Neurocomputing 616 (2025) 128963

Contents lists available at ScienceDirect

Neurocomputing

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

10-minute forest early wildfire detection: Fusing multi-type and multi-source information via recursive transformer

Qiang Zhang [a], Jian Zhu [a], Yushuai Dong [a], Enyu Zhao [a,*], Meiping Song [a], Qiangqiang Yuan [b]

[a] Center of Hyperspectral Imaging in Remote Sensing (CHIRS), Information Science and Technology College, Dalian Maritime University, Dalian, China
[b] School of Geodesy and Geomatics, Wuhan University, Wuhan, China

ARTICLE INFO

Communicated by N. Navarin

Keywords:
Minute-level
Forest early wildfire detection
Himawari-8/9
Near real-time
Information fusion
Recursive transformer

ABSTRACT

Forest wildfire has great impacts on both nature and human society. While disrupts the ecosystems, wildfire leads to significant economic loss and poses a threat to local communities. To detect forest wildfire, remote sensing technology has become an essential and powerful tool. Compared with polar-orbiting satellite, the new generation of geostationary satellite provides higher temporal resolution and faster response capability. In this study, we utilize the near real-time data of Himawari-8/9 satellite, to achieve 10-min forest early wildfire detection. A recursive transformer model is proposed in this work. It fuses multi-type and multi-source information for Himawari-8/9 satellite. By leveraging the spectral, temporal and spatial features of fire pixels and considering land cover information, the proposed method reduces interference factors like cloud and terrain, resulting in minute-level and near real-time detection of forest wildfire. In 21 ground truth forest wildfire scenarios and MODIS-based cross-validation dataset, the proposed method achieves better results compared to the JAXA wildfire product, in terms of overall fire detection accuracy, early fire detection rate, omission rate, and real-time performance. Furthermore, the proposed framework effectively lowers the emergency response time for early forest wildfire detection, thereby reducing the loss caused by forest wildfire.

1. Introduction

Due to the natural factors such as drought, lightning strikes, and the human factors such as illegal outdoor fires and electrical short circuits, forest wildfires occur frequently all over the world [1]. China is also a country with frequent forest wildfires, forest wildfires cause the immeasurable losses and hazards to both the natural environment and human society [2]. Therefore, it is significant to monitor the dynamic conditions of forest wildfire in real time, adopt corresponding prevention and control measures, and establish the near real-time forest wildfire detection system [3].

Currently, forest wildfire detection methods could be categorized into three types: manual surveillance, unmanned aerial vehicle (UAV)-based detection, and satellite-based remote sensing detection. Among them, manual surveillance requires massive manpower and time resources. Moreover, for large forest areas and complex mountainous terrains, manual surveillance is usually inefficient. As for UAV-based detection method, due to the limited flight range and high costs. It is not applicable for large forest areas [4–6].

Satellite-based remote sensing detection method has rapidly developed in recent years for forest wildfire detection [7]. This method utilizes remote sensing satellites to achieve wide-range, high-efficiency, repeatable, and low-cost surface observations [8]. It has become an inevitable way for large-scale forest wildfire detection. Currently, polar-orbiting satellites and geostationary satellites are primarily used for forest wildfire detection. Polar-orbiting satellites like MODIS, AVHRR, and VIIRS, can acquire remote sensing images with the spatial resolution within hundreds of meters [9–11]. These polar-orbiting satellites have been widely applied for forest wildfire detection [12,13]. However, the temporal resolution of polar-orbiting satellite data is low, ranging from several days to more than ten days. It cannot meet the requirements for near real-time wildfire detection [14,15]. In contrast to polar-orbiting satellites, geostationary satellites such as Himawari-8/9 have lower spatial resolutions, ranging from 500 m to 2 km [16]. Nevertheless, they have higher temporal resolutions (10 min), multiple bands (16 bands), and large amounts of data. These provide the possibility for near real-time forest wildfire detection.

Based on the type of information, existing forest wildfire detection methods based on remote sensing satellites could be classified into three categories: spatial information-based, spectral information-based, and temporal information-based methods.

(a) Spatial information-based: These methods analyze the differences in spatial statistical characteristics between target pixels and

* Corresponding author.
E-mail address: zhaoenyu@dlmu.edu.cn (E. Zhao).

https://doi.org/10.1016/j.neucom.2024.128963
Received 29 May 2024; Received in revised form 21 July 2024; Accepted 13 November 2024
Available online 22 November 2024

background pixels [17,18]. They usually classify the target pixels as fire or non-fire points based on the manual thresholds. Background pixels refer to the pixels surrounding the target pixels within a regional window. Forest wildfire could be detected by selecting appropriate statistical features and threshold parameters [19]. For example, Hally et al. [20] studied the impact of window size and changes in the percentage of contextually valid pixels. Jang et al. [21] took the diurnal and seasonal characteristics of forest fires into account, which proposed an adaptive threshold wildfire detection algorithm. Engel et al. [22] estimated dynamic statistical thresholds for different regions and seasons. They categorized all data into day-time and night-time based on reflectance information. This method employed different criteria to detect fires during different periods. Spatial information-based methods generally have faster processing speed [23]. While these methods are sensitive to threshold parameters. Besides, they are usually affected by factors such as spatial heterogeneity, geographical location and weather, which may result in false alarms or missed detection [24,25]. Overall, spatial information-based methods have advantages on large-scale wildfire detection. Nevertheless, they still have the limitations in stability and small-scale or early-stage fire detection.

(b) Spectral information-based: This method relies on the specific characteristics via different bands, such as changes in temperature or smoke [26]. For example, Giglio et al. [27] proposed an improved contextual algorithm using AVHRR sensor data, which achieved a low false alarm rate in complex terrain. However, this algorithm cannot detect wildfire under low solar zenith angle. Subsequently, Giglio et al. [28] used the reflectance information of 0.65 μm and 0.86 μm bands in MODIS sensor, as well as the brightness temperature information of 12 μm band, to eliminate water and cloud pixels. It utilized the spectral differences between 4 μm and 11 μm bands for surface temperature changes. Spectral information-based methods usually require a lot of prior knowledge and are difficult to select the optimal threshold parameters for wildfire detection [29,30]. Additionally, under cloud conditions, the information received by the sensor may be interfered with [31].

(c) Temporal information-based: This category of methods primarily analyzes the time-series brightness temperature changes for forest wildfire detection [32,33]. When the light temperature changes abnormally, it is often accompanied by the occurrence of wildfires. The key for temporal information-based methods is how to use historical time-series data to predict the true background brightness temperature and set the appropriate threshold [34,35]. However, due to the influence of clouds, It may be discontinuous in the historical data, which greatly hinders the estimation of background brightness temperature. For example, Hally et al. [36] constructed a diurnal temperature cycle (DTC) model and introduced singular value decomposition to fit the time-series data. Additionally, Xie et al. [37] developed a robust time-series fitting method. It determines fire points by calculating the deviations between observed and predicted values. Generally, temporal information-based methods could effectively utilize the high-frequency data of geostationary satellites for near real-time wildfire detection [38]. However, these methods are also susceptible to factors such as cloud covering and weather, which severely affects the robustness of forest wildfire detection [39].

In summary, the three information models mentioned above still have their own shortcomings and limitations. Furthermore, most existing wildfire detection methods have a large time postponement, which makes it difficult to meet the urgent needs of emergency and firefighting departments [40]. Therefore, the future direction of forest wildfire detection lies in how to overcome above limitations. Including: considering different types of features and information, building adaptive detection models, and addressing the requirements of high reliability, near real-time and early wildfire detection.

Building on our previous research [41], this work develops a novel method for minute-level forest early wildfire detection. The proposed framework fuses multi-type and multi-source information for Himawari-8/9 satellite. The main innovations of this paper are listed as follows:

(1) The proposed framework integrates multi-type information by leveraging the temporal, spatial, and spectral information of geostationary satellite data. It overcomes the limitations of existing wildfire detection methods that could only use single or dual types of information. It also effectively mitigates the interference factors such as cloud and fog.

(2) The proposed framework considers the multi-source heterogeneous information by introducing land cover products, to distinguish different land surface types. This removes the influence of non-combustible elements, thereby improving the accuracy and robustness of forest early wildfire detection.

(3) The proposed framework constructs a recursive Transformer model that could effectively mine global and contextual features of fire points. In contrast to traditional recursive models, the output at each time step is used to update the input at the next time step. Additionally, this model could achieve 10-minute level forest early wildfire detection, and significantly reduces time consumption with limited samples.

The rest of this paper is organized as follows. Section 2 provides the data descriptions of Himawari-8/9 satellite and MODIS land cover products. Section 3 describes the proposed forest early wildfire detection framework. Section 4 evaluates the wildfire detection results and overall indicators in multiple experimental scenarios. Section 5 discusses the significance of multi-source and multi-type data. Finally, the conclusion is summarized in Section 6.

2. Data description

2.1. Himawari-8/9 satellite AHI data

Himawari-8/9 satellite, designed by the Japan Aerospace Exploration Agency (JAXA), were respectively launched at 2014 and 2022, with the coverage over east Asia and Australia region (60°S~60°N, 80°E~160°W) [42]. As the third-generation geostationary satellite, Himawari-8/9 aims to provide near real-time earth observation data to support applications such as meteorology, environmental monitoring, and natural disaster detection. It has 16 bands, including 3 visible, 3 near-infrared, and 10 infrared bands [43]. The spatial resolution is 0.5 km for visible bands and 1 or 2 km for near-infrared and infrared bands. The full disk observation frequency is every 10 min [5]. Himawari-8/9 is regarded as one of the most advanced meteorological satellites, providing more effective tools for meteorological detection and environmental protection [14].

In this study, we utilize the AHI (Advanced Himawari Imager) data from Himawari-8/9 satellite for near real-time forest early wildfire detection. We download the AHI data of Himawari-8/9, which includes two formats: HSD and NetCDF4. For early wildfire detection, we use the 10-min full disk NetCDF4 data for processing. During the data processing, the 7th band of Himawari-8/9 is sensitive to surface temperature changes, while the 14th band is not. Therefore, we use the brightness temperature data of the 7th and 14th bands as the main data for forest wildfire detection.

2.2. MODIS land cover products

MODIS is a large-scale space sensor developed by NASA, which serves as a medium-resolution imaging spectrometer and provides remote sensing data with multi-spectral and high radiometric sensitivity. Among its wide range of applications, land cover products are extensively utilized [44]. These products utilize the multi-spectral data from MODIS and analyze the reflectance information across different wavelengths to classify various land cover types, such as water bodies, forests, grasslands, croplands, and urban areas. These land cover products have significant application value on land, ocean, and atmospheric

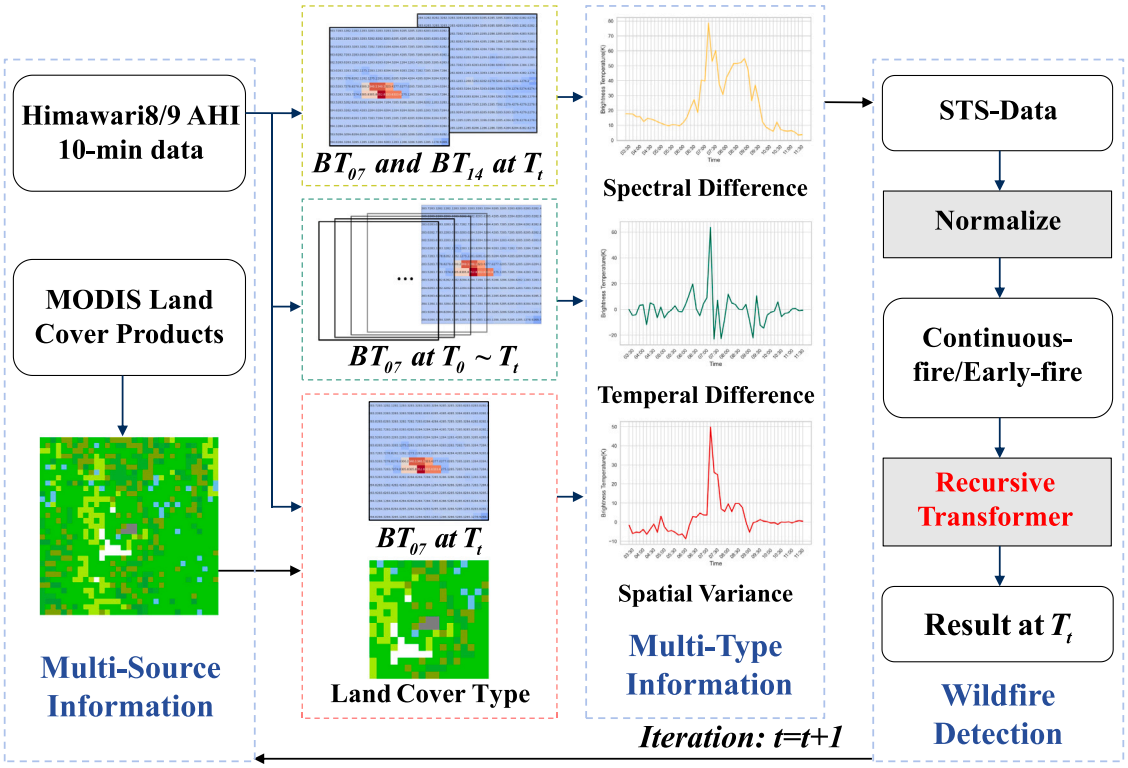


Fig. 1. Flowchart of the proposed framework for minute-level forest early wildfire detection.

monitoring. In this study, different land cover types exhibit varying levels of combustibility, which is an important factor influencing the occurrence of wildfire [45]. The LAADS platform developed by NASA provides access to these products. We download the land cover yearly L3 global 500 m data for the China region, stitching and reprojection by the MODIS Reprojection Tool (MRT) software. Finally, down-sampling is applied to align these land cover products with the Himawari-8/9 data.

3. Methodology

3.1. Overview

When a forest wildfire firstly occurs, there is usually a dramatic change of the surface temperature at this location [46]. Due to the different radiometric reflection characteristics of the heat source in the 7th and 14th bands, the brightness temperature of the fire pixel in the 7th band will significantly increase. While its changes in the 14th band are relatively gentle. Based on these characteristics, we comprehensively consider the spectral, temporal, and spatial features. As shown in Fig. 1, spectral feature is calculated based on the difference between the 7th and 14th bands. Temporal feature is obtained by differentiating the brightness temperature of the fire pixel in the 7th band at different times. Spatial feature is acquired by calculating the variance between the target pixel and the background window. In addition, we take the heterogeneous information into consideration from land cover products, to eliminate the influence of non-combustible material pixels. Then, we normalize these information to amplify feature abnormalities and differences for subsequent learning and prediction. Finally, depending on the type of forest wildfire, we input the data into the recursive Transformer model, which is used for information fusion and forest wildfire detection, as shown in Fig. 1.

3.2. Multi-type and multi-source information

Choosing appropriate information that aligns with the inherent characteristics of forest wildfire, is crucial for forest fire detection. Firstly, forest wildfire typically occurs in combustible areas. Therefore, introducing land cover type information could effectively eliminate non-combustible regions and reduce false alarms. Secondly, for forest wildfire, there is usually a significant change in time-series curve of the fire pixel in the 7th band of Himawari-8/9, during the early-stage wildfire. Thus, incorporating temporal information could enhance the detection accuracy of early-stage fire pixels [47]. Thirdly, during the wildfire occurrence, there is usually an obvious temperature difference between the fire pixel and its spatial neighbors in the 7th band. Hence, spatial information could be used to distinguish the fire pixels from non-fire pixels. Lastly, the 7th band is more sensitive to the changes for land surface temperature, compared with the 14th band. Therefore, spectral information should be introduced to enhance the accuracy of wildfire detection.

In summary, this work fuses multi-type and multi-source information, including land cover information, temporal information, spatial information and spectral information. The specific descriptions are given as follows:

(1) Land cover information: In this study, the 500 m MODIS land cover product is used, which includes 10 different land types. These types are water body, grassland, shrubland, broadleaf cropland, savanna, evergreen broadleaf forest, deciduous broadleaf forest, evergreen needleleaf forest, deciduous needleleaf forest, non-vegetated land, urban and built-up area. To address the inconsistent spatial resolution between these products and Himawari-8/9 data, the land cover products are processed through mosaic reprojection and down-sampling operations, resulting in a consistent spatial resolution of 2 km for all the products and data.

(2) Spatial information: The background pixels are selected within the current contextual window, which have the same land cover type

with the target pixel. Then, the brightness temperature difference between the target in the 7th band and the mean brightness temperature of all the background pixels, is utilized as the spatial information:

$$SP_t = BT_{07}^t(x,y) - \frac{\sum BT_{07}^t(i,j) - BT_{07}^t(x,y)}{N_b} \tag{1}$$

where N_b is the number of background pixels. t represents the current time.

(3) Temporal information: We establish the consecutive time-series brightness temperature curve of the target pixel in the 7th band. Then, the temporal difference between the current and previous moments' brightness temperature is calculated as:

$$TI_t = BT_{07}^t(x,y) - BT_{07}^{t-1}(x,y) \tag{2}$$

where (x, y) represents the position of the target pixel. BT_i stands for the brightness temperature value in the i-th band.

(4) Spectral information: We obtain both the time-series brightness temperature values of the target pixel in the 7th and 14th bands. Then, the spectral difference is determined as:

$$OS_t = BT_{07}^t(x,y) - BT_{14}^t(x,y) \tag{3}$$

Based on above operations, we build up the temporal information TI, spatial information SP and spectral information OS, respectively. Among them, $TI = \{TI_1, \ldots, TI_t\}$, $SP = \{SP_1, \ldots, SP_t\}$, $OS = \{OS_1, \ldots, OS_t\}$. Subsequently, normalization is performed on these temporal, spatial and spectral information, respectively. Then, these information is converted into time-series vectors. The normalization procedure is operated below:

$$X_t' = \frac{X_t - \min(X)}{\max(X) - \min(X) + \varepsilon} \tag{4}$$

where X stands for $\{TI, SP, OS\}$. $\min(X)$ and $\max(X)$ are the minimum and maximum values of the time-series curve, respectively. ε is a tiny parameter to prevent division by zero. X' is stands for the normalized sequence $\{TI', SP', OS'\}$.

3.3. Recursive transformer model

As shown in Fig. 2. In this study, we innovatively adopt an iterative Transformer model, which is different from traditional recursive models. Compared with the method of predicting the subsequent long time series data directly from the original value, our model makes the prediction in an iterative manner. Specifically, we first predict the value of the current moment, overwrite the actual value of the current moment with the resulting predicted value. Then iteratively increase the time step, calling the model to predict the value of the next moment. In this way, we can eliminate the influence of actual outliers on the subsequent time series prediction, and greatly improve the reliability of time series prediction.

This kind of iterative predictive model is called the iterative Transformer model. By iteratively predicting values at the current moment, we are able to more accurately capture small fluctuations in fire development, improving the accuracy and real-time performance of fire monitoring. Compared to traditional methods, our iterative Transformer model shows greater applicability and accuracy when processing fire monitoring data. The proposed model could effectively fuse multi-type and multi-source information. Below, we specify the components of the recursive Transformer model.

The overall structure of the recursive Transformer model is given in Fig. 2. It consists of a window coding layer, a context coding layer and two decoding layers. By adopting adversarial training strategy, the proposed model can extract time-series features from the input data, and output the prediction information at the next time t. W is the input window data. k denotes the window length.

According to the input information, the window data is represented as follows:

$$W_t = \begin{cases} \{TI'_{t-k+1}, \ldots, TI'_t\} \\ \{SP'_{t-k+1}, \ldots, SP'_t\} \\ \{OS'_{t-k+1}, \ldots, OS'_t\} \end{cases} \tag{5}$$

The proposed model converts the input time-series X' to the moving window data $W = \{W_1, \ldots, W_t\}$. We consider the time slice until the current timestamp t of a series X', and denote it as C_t. If t is less than k, we fill it with the last value of the window.

Then, the multivariable sequence W and C are converted into a matrix whose dimensions are time step and feature number m. The scaled dot product of three matrices Q (query), K (key) and V (value) is defined as:

$$\mathrm{Att}(Q,K,V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{m}}\right)V \tag{6}$$

To ensure the training stability, the dot product is scaled (divided by a scaling factor $\sqrt{m}$) to reduce the variance of the weights. Via multiplying each value and weight, we sum them to get the final output vector. For the input matrices Q, K and V, we use the multi-head self-attention mechanism. By passing them to the feedforward layer, Q_i, K_i and V_i are obtained where $i \in \{1, \ldots, h\}$. Then the scaled dot product is utilized as:

$$\mathrm{MultiHA}(Q,K,V) = \mathrm{Concat}(H_1, \ldots, H_h) \tag{7}$$

where H_i denotes $\mathrm{Att}(Q_i, K_i, V_i)$. Multi-head attention mechanism could allow model to pay attention to the information from different representation subspaces at different locations.

After above processing, the recursive Transformer model makes the time-series prediction via two stages.

In the first stage, we take W, C and focus fraction F (an initial zero matrix) as input. We broadcast F to match the dimension of W with appropriate zero padding, and then concatenate the two values. This step uses the zero padding to update F. Next, we apply the positional coding to above result and get the input of the first encoder, denoted as I_1. The first encoder performs the following operations:

$$\begin{aligned} I_1^1 &= \mathrm{LayerNorm}\left(I_1 + \mathrm{MultiHA}\left(I_1, I_1, I_1\right)\right) \\ I_1^2 &= \mathrm{LayerNorm}\left(I_1^1 + \ \mathrm{FeedForward}\ \left(I_1^1\right)\right) \end{aligned} \tag{8}$$

The above operations generate attention weight values using the input time-series window and the complete sequence, to capture the temporal feature of the input data. For window encoder, the recursive Transformer model applies the positional encoding for the input window W, to obtain the second encoder I_2. The second encoder carries out the following operations:

$$\begin{aligned} I_2^1 &= \mathrm{Mask}\left(\mathrm{MultiHA}\left(I_2, I_2, I_2\right)\right) \\ I_2^2 &= \mathrm{LayerNorm}\left(I_2 + I_2^1\right) \\ I_2^3 &= \mathrm{LayerNorm}\left(I_2^2 + \mathrm{MultiHA}\left(I_1^2, I_1^2, I_1^2\right)\right) \end{aligned} \tag{9}$$

Finally, we use two identical decoders (O_1 and O_2) to execute the following operation:

$$O_i = \mathrm{Sigmoid}\left(\mathrm{FeedForward}\left(I_2^3\right)\right) \tag{10}$$

where i denotes 1 and 2, respectively representing two decoders. The $\mathrm{Sigmoid}()$ activation function is utilized to generate the output value, which matches the input window W. To sum up, the input values of the recursive Transformer model are W and C, and the output values are O_1 and O_2.

In order to ensure the training stability, an adversarial training program is designed that uses outputs from two independent decoders.

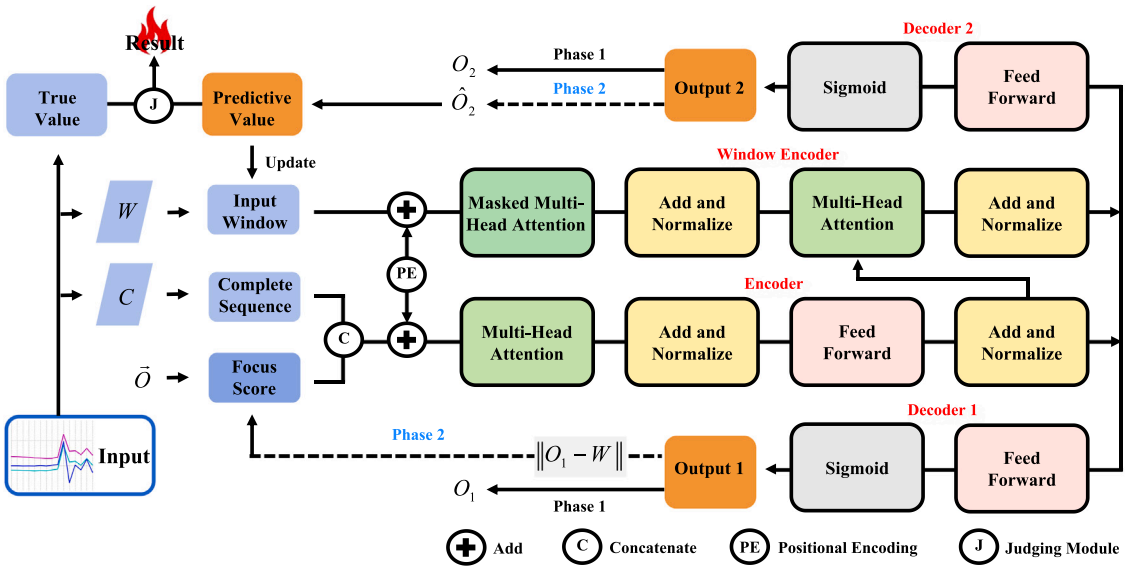


Fig. 2. Overall structure of the recursive Transformer model.

The L_2 norm and the output of the first stage are employed to define the reconstruction loss for each decoder:

$$\begin{aligned}\zeta_1 &= \|O_1 - W\|_2 \\ \zeta_2 &= \|O_2 - W\|_2\end{aligned} \tag{11}$$

In the second stage, we use the reconstruction loss of the first decoder as the focus score. The output value of the second decoder is $\hat{O}_2$.

Then, the adversarial loss obtained from the output of the second stage is introduced. The objective of the first decoder is to minimize the reconstruction error of the self-conditioned output. While the objective of the second decoder is to maximize the reconstruction error of the self-conditioned output. Therefore, based on the output results of the first stage, the training objective of the two decoders is established as follow:

$$\min_{\text{Decoder 1}} \max_{\text{Decoder 2}} \left\|\hat{O}_2 - W\right\|_2 \tag{12}$$

The loss function is defined as follows:

$$\begin{aligned}\zeta_1' &= +\left\|\hat{O}_2 - W\right\|_2 \\ \zeta_2' &= -\left\|\hat{O}_2 - W\right\|_2\end{aligned} \tag{13}$$

Finally, the cumulative loss for each decoder is determined in Eq. (13). In this procedure, the recursive Transformer model simultaneously employs the reconstruction loss and adversarial loss from two stages, resulting in the joint loss function as follows:

$$\begin{aligned}\text{Loss}_1 &= \lambda^{-n}\zeta_1 + (1-\lambda^{-n})\,\zeta_1' \\ \text{Loss}_2 &= \lambda^{-n}\zeta_2 + (1-\lambda^{-n})\,\zeta_2'\end{aligned} \tag{14}$$

where n stands for the training number. λ is the scaling parameter.

Then, the three time-series features are separately predicted by the recursive Transformer model. It outputs the sequence $X_p = \{TI_p, SP_p, OS_p\}$. For time t, we take the predicted value X_p^t for this time from X_p. Behind, the difference between the predicted value and the actual value is calculated to obtain the abnormal deviation. Finally, the abnormal discrimination formula is established to detect the forest early or continuous wildfire points, respectively. The discrimination formula is defined as follows:

$$\begin{aligned}TI_{\text{fire}} &= \left(TI_A^t - TI_P^t\right) > \alpha_1 \cdot \text{Avg}\left(\Gamma_{t=1}^{k-1}\left(TI_A^t - TI_P^t\right)\right) \cap \left(TI_A^t - TI_P^t\right) \geq \beta_1 \\ SP_{\text{fire}} &= \left(SP_A^t - SP_P^t\right) > \alpha_2 \cdot \text{Max}\left(\Gamma_{t=1}^{k-1}\left(SP_A^t - SP_P^t\right)\right) \cap \left(SP_A^t - SP_P^t\right) \geq \beta_2 \\ OS_{\text{fire}} &= \left(OS_A^t - OS_P^t\right) > \alpha_3 \cdot \text{Max}\left(\Gamma_{t=1}^{k-1}\left(OS_A^t - OS_P^t\right)\right) \cap \left(OS_A^t - OS_P^t\right) \geq \beta_3\end{aligned} \tag{15}$$

For early wildfire points, they are classified as the fire points if all three discriminant formulas meet the abnormal conditions; Otherwise, they are classified as the non-fire points. For continuous wildfire points, they are classified as the fire points if both spatial and spectral features meet the abnormal conditions; Otherwise, they are classified as the non-fire points.

4. Experiments and results

4.1. Overview

In this section, we divide our experiments into two parts. The first part is the ground truth fire scenario, and 21 explicitly reported forest wildfire cases are selected as experimental scenarios in Sections 4.2 to 4.4. The second part is a cross-validation dataset constructed based on MODIS fire products in Section 4.5. It uses thousands of fire scenarios to validate the accuracy of the model. The assignment of the individual subsections is detailed below. In Section 4.2, we introduce the metrics to evaluate the performance of the model. In Section 4.3, we show the visual detection results of part of the real fire scene. In Section 4.4, the detection metrics for all 21 real wildfire scenarios are summarized and compared with the following three comparison algorithms: the JAXA WLF L2 products, the multispectral-based thresholding algorithm, and the temporal-spectral-based thresholding algorithm. In Section 4.5, the model accuracy is validated based on the MODIS cross-validation dataset.

4.2. Hyper-parameter selection

The size of background window is an important factor, affecting the efficiency and accuracy of wildfire detection. In order to balance the performance and efficiency, we choose a side length of 7 pixels as the size of the background window. It is used to calculate the spatial variance. The statistical characteristics of spatial information could be well captured in multiple scenes. Thus, effective detection of early wildfire could be achieved for the proposed model.

For the input length, the proposed model takes a sequence of 20 vectors as input from the non-fire period at different time steps. For each time step, the input value consists of a central pixel and its neighboring pixels.

Table 1
The confusion matrix of actual/predicted fire or non-fire.

	Predicted fire	Predicted non-fire
Actual Fire	a	b
Actual Non-Fire	c	d

For the training parameters of the proposed model, the learning rate is fixed as 0.001. Batch size is set as 128. Training epoch is set as 100. The learning rate drop step is set as 5 and the drop multiple is set as 0.9. Window length is set as 10.

For the key parameters α_1, α_2, α_3, β_1, β_2 and β_3 in the proposed model, they are set as 1.2, 1, 1, 5, 0 and 0, respectively.

4.3. Wildfire scenarios and evaluation metrics

We select a total of 21 forest wildfire scenarios from the past six years as experimental data. These wildfire scenarios includes various types, such as small-scale, day-time and night-time wildfires. Among them, the forest wildfires are mainly distributed in Yunnan, Sichuan, Liaoning and other places. We also examined two wildfires in Russia and India.

To evaluate the model performance, we compare the experimental results with the JAXA WLF L2 products. The spatial and temporal resolution of JAXA WLF L2 products is 2 km and 10 minutes, respectively. Five evaluation metrics are employed in this study to assess the accuracy of forest early wildfire detection: Early Fire Accuracy (EFA), Overall Accuracy (OA), Omission Fire Rate (OFR), False Alarm Rate (FAR), and Average Delay Time (ADT) of the early wildfire detection. The confusion matrix demonstrates the correspondence between actual and predicted fire, as listed in Table 1.

Based on this confusion matrix, the above EFA, OA, FAR, OFR and ADT indexes are defined as follows:

$$\mathrm{EFA} = \frac{EF_{\mathrm{predict}}}{EF_{\mathrm{all}}} \tag{16}$$

$$\mathrm{OA} = \frac{a+d}{a+b+c+d} \tag{17}$$

$$\mathrm{OFR} = \frac{b}{a+b} \tag{18}$$

$$\mathrm{FAR} = \frac{c}{a+c} \tag{19}$$

$$\mathrm{ADT} = \frac{\sum_{i=1}^{n} t_i}{n} \tag{20}$$

where EF_{predict} is the early wildfire number predicting correctly. EF_{all} stands for the real early wildfire number. The total number of experimental scenarios denotes n. The early wildfire detection delay time of each scenario refers to t_i, which represents the delay between the earliest wildfire detection time and the actual early wildfire occurrence time. The average value is calculated in all scenarios.

4.4. Results of forest wildfire detection

4.4.1. Forest early wildfire detection

The accuracy of forest early wildfire detection is an important index for evaluating the detection models. Near real-time detecting of potential forest wildfire and taking preventive measures could reduce the loss to a large extent. We present four specific results from the total 21 forest wildfire scenarios, to validate the effectiveness of the proposed framework for forest early wildfire detection.

(1) Scenario 1: Around UTC Time 04:30 at March 13, 2021, a forest wildfire broke out in Eralingou, Machang Village, Yuanzhou District, Guyuan City, Ningxia. The location of this wildfire is approximately 106.1° E, 35.96° N. This wildfire affected area of over 500 acres.

During the initial stage, the time-series spatial variance, temporal difference, and spectral difference curves exhibit a smooth trend. Nevertheless, around UTC Time 04:30, all the three curves generate vibrations to different degrees. It indicates the occurrence of forest wildfire at this time. As shown in Fig. 3, the detection results indicate that both the JAXA WLF L2 products and the proposed method detect the wildfire occurrence around UTC Time 04:30. However, the JAXA WLF L2 products just detect the obvious fire points. While these products ignore the low-temperature fire points. Due to the threshold-based strategy, JAXA WLF L2 products are not sensitive to low-temperature fires. In comparison, the proposed method is able to detect all the fire points well, which is the main reason for the lower omission rate.

(2) Scenario 2: Around UTC Time 4:00 at May 2, 2017, a forest wildfire took place in the Beidahe Forest Farm of Bila River Forestry Bureau, Great Khingan, Inner Mongolia. The geographical location is 123.06° E and 49.52° N. The area affected by this wildfire was 11,500 hectares, whose forest land accounted for 60% and the forest disaster area was 8281.58 hectares. A total of 9430 people and 14 aircrafts were mobilized for firefighting. At the same day, the National Forest Fire Command issued an emergency notice for forest wildfires prevention. A comprehensive deployment of national forest wildfires prevention work was conducted at the afternoon of May 3. It demonstrates the high importance that the country places on preventing forest wildfires. However, this is still a case of hindsight. Therefore, detecting forest wildfire as quickly as possible in the early stage is extremely crucial, to avoid the greater losses.

As shown in Fig. 4, at UTC Time 3:40, the proposed method is capable of promptly detecting forest early wildfire. While JAXA WLF L2 products cannot detect any fire points in the early stage of wildfire. These products only detected the forest wildfire at UTC Time 6:00. Wildfire had already spread extensively at this time. Compared with the proposed method, JAXA WLF L2 products experienced a delay of 140 min, significantly reducing warning value of geostationary satellite for forest wildfire rescue.

(3) Scenario 3: Around UTC Time 10:10 at May 6, 2023, a small-scale forest wildfire occurred in Chahuaqing, Qinglong Town, Anning City, Kunming, Yunnan Province. Its geographical location is 102.32° E, 25.04° N. The Brigade of Kunming City urgently dispatched 122 firefighters, and extinguished this forest wildfire after a day and a night of rescue efforts.

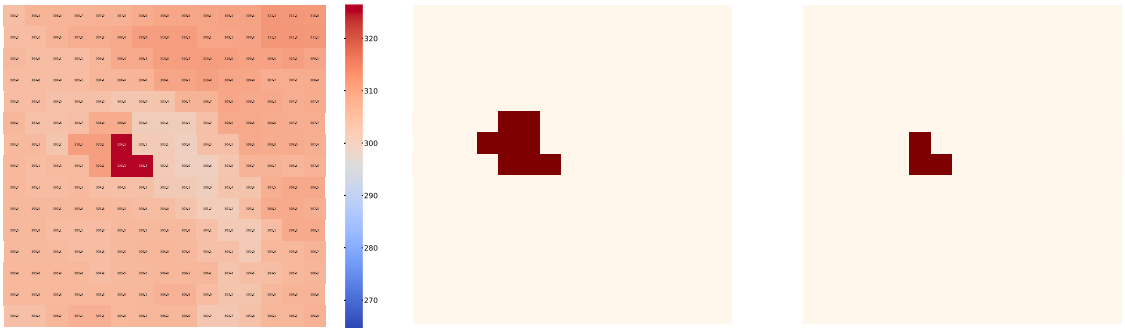
Fig. 5 depicts the forest early wildfire detection results of the proposed method and JAXA WLF L2 products in Scenario 3, during the initial stage of the wildfire. Small-scale wildfires pose a big challenge to wildfire detection algorithms. It is evident that JAXA's products cannot promptly detect the fire point. The primary reason may lie in the use of contextual algorithm for JAXA WLF L2 products. The small-scale wildfire pixels is difficult to exceed the set threshold in contextual algorithm. By fusing spectral–temporal-spatial information and land cover data, the proposed model demonstrates favorable detection capability for small-scale wildfire. This result also validates the robustness and reliability of the proposed framework.

(4) Scenario 4: Around UTC Time 5:50 at April 17, 2019, a forest wildfire broke out in the eastern outskirts of Qipanshan, Shenyang City, Liaoning Province. The location is 123.64° E, 41.92° N. This forest wildfire affected an area of approximately 828 hectares and resulted in a direct economic loss totaling 24.605 million RMB.

As illustrated in Fig. 6, the proposed method is able to accurately detect the occurrence of wildfire in the early stage (UTC Time 6:00). In contrast, JAXA L2 WLF products cannot detect any fire pixels at this moment. This indicates the high sensitivity of the proposed method for forest wildfire. By fusing multi-source and multi-type information, the proposed method could effectively detect forest early wildfire.

4.4.2. Long-term forest wildfire detection

Around UTC Time 7:55 at April 10, 2023, a forest wildfire took place on a hillside near Qionghai in Xichang City, Liangshan Prefecture, Sichuan Province. The location is 102.34° E, 27.78° N. The long-term bright temperature images of the 7th band and the results of wildfire



(a) BT_{07} at UTC 04:30 (b) Proposed Method (c) JAXA WLF L2 Products

Fig. 3. Forest early wildfire detection results of the proposed method and JAXA WLF L2 products in Scenario 1 (Red points in (b) and (c) refer to the detection fire points).

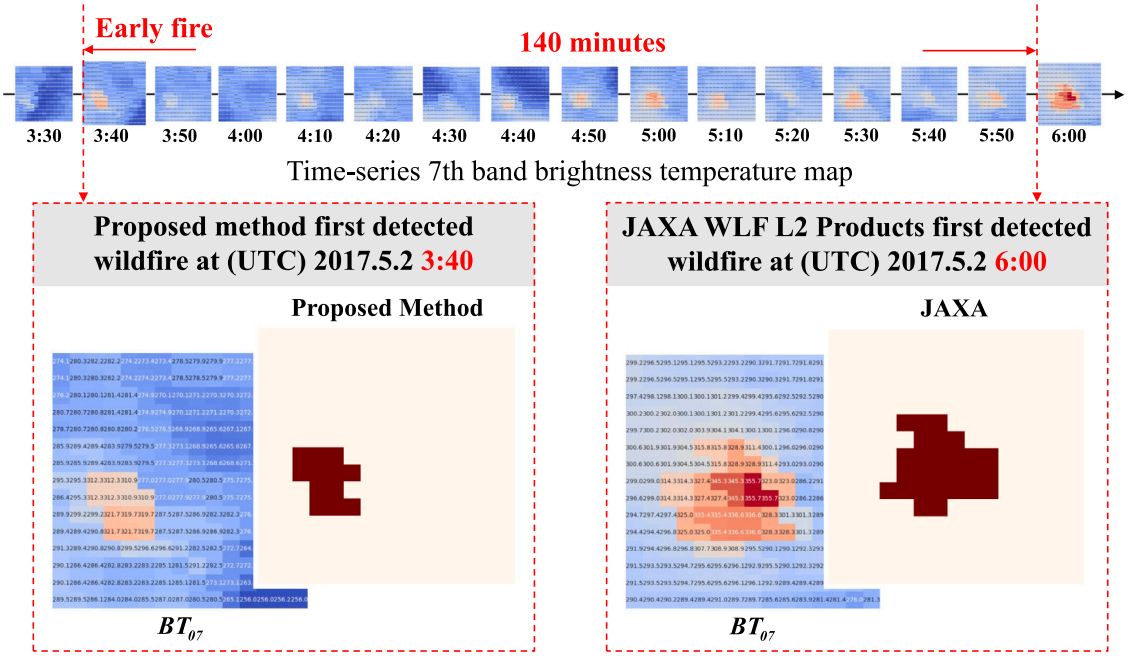
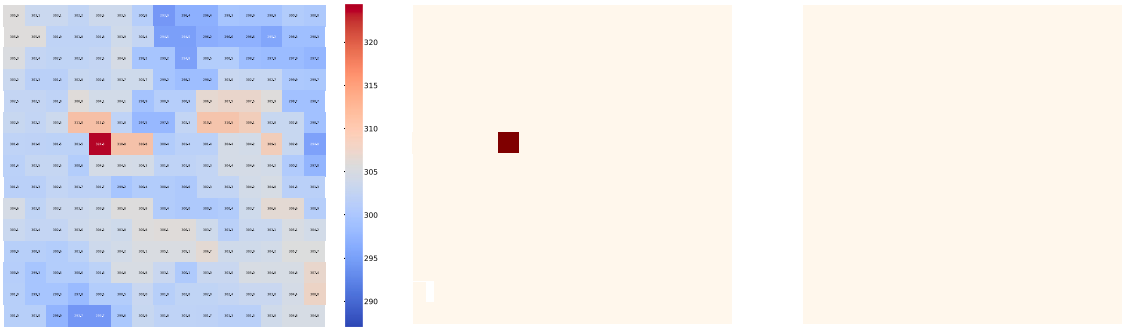


Fig. 4. Time of the first wildfire detection for the proposed method and JAXA WLF L2 products in Scenario 2.

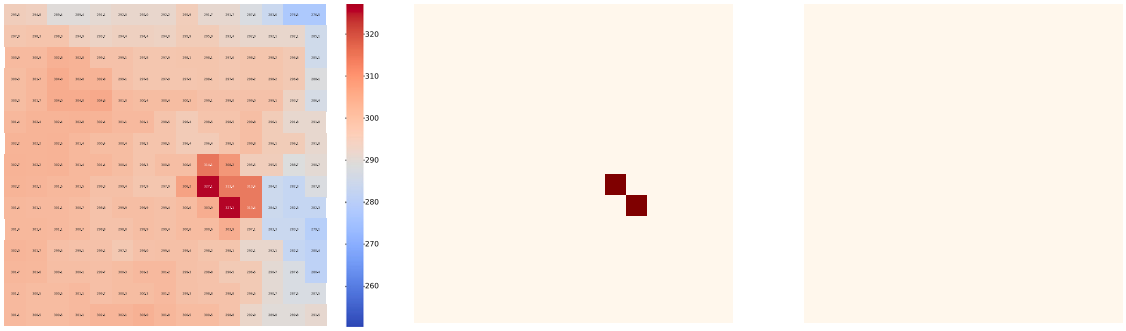


(a) BT_{07} at UTC 08:00 (b) Proposed Method (c) JAXA WLF L2 Products

Fig. 5. Forest early wildfire detection results of the proposed method and JAXA WLF L2 products in Scenario 3.

detection are both shown in Fig. 7. In the detection results, red points indicate the fire pixels detected by two methods. While the other points indicate non-fire pixels. Before UTC Time 8:00, the bright temperature appeared normally in this area. However, starting from UTC Time 8:00, the bright temperature sharply increased, indicating the occurrence of forest wildfire around this moment.

In the last row of Fig. 7, for the forest early wildfire (from UTC Time 07:50 to 08:00), the JAXA WLF L2 products only detect single fire point. In contrast, the proposed method could effectively detect the



(a) BT_{07} at UTC 06:00 (b) Proposed Method (c) JAXA WLF L2 Products

Fig. 6. Forest early wildfire detection results of the proposed method and JAXA WLF L2 products in Scenario 4.

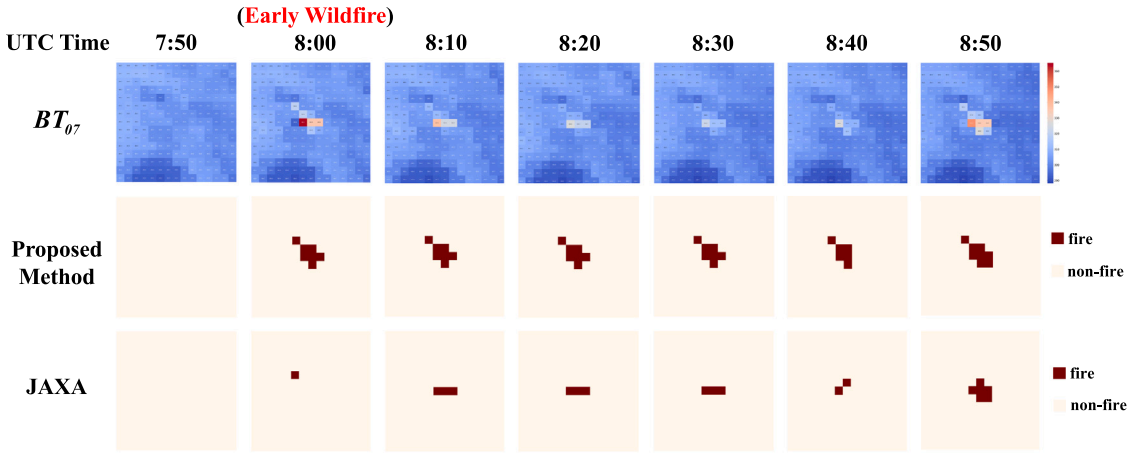


Fig. 7. Long-term forest wildfire detection results of the proposed method and JAXA WLF L2 products.

majority of fire points. In the long-term forest wildfire detection results, the JAXA WLF L2 products also exhibit a large number of missed fire points. Whereas the proposed method could consistently detect the vast majority of fire points. These results fully demonstrate the effectiveness of the proposed method for the long-term forest wildfire detection.

4.5. Overall detection results of forest early wildfire

4.5.1. Comparison algorithms

To test the validity of the proposed model, we chose to compare two comparison approaches and JAXA WLF L2 products.

(1) Multispectral-based threshold algorithm: The improved algorithm first uses Himawari-8's 3.9 μm and 11.2 μm bands to identify potential hotspots [48]. False positives of potential hot spots are suppressed using water masks (2.3 μm band) and cloud masks (0.64 μm, 0.86 μm, and 12.4 μm band). The low albedo in visible band (0.64 μm and 0.86 μm) allows for night recognition. Fig. 8 summarizes the threshold based algorithm proposed using multiple spectral bands. It is represented as MTA in the following content.

(2) Temporal-spectral-based threshold algorithm: This algorithm introduce the time-series information of the fire [21]. The algorithm uses a non-fixed threshold, the first 7% value of each image is used as an adaptive threshold. Multiple temporal components are considered in the threshold-based algorithm, to reduce false detection. Fig. 9 summarizes the temporal-spectral-based threshold algorithm. It is represented as TTA in the following content.

4.5.2. Result and analysis

In this subsection, we summarize the detection results of all forest early wildfire scenarios, and evaluate the overall performance of the proposed method and three comparison algorithms: JAXA, MTA and TTA. Specifically, we conduct the statistical analysis of the Early Fire Accuracy (EFA), Overall Accuracy (OA), Omission Fire Rate (OFR), False Alarm Rate (FAR), and Average Delay Time (ADT) of forest early fire detection in all the experimental scenarios. These indicators not only reflect the comprehensive performance of the model in various types of forest wildfires, but also reveal the model's overall performance

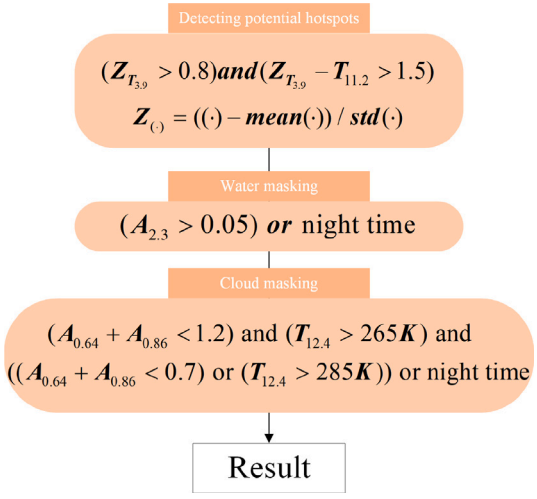


Fig. 8. Multispectral-based threshold algorithm.

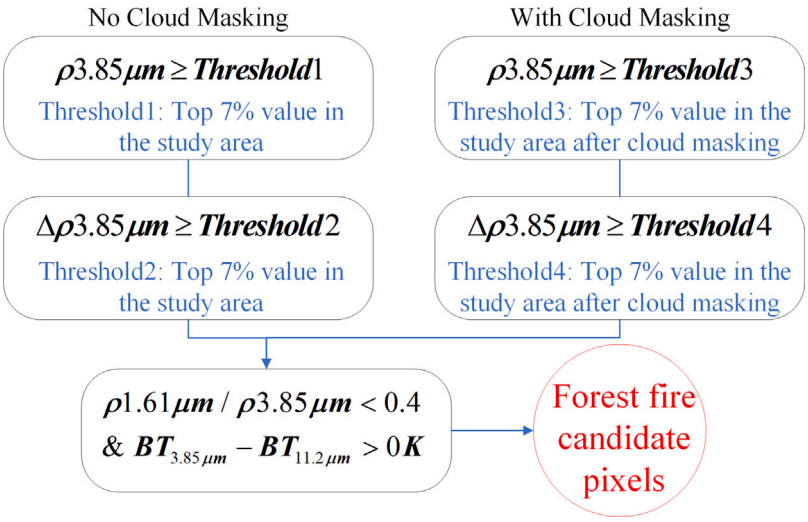


Fig. 9. Temporal-spectral-based threshold algorithm.

Table 2
Quantitative evaluation indexes of the proposed method, JAXA WLF L2 products, MTA and TTA for 21 forest wildfire scenarios.

Products	OA(↑)	EFA(↑)	FAR(↓)	OFR(↓)	ADT(min)(↓)
JAXA	98.65%	56.91%	13.58%	43.09%	15.71
MTA	98.26%	77.47%	22.53%	22.53%	0
TTA	97.38%	75.27%	36.57%	24.73%	0
Proposed	**99.89%**	**96.75%**	**0.83%**	**3.25%**	**0**

Table 3
Wildfire detection results of MODIS cross-validation dataset.

Number of active fire scenes	Detected scenes	Undetected scenes	False detected scenes	Overall accuracy
1606	1522	60	24	94.77%

Table 4
Details of the wildfire example for the MODIS cross-validation dataset.

Date	UTC Time	Longitude and latitude	Number of data
2023–12–23	13:17:00	137.08°E, −18.90°N	35

under different types, geographical environments, and weather conditions. We use the occurrence time of early wildfire in each scenario as the baseline, and conduct experimental analysis on 21 scenarios, totaling 4725 detection pixels. The relevant quantitative evaluation indexes are given in Table 2. The better results are highlighted with bold format.

As shown in Table 2, JAXA WLF L2 products exhibit lower sensitivity for early wildfire, resulting in higher false detection and omission rates of 13.58% and 43.09%, respectively. The false detection rate and missed detection rate of MTA algorithm are 22.53% and 22.53%, respectively. Although the false detection rate of this method is higher than that of JAXA method, the omission detection rate is greatly reduced. Because the spectral information is introduced in the proposed method, the detection accuracy of the model is significantly improved in Table 2. The false detection rate and omission rates of TTA algorithm are 36.57% and 24.73%, respectively. The proposed method introduces the temporal information, improving the overall detection accuracy.

On the contrary, the false detection rate and omission detection rate of the proposed method are 0.83% and 3.25%, respectively. They are lower than the other three comparison algorithms. In comparison with other products, the proposed method demonstrates higher accuracy, with the early wildfire accuracy and overall accuracy of 96.75% and 99.89%, respectively. It has a 39.84% improvement in early wildfire detection accuracy than JAXA WLF L2 products. This difference is mainly attributed to the contextual algorithm used in the JAXA WLF L2 products, which calculates thresholds using spatial information from the surrounding area. Contextual algorithm only considers pixels exceeding the thresholds as fire pixels. However, early and small-scale fire pixels usually have lower brightness temperatures, making it difficult to reach the setting thresholds. This leads to the potential omission of fire pixels and reduces the accuracy of forest wildfire detection.

In contrast, the proposed method comprehensively considers temporal, spectral, spatial and land cover information. It establishes an adaptive discriminant criterion to enhance the sensitivity of wildfire and generalization ability. The overall accuracy of the proposed method also surpasses JAXA WLF L2 products.

More critically, in terms of the ADT for early wildfire detection, although JAXA WLF L2 products could detect forest early wildfires in some scenarios in a timely manner, its ADT achieves 15.71 min. The maximum delay for detecting early wildfire in single scene even exceeds 140 min in Fig. 4. Due to the low brightness temperature and limitations of the contextual algorithm, JAXA WLF L2 products overlooks most forest early wildfires, which is the primary reason for its poor ADT index. In contrast, the proposed method demonstrates higher accuracy in forest early wildfire detection and better ADT index in Table 2.

4.6. MODIS cross-validation dataset

To verify the detection accuracy of the proposed model on long time-series, large dataset, and the portability in different scenarios, a large-scale MODIS cross-validation dataset is constructed. The dataset is based on the MODIS active fire products' overall active fire data in December 2023, which ensures the authenticity of the fire points. Specifically, we reserve all fire points with a confidence level above 80% to ensure the reliability. Clustering with a space radius of 200 km and a time span of 7 days are performed for better validation. Wildfire detection results of MODIS cross-validation dataset are listed in Table 3.

As shown in Table 3, the model demonstrates significant efficacy on long time-series and the portability in different scenarios. It achieves an impressive overall fire detection rate of 96.21%. Furthermore, its robust performance underscores its adaptability to diverse geographical regions and climates, affirming its strong portability. We take one of the MODIS cross-validation dataset as an example and visualize its detection results. The details of this example are shown in Table 4.

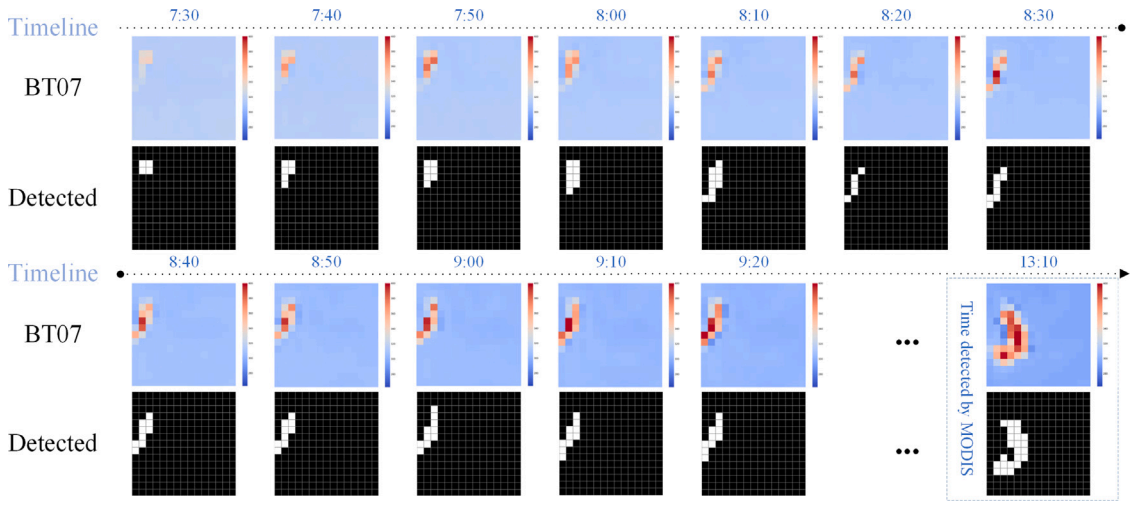


Fig. 10. Long time-series detection results for MODIS cross-validation dataset.

The wildfire detection results are shown in Fig. 10. As a part of the MODIS cross-validation dataset, the fire spreading in this area could be visually interpreted based on the brightness temperature of the 7th band. As displayed in Fig. 10, the proposed model has a high sensitivity to capture fire. Therefore, the spread of forest fire could be clearly identified and detected.

In summary, through the fusion of multi-type and multi-source information, the proposed method achieves reliable results in near-real-time detection of forest wildfires. Particularly, the proposed model exhibits favorable performance in detection of long-term, early and small-scale forest wildfires.

5. Discussions

5.1. Analysis of spectral-temporal-spatial information

Fig. 11 presents the time-series curves of spatial variation, temporal difference, and spectral difference for the forest wildfire in Chahuaqing, Qinglong Town, Anning City, Kunming, Yunnan Province at May 6, 2023. As shown in Fig. 11, different amplitude fluctuations of the three types of information appear at UTC 1:00 and 5:00. Nevertheless, the proposed method does not classify them as forest wildfires at these moments. Because the model integrates spectral–temporal-spatial information, it only identifies the early wildfire points when the obvious changes simultaneously occur in these three categories of information. Through this integration and comprehensive judgment of spectral–temporal-spatial information, the proposed model effectively distinguishes early wildfire pixels from non-fire pixels. It could exclude interference factors such as cloud and fog, thereby enhancing the accuracy and reliability of wildfire detection. This is also one of the main reasons why the proposed model exhibits high robustness for forest early wildfire detection [49].

Through supplementing the ablation experiments, we conduct a comprehensive analysis of spatial and temporal features for the proposed model. Furthermore, by analyzing the experimental results of different algorithms, we underscore the significance and contributions of spatial and temporal features to the final results.

In the ablation experiments, we used the forest fire in Sichuan Province, at April 10, 2023 as the study case. We respectively eliminate the spatial and temporal information from the proposed model, and then evaluate the wildfire detection performance. These ablation experiments aim to assess the importance of two types of information in wildfire detection. The results of the ablation experiments are summarized in Table 5 below.

Table 5
The impact of spatial and temporal information for the proposed method.

Detection model	OA(↑)	EFA(↑)	FAR(↓)	OFR(↓)
Without spatial feature	99.42%	85.94%	6.78%	14.06%
Without temporal feature	98.84%	90.63%	25.64%	9.37%
Proposed method	**99.78%**	**92.19%**	**0.00%**	**7.81%**

As shown in Table 5, eliminating the spatial information from the proposed model significantly increases the omission rate of fire detection, particularly for detecting small and low-temperature fires. This highlights the crucial role of spatial information in accurately identifying such incidents. Moreover, eliminating the temporal information leads to a higher rate of false alarms, reaching 25.64% in Table 5. The reason is that, the continuous and dense temporal information not only provides the distinctive early fire features, but also helps the proposed model distinguish the incidental events such as smoke or illumination changes, thereby reducing the false alarms. Models lacking temporal information may struggle to differentiate these incidental events, leading to higher false alarm rates.

The proposed model integrates temporal, spatial, and spectral information, allowing for the better learning of forest fire characteristics, especially in early-stage wildfire detection. As listed in Table 2, our model achieves optimal performance in overall detection accuracy, false alarms, and omissions.

Additionally, we compare the proposed model with three contrasting algorithms. The MTA algorithm utilizes the spectral and spatial information while neglecting the temporal information. The TTA algorithm employs spectral and temporal information while neglecting the spatial information. As indicated in Table 2, across all 21 scenarios, both of these comparative algorithms notably underperform compared to the proposed model, which integrates spectral, temporal, and spatial information. This also indirectly demonstrates the significant role and contribution of both temporal and spatial information for wildfire detection.

5.2. Analysis of land cover information

Land cover products provide the information for land cover types. It could be more accurate to determine whether the abnormal hot spot is caused by a fuel fire. Particularly in complex terrain such as mountainous areas, considering land cover type information could better eliminate non-fire pixels. It could improve the quality of spatial feature, thereby reducing the false alarm rate and improving overall fire detection accuracy for forest wildfire detection.

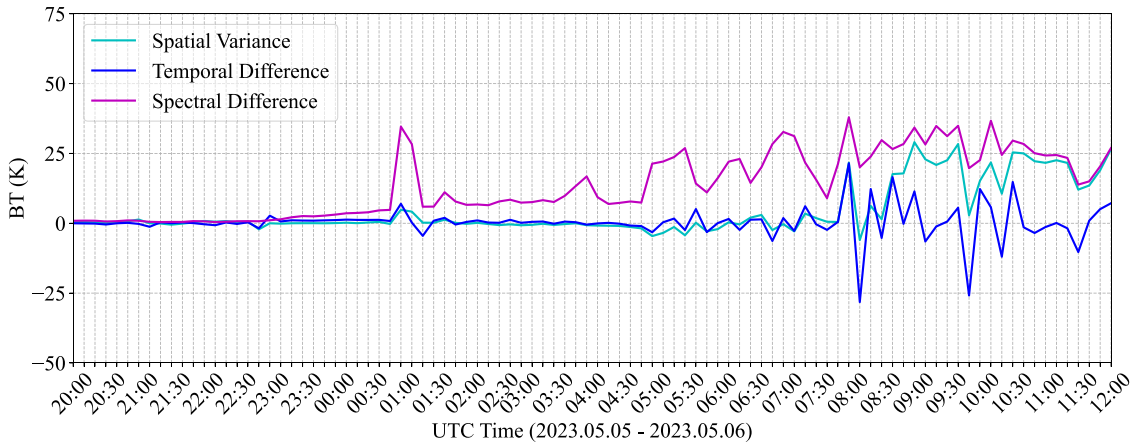


Fig. 11. Time-series curves of spatial variation, temporal difference, and spectral difference.

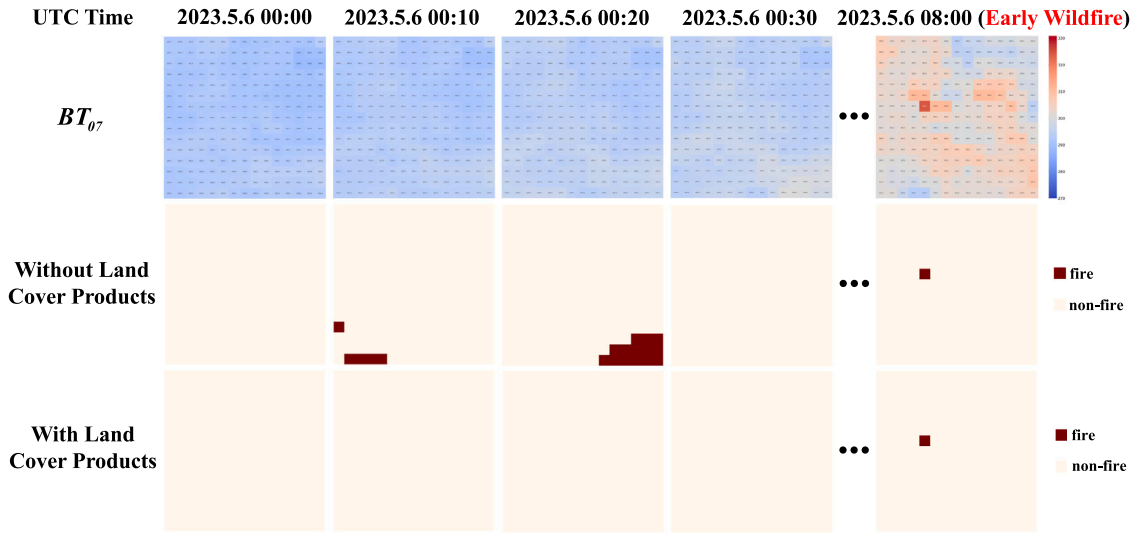


Fig. 12. Effectiveness analysis of land cover information.

To assess the efficacy of land cover products for forest wildfire detection, we utilize two types of models: one incorporating land cover products and another excluding them. To illustrate the difference between the two models, we analyze the detection results for the forest wildfire in Anning City, Kunming, Yunnan Province, at May 6, 2023. Fig. 12, the brightness temperature in the 7th band and the visualized fire point detection results for the two models are depicted at corresponding times. From Fig. 12, the brightness temperature in the 7th band remained at a low level before the forest wildfire occurred. Obviously, the model without the use of land cover products exhibits false fire point detection at UTC time 00:10 and 00:20 at May 6, 2023. In contrast, after integrating land cover products, the proposed model effectively suppresses false fire point detection, significantly enhancing the accuracy. This result availably demonstrates the importance and effectiveness of fusing land cover information for forest wildfire detection.

6. Conclusion

This work utilizes near real-time data of Himawari-8/9 satellite for 10-min forest early wildfire detection. The proposed recursive Transformer model fuses multi-type and multi-source information. This model is capable of simultaneously extracting spectral, temporal and spatial features of fire pixels. Experimental results demonstrate the applicability of the proposed method for forest early wildfire detection under various scenarios. In comparison with the JAXA WLF L2 products, the proposed method achieves quicker and more reliable results. In the future work, our target is to combine geostationary and polar-orbiting satellites, which could simultaneously utilize high temporal and spatial resolution data for forest wildfire detection.

CRediT authorship contribution statement

Qiang Zhang: Validation, Methodology. **Jian Zhu:** Writing – original draft, Methodology, Data curation. **Yushuai Dong:** Validation, Software. **Enyu Zhao:** Writing – review & editing, Validation, Supervision, Methodology. **Meiping Song:** Resources, Project administration. **Qiangqiang Yuan:** 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: Qiang Zhang reports financial support was provided by National Natural 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.

Acknowledgments

This work is supported in part by the Fundamental Research Funds for the Central Universities under Grant 3132024262; And in part by the China Postdoctoral Science Foundation under Grant 2023M740460; And in part by the National Natural Science Foundation of China under Grant 42271355 and 62401095.

Data availability

Data will be made available on request.

References

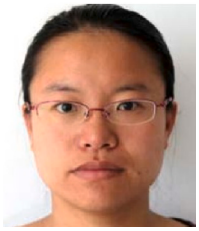
[1] M.A. Moritz, E. Batllori, R.A. Bradstock, A.M. Gill, J. Handmer, P.F. Hessburg, J. Leonard, S. McCaffrey, D.C. Odion, T. Schoennagel, et al., Learning to coexist with wildfire, Nature 515 (7525) (2014) 58–66.
[2] E. Zervas, A. Mpimpoudis, C. Anagnostopoulos, O. Sekkas, S. Hadjiefthymiades, Multisensor data fusion for fire detection, Inf. Fusion 12 (3) (2011) 150–159.
[3] E.H. Chowdhury, Q.K. Hassan, Operational perspective of remote sensing-based forest fire danger forecasting systems, ISPRS J. Photogramm. Remote Sens. 104 (2015) 224–236.
[4] Q. Zhang, Y. Dong, Q. Yuan, M. Song, H. Yu, Combined deep priors with low-rank tensor factorization for hyperspectral image restoration, IEEE Geosci. Remote Sens. Lett. 20 (2023) 1–5.
[5] G.H. de Almeida Pereira, A.M. Fusioka, B.T. Nassu, R. Minetto, Active fire detection in Landsat-8 imagery: A large-scale dataset and a deep-learning study, ISPRS J. Photogram. Remote Sens. 178 (2021) 171–186.
[6] Y. Gao, M. Hao, Y. Wang, L. Dang, Y. Guo, Multi-scale coal fire detection based on an improved active contour model from Landsat-8 Satellite and UAV images, ISPRS Int. J. Geo-Inf. 10 (7) (2021) 449.
[7] S.W. Maier, J. Russell-Smith, A.C. Edwards, C. Yates, Sensitivity of the MODIS fire detection algorithm (MOD14) in the savanna region of the Northern Territory, Australia, ISPRS J. Photogramm. Remote Sens. 76 (2013) 11–16.
[8] D. Ienco, R. Gaetano, R. Interdonato, A constrastive semi-supervised deep learning framework for land cover classification of satellite time series with limited labels, Neurocomputing 567 (2024) 127031.
[9] W. Xu, M.J. Wooster, Sentinel-3 SLSTR active fire (AF) detection and FRP daytime product-Algorithm description and global intercomparison to MODIS, VIIRS and Landsat AF data, Sci. Remote Sens. 7 (2023) 100087.
[10] S.T. Seydi, V. Saeidi, B. Kalantar, N. Ueda, A.A. Halin, Fire-Net: A deep learning framework for active forest fire detection, J. Sensors 2022 (2022) 1–14.
[11] W. Schroeder, P. Oliva, L. Giglio, I.A. Csiszar, The New VIIRS 375 m active fire detection data product: Algorithm description and initial assessment, Remote Sens. Environ. 143 (2014) 85–96.
[12] P. Kennedy, A. Belward, J. Gregoire, An improved approach to fire monitoring in West Africa using AVHRR data, Int. J. Remote Sens. 15 (11) (1994) 2235–2255.
[13] S. Flasse, P. Ceccato, A contextual algorithm for AVHRR fire detection, Int. J. Remote Sens. 17 (2) (1996) 419–424.
[14] D. Zhang, C. Huang, J. Gu, J. Hou, Y. Zhang, W. Han, P. Dou, Y. Feng, Real-time wildfire detection algorithm based on VIIRS fire product and Himawari-8 data, Remote Sens. 15 (6) (2023) 1541.
[15] J. Chen, Q. Lv, S. Wu, Y. Zeng, M. Li, Z. Chen, E. Zhou, W. Zheng, C. Liu, X. Chen, et al., An adapted hourly Himawari-8 fire product for China: principle, methodology and verification, Earth Syst. Sci. Data Discuss. 2023 (2023) 1–32.
[16] K. Chatzopoulos-Vouzoglanis, K.J. Reinke, M. Soto-Berelov, S.D. Jones, One year of near-continuous fire monitoring on a continental scale: Comparing fire radiative power from polar-orbiting and geostationary observations, Int. J. Appl. Earth Obs. Geoinf. 117 (2023) 103214.
[17] C. Liu, R. Chen, B. He, Integrating machine learning and a spatial contextual algorithm to detect wildfire from Himawari-8 data in southwest China, Forests 14 (5) (2023) 919.
[18] S. Movaghati, F. Samadzadegan, A. Azizi, An agent-based algorithm for forest fire detection, in: ISPRS Congress Beijing, Vol. 37, 2008, pp. 631–634.
[19] T. Wang, Y. Wang, F. Zhao, H. Feng, J. Liu, L. Zhang, N. Zhang, G. Yuan, D. Wang, A spatio-temporal temperature-based thresholding algorithm for underground coal fire detection with satellite thermal infrared and radar remote sensing, Int. J. Appl. Earth Obs. Geoinf. 110 (2022) 102805.
[20] B. Hally, L. Wallace, K. Reinke, S. Jones, C. Engel, A. Skidmore, Estimating fire background temperature at a geostationary scale—An evaluation of contextual methods for AHI-8, Remote Sens. 10 (9) (2018) 1368.
[21] E. Jang, Y. Kang, J. Im, D.-W. Lee, J. Yoon, S.-K. Kim, Detection and monitoring of forest fires using Himawari-8 geostationary satellite data in South Korea, Remote Sens. 11 (3) (2019) 271.
[22] C.B. Engel, S.D. Jones, K. Reinke, A seasonal-window ensemble-based thresholding technique used to detect active fires in geostationary remotely sensed data, IEEE Trans. Geosci. Remote Sens. 59 (6) (2020) 4947–4956.
[23] Y. Xiao, Q. Yuan, J. He, Q. Zhang, J. Sun, X. Su, J. Wu, L. Zhang, Space-time super-resolution for satellite video: A joint framework based on multi-scale spatial-temporal transformer, Int. J. Appl. Earth Obs. Geoinf. 108 (2022) 102731.
[24] J. Carrasco, D. López, I. Aguilera-Martos, D. García-Gil, I. Markova, M. García-Barzana, M. Arias-Rodil, J. Luengo, F. Herrera, Anomaly detection in predictive maintenance: A new evaluation framework for temporal unsupervised anomaly detection algorithms, Neurocomputing 462 (2021) 440–452.
[25] X. Wu, X. Lu, H. Leung, Video smoke separation and detection via sparse representation, Neurocomputing 360 (2019) 61–74.
[26] C. Maffei, R. Lindenbergh, M. Menenti, Combining multi-spectral and thermal remote sensing to predict forest fire characteristics, ISPRS J. Photogram. Remote Sens. 181 (2021) 400–412.
[27] L. Giglio, J. Kendall, C. Justice, Evaluation of global fire detection algorithms using simulated AVHRR infrared data, Int. J. Remote Sens. 20 (10) (1999) 1947–1985.
[28] L. Giglio, J. Descloitres, C.O. Justice, Y.J. Kaufman, An enhanced contextual fire detection algorithm for MODIS, Remote Sens. Environ. 87 (2–3) (2003) 273–282.
[29] Q. Zhang, Q. Yuan, J. Li, F. Sun, L. Zhang, Deep spatio-spectral Bayesian posterior for hyperspectral image non-iid noise removal, ISPRS J. Photogramm. Remote Sens. 164 (2020) 125–137.
[30] Q. Zhang, Q. Yuan, M. Song, H. Yu, L. Zhang, Cooperated spectral low-rankness prior and deep spatial prior for HSI unsupervised denoising, IEEE Trans. Image Process. 31 (2022) 6356–6368.
[31] Q. Zhang, Q. Yuan, Z. Li, F. Sun, L. Zhang, Combined deep prior with low-rank tensor SVD for thick cloud removal in multitemporal images, ISPRS J. Photogramm. Remote Sens. 177 (2021) 161–173.
[32] S. Strydom, M.J. Savage, A spatio-temporal analysis of fires in South Africa, South Afr. J. Sci. 112 (11–12) (2016) 1–8.
[33] S. Chen, Y. Cao, X. Feng, X. Lu, Global2Salient: Self-adaptive feature aggregation for remote sensing smoke detection, Neurocomputing 466 (2021) 202–220.
[34] L. He, X. Gong, S. Zhang, L. Wang, F. Li, Efficient attention based deep fusion CNN for smoke detection in fog environment, Neurocomputing 434 (2021) 224–238.
[35] S. Tuli, G. Casale, N.R. Jennings, Tranad: Deep transformer networks for anomaly detection in multivariate time series data, 2022, arXiv preprint arXiv:2201.07284.
[36] B. Hally, L. Wallace, K. Reinke, S. Jones, A. Skidmore, Advances in active fire detection using a multi-temporal method for next-generation geostationary satellite data, Int. J. Digit. Earth (2018).
[37] Z. Xie, W. Song, R. Ba, X. Li, L. Xia, A spatiotemporal contextual model for forest fire detection using Himawari-8 satellite data, Remote Sens. 10 (12) (2018) 1992.
[38] H. Feng, L. Zhang, X. Yang, Z. Liu, Enhancing class-incremental object detection in remote sensing through instance-aware distillation, Neurocomputing 583 (2024) 127552.
[39] Q. Zhang, Q. Yuan, J. Li, Z. Li, H. Shen, L. Zhang, Thick cloud and cloud shadow removal in multitemporal imagery using progressively spatio-temporal patch group deep learning, ISPRS J. Photogramm. Remote Sens. 162 (2020) 148–160.
[40] P. Jain, S.C. Coogan, S.G. Subramanian, M. Crowley, S. Taylor, M.D. Flannigan, A review of machine learning applications in wildfire science and management, Environ. Rev. 28 (4) (2020) 478–505.
[41] Q. Zhang, J. Zhu, Y. Huang, Q. Yuan, L. Zhang, Beyond being wise after the event: Combining spatial, temporal and spectral information for Himawari-8 early-stage wildfire detection, Int. J. Appl. Earth Obs. Geoinf. 124 (2023) 103506.
[42] K. Bessho, K. Date, M. Hayashi, A. Ikeda, T. Imai, H. Inoue, Y. Kumagai, T. Miyakawa, H. Murata, T. Ohno, et al., An introduction to Himawari-8/9—Japan's new-generation geostationary meteorological satellites, J. Meteorol. Soc. Japan. Ser. II 94 (2) (2016) 151–183.
[43] Z. Hong, Z. Tang, H. Pan, Y. Zhang, Z. Zheng, R. Zhou, Z. Ma, Y. Zhang, Y. Han, J. Wang, et al., Active fire detection using a novel convolutional neural network based on Himawari-8 satellite images, Front. Environ. Sci. 10 (2022) 794028.
[44] Y. Vetrita, M.A. Cochrane, Fire frequency and related land-use and land-cover changes in Indonesia's peatlands, Remote Sens. 12 (1) (2019) 5.
[45] K. Muhammad, J. Ahmad, S.W. Baik, Early fire detection using convolutional neural networks during surveillance for effective disaster management, Neurocomputing 288 (2018) 30–42.
[46] J. Wu, W. Zeng, F. Yan, Hierarchical Temporal Memory method for time-series-based anomaly detection, Neurocomputing 273 (2018) 535–546.
[47] Z. Liu, A.P. Ballantyne, L.A. Cooper, Biophysical feedback of global forest fires on surface temperature, Nature Commun. 10 (1) (2019) 214.
[48] G. Xu, X. Zhong, Real-time wildfire detection and tracking in Australia using geostationary satellite: Himawari-8, Remote Sens. Lett. 8 (11) (2017) 1052–1061.
[49] Q. Zhang, Y. Zheng, Q. Yuan, M. Song, H. Yu, Y. Xiao, Hyperspectral image denoising: From model-driven, data-driven, to model-data-driven, IEEE Trans. Neural Netw. Learn. Syst. (2023).

Qiang Zhang 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 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 twenty journal articles in the IEEE TIP, IEEE TNNLS, IEEE TGRS, ISPRS P&RS, ESSD, and JAG. His research interests include remote sensing information processing, computer vision, and machine learning.

Jian Zhu received the B.S. degree in Computer Science from the Anhui University of Finance and Economics, Bengbu, China, in 2021. He is currently pursuing the M.S. degree with the School of Information Science and Technology College, Dalian Maritime University, Dalian, China. His research interests include wildfire detection and machine learning.

Yushuai Dong received the B.S. degree in Food Science and Engineering from the Huazhong Agricultural University, Wuhan, China, in 2022. He is currently pursuing the M.S. degree with the School of Information Science and Technology College, Dalian Maritime University, Dalian, China. His research interests include hyperspectral image processing and machine learning.

Enyu Zhao 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 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.



Meiping Song received the Ph.D. degree from the College of Computer Science and Technology, Harbin Engineering University, Harbin, China, in 2006. She is a Professor with the College of Information Science and Technology, Dalian Maritime University, Dalian, China, since 2020. Her research includes remote sensing and hyperspectral image processing.

Qiangqiang Yuan received the B.S. degree in surveying and mapping engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2006 and 2012, respectively. In 2012, he joined the School of Geodesy and Geomatics, Wuhan University, where he is currently a professor. His current research interests include image reconstruction, remote sensing image processing and application, and data fusion.