications, Xi'an, China, in 2006, the M.S. degree in computer science from the Academy of Opto-Electronics, Chinese Academy of Sciences, Beijing, China, in 2009, and the Ph.D. degree in cartography and geography information system from the University of Chinese Academy of Sciences, Beijing, in 2012.

She is currently an Associate Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences. Her research interests include in-orbit calibration and validation of optical sensors, and the retrieval of surface temperature and emissivity.



Si-Bo Duan (Senior Member, IEEE) received the Ph.D. degree in cartography and geographical information systems from the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China, in 2014.

He is a Professor with the Key Laboratory of Agricultural Remote Sensing, Ministry of Agriculture and Rural Affairs/Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences. His research interest includes the retrieval and validation of land surface temperature.



Jian Zeng received the M.S. degree in cartography and geographic information systems from Beijing Forestry University, Beijing, China, in 2019.

He is currently a middle Engineer with China Centre for Resources Satellite Data and Application, Beijing. His research interests include radiometric calibration of optical and thermal infrared sensors for on-orbit satellites, surface reflectance, and temperature retrieval from remote sensing data.



Qiang Zhang (Member, IEEE) received the B.E. degree in surveying and mapping engineering and the M.E. and Ph.D. degrees in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2017, 2019, and 2022, respectively.

He is currently an Xinghai Associate Professor with the Center of Hyperspectral Imaging in Remote Sensing (CHIRS), Information Science and Technology College, Dalian Maritime University, Dalian, China. He has authored more than 20 journal articles in *IEEE TRANSACTIONS ON IMAGE PROCESSING*, *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, *Earth System Science Data*, and *ISPRS Journal of Photogrammetry and Remote Sensing*. His research interests include remote sensing information processing, computer vision, and machine learning. More details could be found at <https://qzhang95.github.io/>.