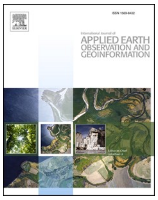
Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

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



Near-real-time wildfire detection approach with Himawari-8/9 geostationary satellite data integrating multi-scale spatial–temporal feature

Lizhi Zhang [a], Qiang Zhang [b], Qianqian Yang [c], Linwei Yue [d,*], Jiang He [a], Xianyu Jin [a], Qiangqiang Yuan [a,e]

[a] School of Geodesy and Geomatics, Wuhan University, Wuhan, China
[b] Center of Hyperspectral Imaging in Remote Sensing (CHIRS), Information Science and Technology College, Dalian Maritime University, Dalian, China
[c] Department of Geography, Faculty of Arts and Social Sciences, Hong Kong Baptist University, 999077, Hong Kong Special Administrative Region of China
[d] School of Geography and Information Engineering, China University of Geosciences Wuhan, China
[e] Key Laboratory of Polar Environment Monitoring and Public Governance, Ministry of Education, Wuhan University, Wuhan, China

ARTICLE INFO

Keywords:
Near-real-time wildfire detection
Multi-scale
Spatial–temporal feature fusion
Deep learning
Himawari-8/9

ABSTRACT

Wildfires pose a great threat to the ecological environment and human safety. Therefore, rapid and accurate detection of wildfires holds significant importance. However, existing wildfire detection methods neglect the full integration of spatial–temporal relationships across different scales, and thus suffer from issues of low robustness and accuracy in varying wildfire scenes. To address this, we propose a deep learning model for near-real-time wildfire detection, where the core idea is to integrate multi-scale spatial–temporal features (MSSTF) to efficiently capture the dynamics of wildfires. Specifically, we design a multi-kernel attention-based convolution (MKAC) module for extracting spatial features representing the differences between fire and non-fire pixels within multi-scale receptive fields. Moreover, a long short-term Transformer (LSTT) module is used to capture the temporal differences from the image sequences with different window lengths. The two modules are combined into multiple streams to integrate the multi-scale spatial–temporal features, and the multi-stream features are then fused to generate the fire classification map. Extensive experiments on various fire scenes show that the proposed method is superior to JAXA Wildfire products and representative deep learning models, achieving the best accuracy scores (i.e., average fire accuracy (FA): 88.25%, average false alarm rate (FAR): 20.82%). The results also show that the method is sensitive to early-stage fire events and can be applied in the task of near-real-time wildfire detection with 10-minute Himawari-8/9 satellite data. The data and codes used in the study are detailed in: https://github.com/eagle-void/MSSTF.

1. Introduction

As one of the most serious extreme natural disasters, wildfires occur frequently all over the world, posing significant threats to agricultural production, air pollution, climate change, and human life and property safety (Rogers et al., 2015; Jethva et al., 2019; Guo et al., 2020; Zhang et al., 2023c). To minimize the losses caused by wildfires, the rapid and accurate detection of fire events across large ground areas holds significant importance.

Satellite remote sensing has long been recognized for its macroscopic observation capabilities, which is a pivotal technique for wildfire detection (Xiao et al., 2016; Xiao et al., 2019; Hong et al., 2022; Yang et al., 2024). Among various types of satellites, geostationary satellites stand out for their high temporal resolution and stable capability of large-scale monitoring. Himawari-8/9 geostationary satellites are the third generation geostationary designed for the monitoring of weather and disasters, with significant improvements in frequency and resolution. Specifically, Himawari-8/9 satellites perform a high temporal resolution of up to 10 min, spanning East Asia and Southeast Asia region (Wickramasinghe et al., 2016; Liu et al., 2018). Moreover, the onboard Advanced Himawari Imagers (AHI) sensor provides abundant spectral information across 16 bands. Given the high temporal resolution and multi-spectral imaging, Himawari-8/9 satellites are the primary data source supporting high-frequency and detailed observation of wildfire events.

The dynamics of wildfires are influenced by the complex interactions

* Corresponding author.
E-mail address: yuelw@cug.edu.cn (L. Yue).

https://doi.org/10.1016/j.jag.2025.104416
Received 8 October 2024; Received in revised form 2 February 2025; Accepted 8 February 2025
Available online 15 February 2025

of physical factors, such as temperature, humidity, velocity and direction of wind, fuels and so on. As a result, the growth of wildfires is generally characterized by heterogeneous changes in spatial, temporal and spectral dimensions. Based on these facts, researchers have been devoted to propose abundant algorithms for wildfire detection based on remote sensing data (Xiao et al., 2019; Yang et al., 2024). Early attempts included the traditional threshold-based methods based on multi-temporal information (Filizzola et al., 2016; Hally et al., 2018, 2019; Yan et al., 2020) or spatial contextual information (Giglio et al., 2003; Liew, 2019; Parto et al., 2020; Wooster et al., 2012; Zhang et al., 2023b). Multi-temporal information threshold-based methods leverage the brightness temperature (BT) difference from different observation moments to detect wildfires (Laneve et al., 2006; van den Bergh et al., 2009; Xie et al., 2018; Yan et al., 2020). Considering the spatial information, contextual algorithms aim to find differences in context information, such as thermal characteristics and albedo, between active fire and its surrounding background pixels (Wooster et al., 2012; Chen et al., 2017; Liew et al., 2019). In the general cases, these methods also need the thresholds for spatial contrast information based on local statistics (e.g., average and standard deviation, and variance values) (Wickramasinghe et al., 2016; Xu and Zhong, 2017; Yan et al., 2020). Researchers also explored using adaptive-size windows around candidate fire pixel to improve the model's generalization performance in different fire events (Giglio et al., 2016; Chen et al., 2022). The threshold-based methods are easy to be implemented and are widely used for wildfire detection with Himawari-8/9 data (Chen et al., 2021). However, it is always challenging to select appropriate thresholds to adapt to different wildfire cases with complex dynamic variations and spatial heterogeneity.

To enhance the accuracy and robustness of wildfire detection, deep learning has been introduced in wildfire detection as a robust spatial–temporal modelling tool (Barmpoutis et al., 2020). Until recently, deep neural networks, e.g., Convolutional Neural Networks (CNNs) (Krizhevsky et al., 2012), Recurrent Neural Networks (RNNs) (Elman, 1990), and Transformers (Ashish, 2017), have been employed in wildfire detection tasks (Majid et al., 2022; Zheng et al., 2024). These works demonstrate that effective learning of temporal and spatial information from remote sensing imagery is the key aspect of robust detection of wildfire. Over the past few years, the development of deep learning models has demonstrated significant potential in feature fusion for wildfire detection (Barmpoutis et al., 2020; Gong et al., 2021; Majid et al., 2022; Zhao et al., 2023; Zheng et al., 2024). Accordingly, researchers explored learning of multi-dimensional features from geostationary satellite data to enhance detection performance. For example, Phan et al. (2020) and Hong et al., (2024) integrated spatial and temporal features to detect wildfire in satellite images using deep neural networks. Furthermore, Kang et al. (2022) combined CNNs and random forest (RF) models to integrate multiple feature groups (i.e., spatial, temporal and spectral) for identifying fire pixels. Zhang et al. (2023c) proposed a prediction model based on RNNs to learn the spatial–temporal-spectral representations from Himawari-8 geostationary data, and achieved good results in study cases. These studies have preliminarily demonstrated that the integration of multi-dimensional information is effective in enhancing fire detection accuracy. However, existing deep learning models generally neglect the full integration of spatial–temporal relationships across various scales, which is crucial for adapting to wildfires that exhibit diverse scale characteristics and growth patterns. It is then interesting to fully explore multi-scale spatial and temporal features to achieve robust detection results in various wildfire cases.

To address the above issues, we design a novel deep learning framework integrating multi-scale spatial–temporal features for wildfire detection. The major contributions can be summarized as:

1) We propose a deep learning framework for near-real-time wildfire detection with Himawari-8/9 geostationary satellite data by integrating multi-scale spatial–temporal features to delineate the complex characteristics of wildfire events.
2) For highlighting the spatial differences of wildfire, we use a multi-kernel attention-based convolution (MKAC) module for extracting spatial features representing the difference of fire and non-fire pixels within multi-scale receptive fields.
3) To fully utilize the high-resolution temporal information contained in Himawari-8/9 geostationary satellite data, we design a long short-term Transformer (LSTT) module for learning multi-scale temporal features from the image sequences with different window lengths.

2. Study area and dataset

2.1. Study area

To cover a wider area, six representative forest fire events from the southwestern region of China are analyzed. Southwest China, known for its high forest coverage rate and rich ecosystem diversity, ranges from tropical rainforests to alpine forests. The influence of the monsoon climate results in scant rainfall during the dry season, contributing to a high incidence of wildfires. Furthermore, the complex terrain of the study area poses challenges for efficient fire response and rescue operations. Detailed information about the six fire events is presented in Fig. 1.

2.2. Himawari-8/9 satellite AHI data

Himawari-8/9 satellite is designed to provide near-real-time earth observations every 10 min. With a high temporal resolution, the Himawari-8/9 data can support applications in meteorology, environmental monitoring, and natural disaster detection (Bessho et al., 2016). The satellite's coverage includes East Asia and Australia, ranging from 60°S to 60°N and 80°E to 160°W. The onboard AHI sensor is capable of capturing data across 16 spectral bands, comprising 3 visible bands, 3 near-infrared (NIR) bands, and 10 infrared bands. The detailed information about the bands of Himawari-8/9 AHI sensor used is presented in Table 1.

In this study, we employ Himawari-8/9 AHI data as the main data source. The Japan Meteorological Agency (JMA) offers two data formats for Himawari-8/9 AHI data, i.e., HSD and NetCDF4. We utilize the 10-minute full disk NetCDF4 data in the tests. Given the thermal sensitivity difference of the MIR and LWIR bands (Dozier, 1981; Sullivan et al., 2003; Kaufman et al., 1998; Dennison et al., 2006), the BT_{07} and BT_{14} data from the Himawari-8/9 AHI dataset serve as the primary inputs for fire detection.

To address missing data due to the house-keeping of the Himawai-8 and −9 satellites (JAXA Himawari Monitor (P-Tree System), 2024) or major errors in the processing stage, we use the average value between the data achieved before and after the missing time to obtain seamless time series data within the time period for model training and testing. To deal with the cloud pixels, we employ a threshold-based algorithm to create a cloud mask (Zhang et al., 2023a). The cloud pixels are then excluded from the subsequent fire detection process, which aims to mitigate the false alarms caused by clouds. During the daytime, the cloud pixels are defined as follows:

$$\begin{aligned}&\mathrm{BT}_{15} < 265\mathrm{K} \text{ and } \mathrm{R}_3 + \mathrm{R}_4 > 1.2 \text{ or}\\&((\mathrm{R}_3 + \mathrm{R}_4 > 0.7) \text{ and } (\mathrm{BT}_{15} < 285\mathrm{K}))\end{aligned} \tag{1}$$

For the night pixels, the cloud pixels are defined as follows:

$$\mathrm{BT}_{15} < 265\mathrm{K} \text{ and } \mathrm{BT}_{07} < 285\mathrm{K} \tag{2}$$

where R_3 and R_4 are reflectivity data of band 3 and band 4 of Himawari-8/9 satellite, and BT_{07} and BT_{15} is brightness temperature data of band 7 and band 15 of Himawari-8/9 satellite.

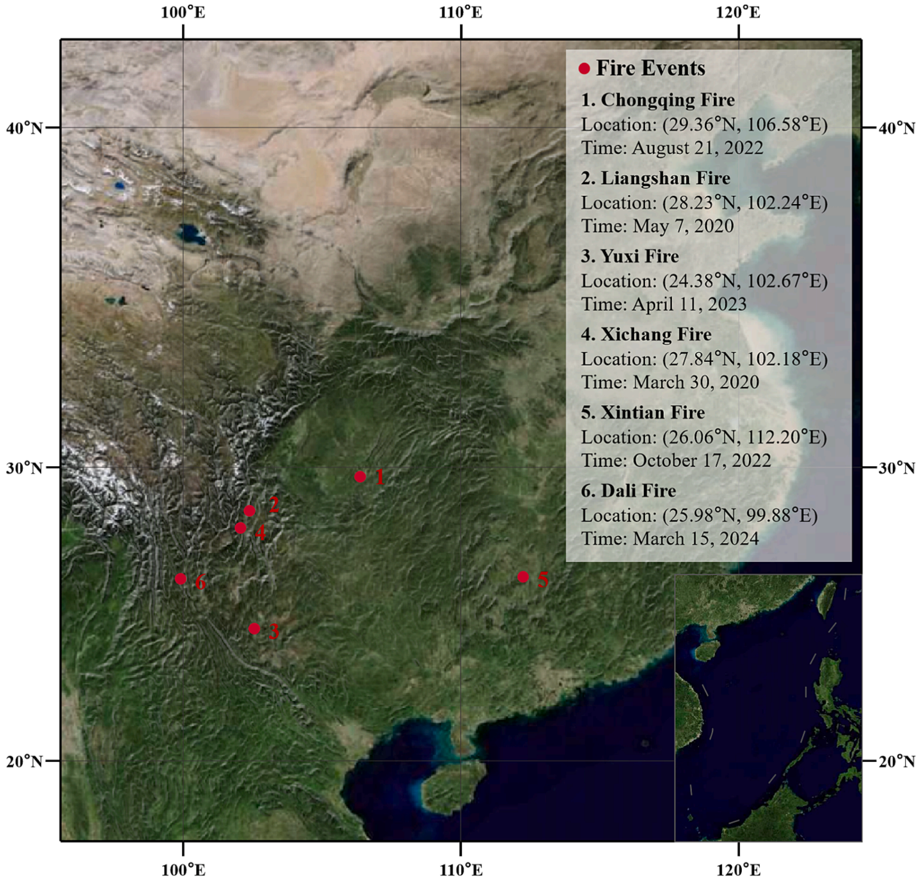


Fig. 1. Information about the study area.

Table 1
Information about the bands of Himawari-8/9 AHI sensor used in this study.

Band	Band type	Center wavelength	Spatial resolution	Sensitive objects
3	Reflectivity	0.64 μm	0.5 km	Land, clouds, etc.
4		0.86 μm	1 km	Marine water color, phytoplankton, etc.
7	Brightness temperature	3.9 μm	2 km	Surface temperature, cloud top temperature
14		11.2 μm	2 km	Surface temperature, cloud top temperature
15		12.3 μm	2 km	Surface temperature, cloud top temperature

2.3. *Experimental dataset*

Current widely utilized active fire detection datasets, such as MCD14ML and MCD64A1 of MODIS, and VNP14DL of VIIRS, fall short of the high temporal and spatial resolution demands for active fire detection and monitoring (Wooster et al., 2021). Moreover, the existing satellite products often ignore small fires, leading to the omission of active fire events (Jones et al., 2022). For example, the JAXA Wild Fire (WLF) L2 products of Himawari-8/9 satellite use statistics of time series of BT difference between BT_{07} and BT_{14} to classify fire pixels. As shown in Fig. 2, JAXA WLF L2 products omit numerous true fire pixels. Therefore, we construct a manually annotated fire dataset as the ground truth for validation, to ensure a comprehensive assessment of the detection results. A visual example of the manual fire labels for Xichang fire is shown in Fig. 2.

Given the limited data amount of the manual samples, we further created a simulated training dataset for model training. For the generation of simulated data patches, we randomly selected N pixels as the fire seeds, and randomly label four to eight pixels within the 3 × 3 local neighbor of each seed pixel as fire pixels. Labeling the nearest pixels is aimed to simulate the actual fire diffusion, adhering to the physical characteristics of fire spread in the spatial dimension. After determining the location of fire pixels, we assign a random ignition time to each fire pixels, in order to extend the process of fire pixel marking to the temporal dimension. Specifically, for labeled k fire pixels $(\mathrm{x_i},\mathrm{y_i}),\ \mathrm{i}\in\{1,2\cdots,\mathrm{k}\}$ and their ignition time $t_i,\ i\in\{1,\ 2,\ \cdots,\ k\}$, the rules for assigning characteristic value to them are as follows:

$$\mathrm{BT_{dif}}(\mathrm{t},\mathrm{x_i},\mathrm{y_i}) = \mathrm{BT_{dif_{fire}}},\ \ \mathrm{t_i} \le \mathrm{t} \le \mathrm{T},\ \ \mathrm{i} \in \{1,2,\cdots,\mathrm{k}\} \tag{3}$$

where T represents the length of the temporal dimension and BT_{dif_fire} represent the characteristic value of fire pixels based on statistical features. For other unlabeled pixels, we regard them as non-fire pixels and assign $BT_{dif\_non\_fire}$ as their characteristic values. The feature values for fire and non-fire pixels are determined with the statistics from the manually true labels. Our statistic analysis encompasses a broad range of fire events, extending beyond those mentioned in this paper, and leverages the extensive time series data from Himawari satellite observations, which are free from fire occurrences. In this study, we utilize the $BT_{07} - BT_{14}$ as the input feature for model training and choosing the feature values, and the value ranges of BT_{dif_fire} and $BT_{dif\_non\_fire}$ are set as $[5,\ 35]$ and $[0,\ 2]$, respectively. Therefore, we generate a customizable number of data with the size of $(\mathrm{T},1,\mathrm{H},\mathrm{W})$ for training the detection model, where T represents the length of the temporal dimension. The generation process of the virtual dataset is shown in Fig. 3.

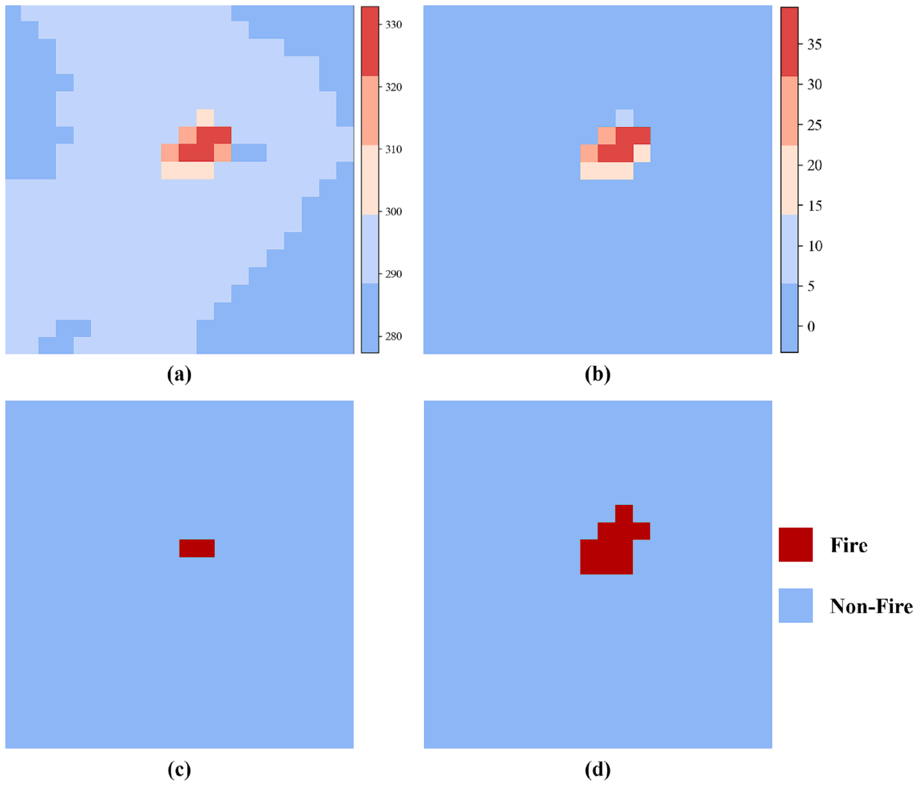


Fig. 2. Visualization of sub-region BT values and the corresponding fire labels in Xichang fire at 13: 20 (UTC time) on March 20, 2020: (a)BT_{07} data of Himawari-8 AHI; (b) Difference between BT_{07} and BT_{14} data of Himawari-8 AHI; (c) Fire labels of JAXA WLF L2 products; (d) Manually labeled dataset.

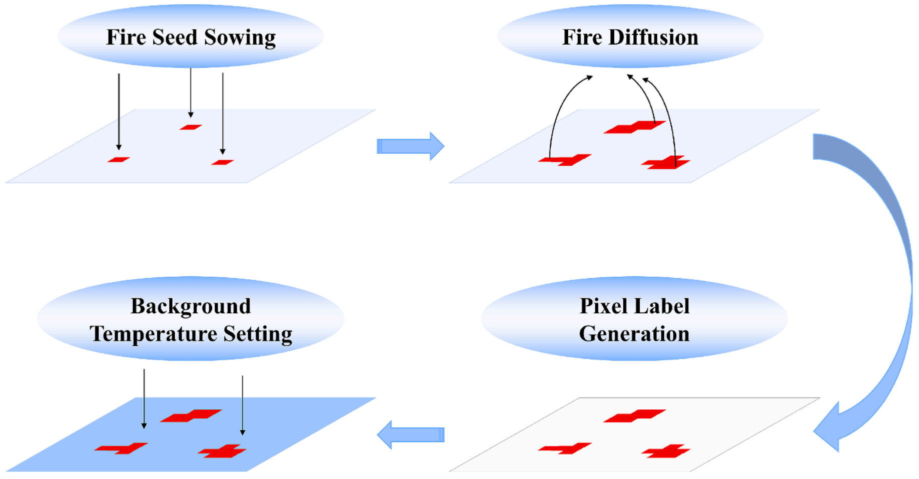


Fig. 3. The generation process of the virtual dataset for model training.

3. Methodology

3.1. Overview

As discussed previously, forest fires are characterized by complex temporal dynamics and spatial heterogeneity. In the light of this, we propose a multi-scale spatial–temporal feature fusion (MSSTF) model for near-real-time wildfire detection with Himawari-8/9 AHI data, as shown in Fig. 4. The essence of the MSSTF model is the fusion of multi-scale spatial–temporal features from BT differences for wildfire detection.

The MSSTF model receives the normalized difference between BT_{07} and BT_{14} of Himawari data as temporal series input. Specifically, the input features are successively processed with the spatial and temporal feature extraction modules. For spatial feature learning, multi-kernel attention-based convolution (MKAC) is employed to capture the spatial difference features within multi-level spatial receptive fields. To deal with the complex temporal changes of wildfire events, the multi-scale temporal features are learned using a Transformer-based module with long and short-term Himawari-8/9 image sequences as the input, respectively. Finally, the spatial and temporal features are integrated across scales to achieve the task of wildfire detection.

As shown in Fig. 4, there are three distinct branches for spatial–temporal feature fusion, which include: 1) local attention-based convolution combined with long-term temporal feature extraction, focusing on temporal modelling of fire dynamics; 2) further use of local attention-based convolution to enhance local spatial difference features

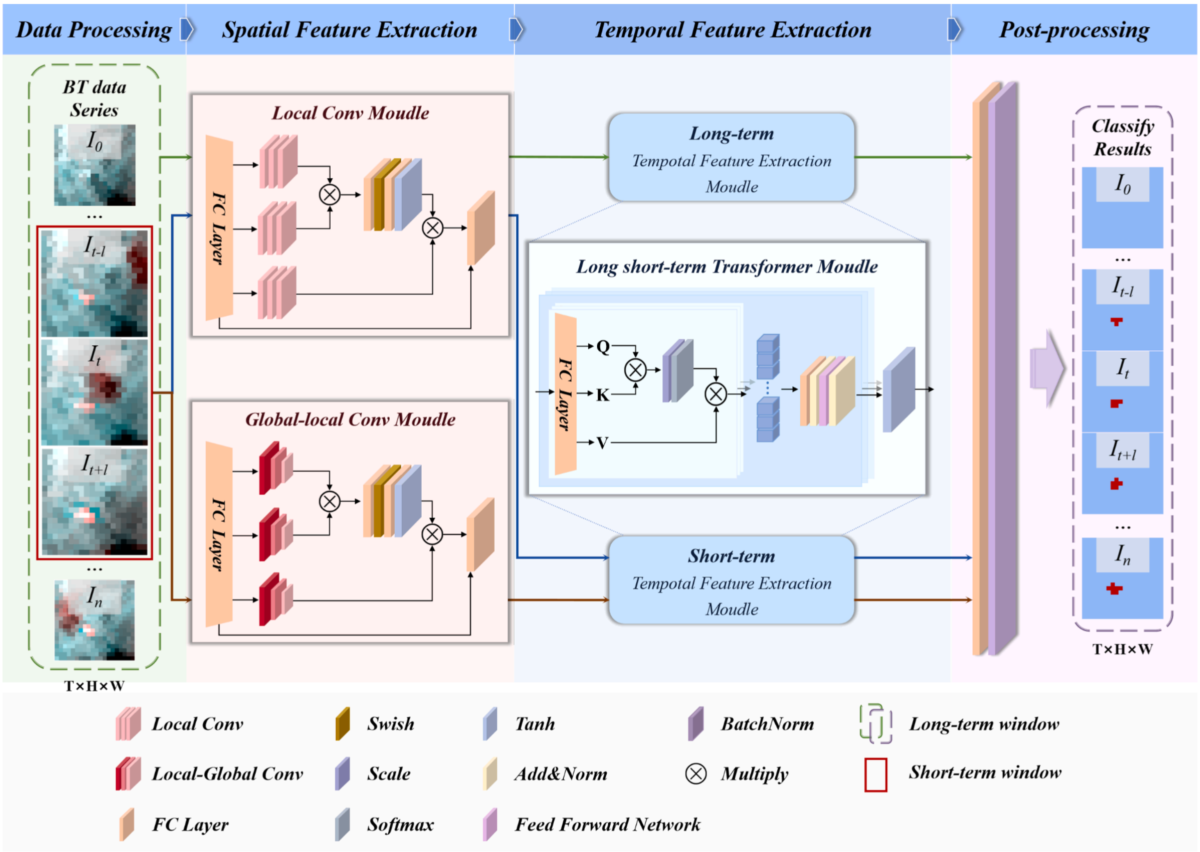
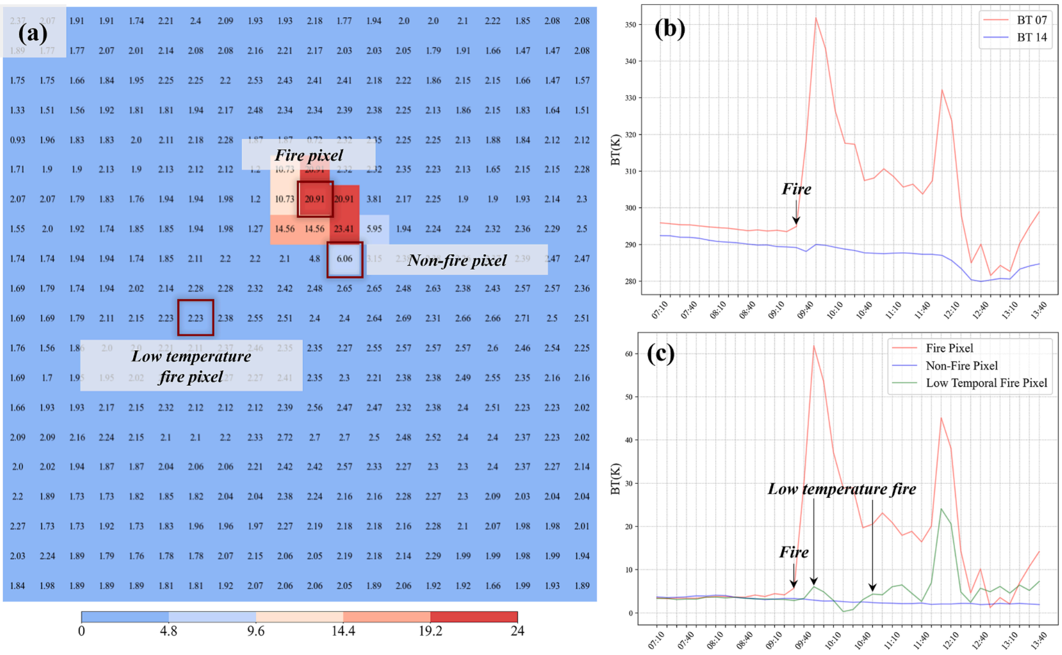


Fig. 4. The structure of the proposed MSSTF Framework.

in short-term sequences, utilizing the relationships among spatially and temporally neighboring pixels; 3) multi-kernel attention-based convolution integrated with short-term temporal feature learning, adapting to fire areas of varying sizes. The multi-stream information is merged, and subsequently processed with a fully-connected layer and batch normalization. Finally, the binary classification map can be generated after the output features are processed with an activation function.



2.37	2.07	1.91	1.91	1.74	2.21	2.4	2.09	1.93	1.93	2.18	1.77	1.94	2.0	2.0	2.1	2.22	1.85	2.08	2.08
1.89	1.77	1.77	2.07	2.01	2.14	2.08	2.08	2.16	2.21	2.17	2.03	2.03	2.05	1.79	1.91	1.66	1.47	1.47	2.08
1.75	1.75	1.66	1.84	1.95	2.25	2.25	2.2	2.53	2.43	2.41	2.41	2.18	2.22	1.86	2.15	2.15	1.66	1.47	1.57
1.33	1.51	1.56	1.92	1.81	1.81	1.94	2.17	2.48	2.34	2.34	2.39	2.38	2.25	2.13	1.86	2.15	1.83	1.64	1.51
0.93	1.96	1.83	1.83	2.0	2.11	2.18	2.28	1.87	1.87	0.72	2.32	2.35	2.25	2.25	2.13	1.88	1.84	2.12	2.12
1.71	1.9	1.9	2.13	1.9	2.13	2.12	2.12	1.2	10.73	20.91	2.32	2.32	2.35	2.23	2.13	1.65	2.15	2.15	2.28
2.07	2.07	1.79	1.83	1.76	1.94	1.94	1.98	1.2	10.73	20.91	20.91	3.81	2.17	2.25	1.9	1.9	1.93	2.14	2.3
1.55	2.0	1.92	1.74	1.85	1.85	1.94	1.98	1.27	14.56	14.56	23.41	5.95	1.94	2.24	2.24	2.32	2.36	2.29	2.5
1.74	1.74	1.94	1.94	1.74	1.85	2.11	2.2	2.2	2.1	4.8	6.06	3.15	[illegible]	[illegible]	[illegible]	[illegible]	2.39	2.47	2.47
1.69	1.79	1.74	1.94	2.02	2.14	2.28	2.28	2.32	2.42	2.48	2.65	2.65	2.48	2.63	2.38	2.43	2.57	2.57	2.36
1.69	1.69	1.79	2.11	2.15	2.23	2.23	2.38	2.55	2.51	2.4	2.4	2.64	2.69	2.31	2.66	2.66	2.71	2.5	2.51
1.76	1.56	1.86	2.0	2.0	2.21	2.11	2.37	2.46	2.35	2.35	2.27	2.55	2.57	2.57	2.57	2.6	2.46	2.54	2.25
1.69	1.7	1.95	1.95	2.02	[illegible]	[illegible]	[illegible]	2.27	2.41	2.35	2.3	2.21	2.38	2.38	2.49	2.55	2.35	2.16	2.16
1.66	1.93	1.93	2.17	2.15	2.32	2.12	2.12	2.12	2.39	2.56	2.47	2.47	2.32	2.38	2.4	2.51	2.23	2.23	2.02
2.09	2.09	2.16	2.24	2.15	2.1	2.1	2.2	2.33	2.72	2.7	2.7	2.5	2.48	2.52	2.4	2.4	2.37	2.23	2.02
2.0	2.02	1.94	1.87	1.87	2.04	2.06	2.06	2.21	2.42	2.42	2.57	2.33	2.27	2.3	2.3	2.4	2.37	2.27	2.14
2.2	1.89	1.73	1.73	1.82	1.85	1.82	2.04	2.04	2.38	2.24	2.16	2.16	2.28	2.27	2.3	2.09	2.03	2.04	2.04
2.27	1.73	1.73	1.92	1.73	1.83	1.96	1.96	1.97	2.27	2.19	2.18	2.18	2.16	2.28	2.1	2.07	1.98	1.98	2.01
2.03	2.24	1.89	1.79	1.76	1.78	1.78	2.07	2.15	2.06	2.05	2.19	2.18	2.14	2.29	1.99	1.99	1.98	1.99	1.94
1.84	1.98	1.89	1.89	1.89	1.81	1.81	1.92	2.07	2.06	2.06	2.05	1.89	2.06	1.92	1.92	1.66	1.99	1.93	1.89

Fig. 5. Visualization of time-series BT curves for fire and non-fire pixels in Xintian fire on October 17, 2022: (a) Difference of Himawari-8/9 AHI data between BT_{07} and BT_{14} at 11: 10 (UTC time); (b) Time-series BT_{07} and BT_{14} curves for the highlighted fire pixel in (a); (c) BT difference curves of BT_{07} and BT_{14} for the fire pixel, the non-fire pixel and the low-temperature fire pixels highlighted in (a).

3.2. Brightness temperature difference in spectral dimension

The Himawari-8/9 satellite can capture both the MIR (BT_{07}, 3.9 μm) and LWIR (BT_{14}, 11.2 μm) data. As shown in Fig. 5 (b), BT_{07} band exhibits high sensitivity to thermal anomaly, while BT_{14} band records relatively stable measurements of surface temperature variations. Therefore, the difference between BT_{07} and BT_{14} can be a good indicator to distinguish the fire and non-fire pixels, as shown in Fig. 5 (c). Based on the fact, we use the difference feature of BT_{07} and BT_{14} for the modelling of fire pixels, which is expressed as:

$$BT^{n}_{dif}(x,y) = BT^{n}_{07}(x,y) - BT^{n}_{14}(x,y) \tag{4}$$

where n stands for the T_n moment in Himawari-8/9 AHI data. (x, y) represents the position of each pixel. $BT^{n}_{dif}(x,y)$ represents the spectral characteristics of the position of pixel (x, y) at the T_n moment.

3.3. Spatial feature extraction with multi-kernel attention-based convolution

The goal of spatial feature extraction is to learn the neighboring difference between BT^{n}_{dif} values of fire and non-fire pixels. With powerful ability to capture local spatial patterns, convolution process is highly effective for this task. However, the early-stage fire pixels exhibit weak characteristics of thermal anomalies in the AHI images, due to the relatively low temperature for the fire region at this stage and coarse data resolution. Moreover, the fire regions might expand to varying scales as time progresses. In those cases, it is difficult to effectively delineate the difference features between fire and non-fire pixels with a single-scale convolution.

Based on these facts, we propose to use a multi-kernel attention-based convolution (MKAC) for extracting spatial features representing the difference of fire and non-fire pixels within multi-scale receptive fields. The module integrates a context-aware attention mechanism (Fan et al., 2023) to enhance local features through context-aware weighting, while large and small kernel sizes provide flexible receptive scales to aggregate spatial information, as shown in Fig. 6.

In the MKAC module, we initiate the process by applying a linear transformation to the input feature tensor X. Then a simple depth-wise convolution (DWconv) is used for aggregating local spatial information to derive the Q_s (query), K_s (key), and V_s (value) matrices. Multi-kernel convolution processing is used in DWconv to extract multi-scale spatial information from the wildfire scenes. The process is expressed as follows:

$$Q_s, K_s, V_s = DWConv(FC(X)) \tag{5}$$

where F C is the fully-connected layer for linear transformation. After that, we make a series of processing on Q_s and K_s matrices to generate the context-aware weight, aiming to enhance the ability of the model to perceive the correlation between input data. Specifically, we multiply the Q_s matrices by the K_s matrices, and get the attention map through a series of nonlinear activation functions and fully connected layers. Ultimately, we multiply the attention map and V_s to apply the shared weight to enhance the self-attention of the input tensor X. In this way, the model will give higher weight to the high-value parts caused by thermal anomalies. The process is expressed as follows:

$$\begin{gathered} Attn_t = FC(Swish(FC(Q_s \otimes K_s))) \\ \mathrm{Attn} = \mathrm{Tanh}\left(\frac{\mathrm{Attn_t}}{\sqrt{\mathrm{d}}}\right) \\ X_{attn} = Attn \otimes V_s \end{gathered} \tag{6}$$

where d is the channel of the input data and $\otimes$ indicates the element-wise product. Swish is a special nonlinear activation function defined as $Swish(x) \ = x \bullet sigmoid(x)$, which can alleviate the problem of gradient disappearance in the process of back propagation. The attention-based convolution module uses a set of nonlinear transformations (e.g., Tanh and Swish) in the generation of context-aware weights. Therefore, stronger nonlinearity can be represented in the feature mapping process, and obtains the attention weights for delineation of the local fire char-

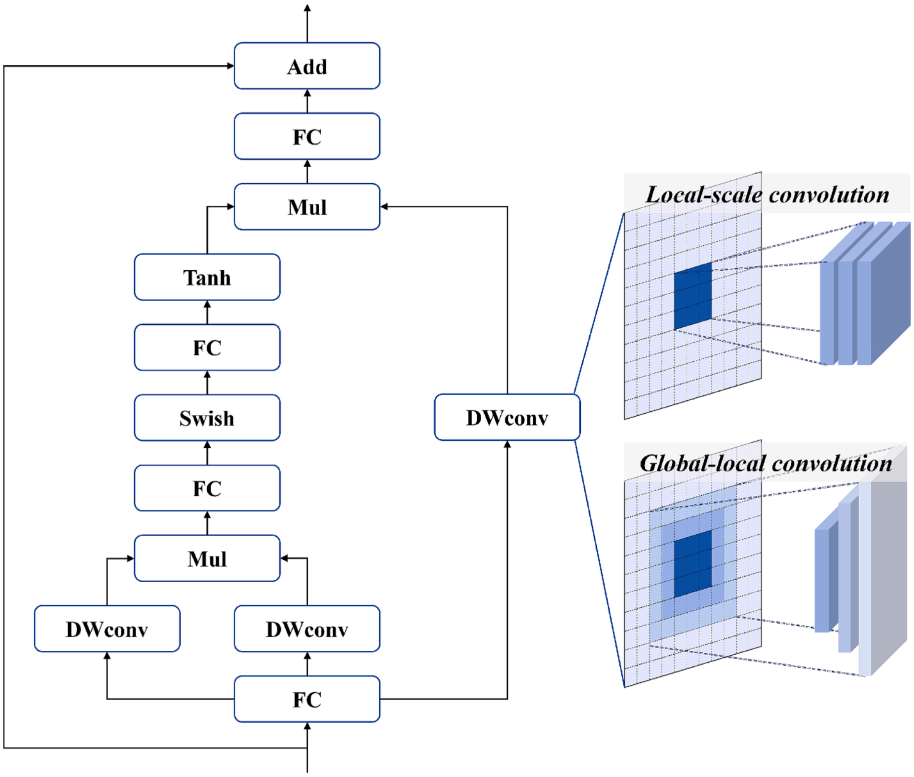


Fig. 6. Overall architecture of MKAC module.

acteristics. The final output is derived through a fully-connected layer, followed by a residual connection block. The process is expressed as follows:

$$X_s = X + FC(X_{attn}) \tag{7}$$

For providing flexible receptive scales to aggregate spatial information, we adopt two MKAC modules with different kernel sizes. One uses multiple groups of convolution kernels with different sizes to integrate global and local information, while the other emphasizes the role of local spatial information in detecting fire pixels through small-scale convolution kernels, as shown in Fig. 4.

3.4. Temporal feature extraction with long short-term Transformer

With a 10-minute temporal resolution, the Himawari-8/9 AHI data provide a wealth of temporal information as the auxiliary clues in addition to spatial variations to identify fire pixels. To fully use the temporal characteristics of wildfire pixels within different time periods, we segment the output features of MKAC module into long and short-term subsets along the temporal dimension to generate multi-scale data for temporal feature learning. The extraction of multi-scale temporal features is then achieved by employing a long short-term Transformer (LSTT) module, as shown in Fig. 7.

Similar to MKAC module, we apply a linear transformation to derive the Q_t, K_t, and V_t matrices and then calculate the attention scores:

$$\begin{gathered} Q_t, K_t, V_t = FC(X_s) \\ \mathrm{Attention}(\mathrm{Q_t}, \mathrm{K_t}, \mathrm{V_t}) = \mathrm{softmax}\left(\frac{\mathrm{Q_t} \bullet \mathrm{K_t^T}}{\sqrt{\mathrm{d}}}\right)\mathrm{V_t} \end{gathered} \tag{8}$$

where X_s is the output of MKAC module divided into long and short time windows, and d is the dimension of Q_t and K_t. The above operation is performed for each self-attention head, which refers to $h_i = Attention(Q_t, K_t, V_t)$. The outputs of multiple heads are integrated by a fully-connected layer to integrate multiple attention information, as depicted in Eq. (9):

$$\mathrm{H} = \mathrm{FC}\left(\begin{bmatrix} \mathrm{h_1} \\ \vdots \\ \mathrm{h_n} \end{bmatrix}\right) \tag{9}$$

where H stands for the output of fused multi-head attention. Then the skip connection and layer normalization are combined to preserve the high-frequency information, and highlight the temporal feature of fire pixels. The process can be expressed as:

$$\begin{gathered} X_{norm} = LayerNorm(X_s + H(X_s)) \\ X_t = LayerNorm(X_{norm} + FFN(X_{norm})) \end{gathered} \tag{10}$$

where FFN denotes feed-forward neural network, which is employed to enhance the expressive ability of temporal features through nonlinear transformation and mapping. At the end of each Transformer layer, a softmax layer is used to generate the layer output from X_t. By stacking multiple Transformer layers, we can obtain the final output of the LSTT module.

3.5. Loss term

As for loss function, we use BCEWithLogitsLoss to optimize the MSSTF model, which is a commonly used loss function in deep learning used for the binary classification tasks (Li et al., 2024). The BCEWithLogitsLoss function is described as:

$$\begin{gathered} \mathrm{Loss} = \frac{1}{\mathrm{N}}\sum\nolimits_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{y_i} \bullet \log(\sigma(\mathrm{p_i})) + (1 - \mathrm{y_i}) \bullet \log(1 - \sigma(\mathrm{p_i}))) \\ \sigma(\mathrm{p_i}) = \frac{1}{1 + e^{-p_i}} \end{gathered} \tag{11}$$

where $y_i \in \{0, 1\}$ represents fire labels of each pixel and $p_i \in [0, 1]$ represents the outputs of MSSTF model of each pixel. N is the total number of pixels. The model is then optimized with Adagrad algorithm (Duchi et al., 2011), which can optimize the learning rate adaptively to accelerate the training of the model.

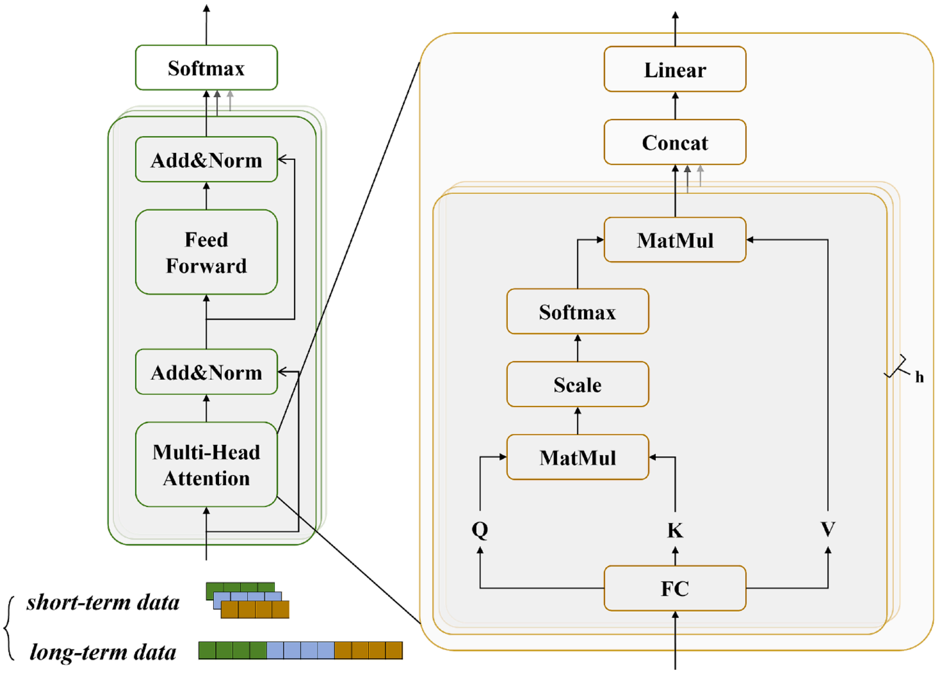


Fig. 7. Overall architecture of LSTT module.

4. Experiments and results

4.1. Implementation details and evaluation

In the experiments, the model setting is detailed as follows. The optimal settings of model hyperparameters (e.g., convolution kernel sizes and temporal windows size) were determined via empirical trials. For the MKAC module, we configure two sets of convolutional kernels of sizes [7, 5, 3] and [3, 3, 3]. The larger kernels are employed to capture global-scale information, while the smaller kernels focus more on fine-grained local-scale details. In the LSTT module, we establish a long-term window length of 72 and a short-term window length of 6, which is determined through the statistical analysis of the temporal curves of fire pixels. The detailed descriptions about the hyperparameters are presented in Table 2.

Four metrics are employed for the quantitative assessment of the detection results, including fire accuracy (FA), overall accuracy (OA), false alarm rate (FAR), and overall false alarm rate (OFR). The definitions are as follows:

$$\begin{aligned} FA &= \frac{TP}{TP + FN} \\ OA &= \frac{TP + TN}{TP + FN + FP + TN} \\ FAR &= \frac{FP}{TP + FP} \\ OFR &= \frac{FN}{TP + FN} \end{aligned} \tag{12}$$

where TP, FN, FP, and TN represent the number of true positive, false negative, false positive, and true negative predictions, respectively.

For comparing the detection accuracy of MSSTF model, we include JAXA WLF L2 Products (https://www.eorc.jaxa.jp/ptree/index.html) for comparison. Furthermore, given the high temporal resolution and low spatial resolution of Himawari-8/9 AHI data, the traditional CNN networks relying solely on spatial information fail to effectively delineate the dynamic features among the observation data. Therefore, the baseline models for comparison include Gated Recurrent Unit (GRU) (Dey and Salem, 2017), RNN (Elman, 1990), LSTM (Hochreiter, 1998), Transformer (Vaswani et al., 2017) and ConvLSTM (Shi et al., 2015). Among them, GRU, RNN, LSTM and Transformer focus to learn the temporal dependencies, while ConvLSTM integrate spatial and temporal feature learning with the combination of CNN and LSTM modules.

4.2. Quantitative evaluation results

In the tests, we conducted experiments using Himawari-8/9 AHI data over six study areas in Fig. 1 to test the method performance in detection of wildfire events with various scales. Table 3 presents a comparative analysis of the proposed model against JAXA WLF L2 products and the comparison models. The evaluation scores show that MSSTF model achieves a robust performance over different wildfire scenarios, in terms of the highest fire accuracy and lowest false positive rates.

Comparatively, JAXA's wildfire products show unstable FA scores in fire events with different spatial scales, which might be due to the low adaptability of the threshold-based method used by JAXA WLF products to various fire scenes. Moreover, serious omissions are observed in the results for GRU, RNN, and Transformer models in Chongqing fire, whose background non-fire pixels have relatively high BT values at the early stage of the fire. This indicates that it is difficult to accurately identify fire pixels only using temporal information. The blending of spatial and temporal features enables the ConvLSTM model to achieve a good performance in this task. However, the ConvLSTM only incorporates single-scale information, and omission errors and false alarms easily occur in fire cases with heterogeneous spatial and temporal dynamics. Overall, the proposed MSSTF model integrates multi-scale spatial–temporal information to capture the changing characteristics of wildfire across different scales and dimensions, and achieves an average FA of over 88 % and OA of over 99 % in various fire events.

Table 2
Parameter setting.

Section	Parameter	Value	Description
MKAC module	Kernel_sizes_small	[3, 3, 3]	Small-scale convolution kernel size
	Kernel_sizes_large	[7, 5, 3]	Large-scale convolution kernel size
LSTT module	Win_size_global	72	Global window size
	Win_size_local	6	Local window size
Train	Batch_train	4	Training data batch size
	Epoch	50	Iterations
	lr	0.04	Initial learning rate

4.3. Visual evaluation results

To visually evaluate the model performance, we first present the frequency maps of the detected fire pixels within a 24-hour period over different study areas in Fig. 8, which reflect the overall spatial and temporal patterns of the six fire events. The methods with considerable quantitative results in Section 4.2 (i.e., LSTM, ConvLSTM, and Transformers) are involved for visual comparison. The results shown in Fig. 8 indicate that the JAXA WLF product often tends to miss fire pixels and leads to high OFR. Without the proper integration of spatial features, both the LSTM and Transformers models are susceptible to significant omission and commission errors, especially in the Chongqing fire. Particularly, the LSTM model is limited in using a single-scale perception field along the temporal dimension and extracting evolving features within long time series, thus often leading to a large number of false alarms in non-fire areas. Moreover, ConvLSTM tends to overestimate the fire areas, probably because the single-scale receptive field in the ConvLSTM module fails to efficiently capture the growth of fires in spatial and temporal dimensions. Comparatively, MSSTF model achieves the most accurate fire mapping results across different scales and temporal dynamics. This supports that multi-scale spatial–temporal information fusion is beneficial for fire detection.

In Fig. 9, we further present the temporal detection results at the early-stage of fire events. The six fire events exhibit different patterns of spatial–temporal variations. For example, due to the high BT values of the background pixels in Chongqing fire, there is no significant difference in the BT values between fire and non-fire pixels. For Liangshan, Yuxi and Dali fire with relatively small scales, the spatial extent remains stable in the early stage, as shown in Fig. 9 (b) and (c). For Xichang and Xintian fire, the fire pixels are characterized with low BT values in the early stage. Generally, the MSSTF model has a robust performance in identifying the fire events with different scales and changing characteristics.

In Fig. 10, we perform a visual analysis on the detection results of Yuxi and Chongqing wildfire scenarios at different stages. The Yuxi fire exhibits a dramatic progression, expanding to over 15 pixels (i.e., approximately 60 km^2) in just 12 h. Conversely, the Chongqing fire shows minor changes in spatial scales, yet it has an overall high BT for the background pixels at the beginning of the fire. This brings difficulties in identifying the fire pixels based on their distinctive thermal properties. Regarding the detection results, JAXA's wildfire products tend to miss the fire pixels with relatively low BT differences. Comparatively, the proposed MSSTF model achieves good detection accuracy at various stages of the fire events, and can effectively adapt to fire scenes with complex backgrounds and temporal dynamics.

Table 3
Accuracy results of MSSTF and comparison methods in six study sites. For each study area, the bold values in the table represent the best model accuracy results, while the underlined values represent the second best model accuracy results.

		JAXA WLF L2 Products	Conv LSTM	GRU	RNN	LSTM	Transformer	MSSTF
Chongqing Fire	**FA**	0.7418	0.7295	0.3053	0.2541	0.4221	0.1885	**0.8238**
	OA	**0.9977**	0.9966	0.8929	0.9135	0.8454	0.9555	0.9967
	FAR	**0.0243**	0.1483	0.9751	0.9738	0.9767	0.9593	0.2055
	OFR	0.2582	0.2705	0.6947	0.7459	0.5779	0.8115	**0.1762**
Liangshan Fire	**FA**	0.0989	0.9219	0.8278	0.8038	0.8469	0.7656	**0.9282**
	OA	0.9902	0.9854	0.9388	0.9134	0.9061	0.9915	**0.9969**
	FAR	**0.0313**	0.5781	0.8682	0.9062	0.9092	0.4175	0.1872
	OFR	0.9011	0.0782	0.1722	0.1962	0.1531	0.2345	**0.0718**
Xichang Fire	**FA**	0.1318	0.8748	0.8193	0.7740	0.8840	0.8168	**0.8958**
	OA	0.9819	0.9833	0.9923	0.9932	0.9866	0.9950	**0.9933**
	FAR	0.0930	0.4382	0.1895	0.1161	0.3753	**0.0690**	0.1898
	OFR	0.8682	0.1252	0.1807	0.2261	0.1160	0.1832	**0.1042**
Xintian Fire	**FA**	0.4601	0.4093	0.7944	0.6542	**0.9140**	0.8617	0.8738
	OA	0.9949	0.9898	0.9516	0.9239	0.9178	0.9954	**0.9972**
	FAR	**0.0159**	0.5558	0.8629	0.9230	0.9056	0.2919	0.1664
	OFR	0.5399	0.5907	0.2056	0.3458	**0.0860**	0.1383	0.1262
Yuxi Fire	**FA**	0.6320	0.8902	0.8882	0.8493	**0.9132**	0.8713	0.8940
	OA	0.9932	0.9720	0.9928	0.9854	0.9831	**0.9971**	0.9931
	FAR	**0.0366**	0.6272	0.2521	0.4481	0.4917	0.0428	0.2406
	OFR	0.3680	0.1098	0.1118	0.1507	**0.0868**	0.1287	0.1060
Dali Fire	**FA**	0.3704	0.4706	0.3498	0.3100	0.4467	0.4229	**0.8792**
	OA	0.9930	0.9926	0.8583	0.8569	0.8082	0.9560	**0.9953**
	FAR	**0.0372**	0.2449	0.9724	0.9756	0.9744	0.8909	0.2597
	OFR	0.6296	0.5294	0.6502	0.6900	0.5533	0.5771	**0.1208**
Average	**FA**	0.4058	0.7161	0.6641	0.6076	0.7378	0.6545	**0.8825**
	OA	0.9918	0.9866	0.9378	0.9311	0.9079	0.9818	**0.9954**
	FAR	**0.0397**	0.4321	0.6867	0.7238	0.7722	0.4452	0.2082
	OFR	0.5942	0.2840	0.3359	0.3925	0.2622	0.3456	**0.1175**

4.4. Ablation study

We try to explore the influence of different modules on the detection results by adjusting the model components. As shown in Table 4 and Fig. 11, Models A to E represent the combination of different spatial–temporal feature extraction modules. Overall, the proposed model yields the optimal results with the combination of multi-scale spatial and temporal information. In terms of temporal learning, the longer time window can better reflect the fire characteristics observed by Himawari-8/9 AHI data. The FA value is enhanced by up to 7.5 % (i.e., from 0.8134 for model C to 0.8884 for model A) with the incorporation of long short-term information along the temporal dimension, while the FAR is decreased by 21.66 %. This indicates that the incorporation of long short-term information can improve the model's ability to detect fire pixels that are easy to miss and greatly reduce the false alarm interference caused by short-term fluctuations in BT values.

For spatial modelling, when long short-term temporal information is employed, the incorporation of global local-scale spatial information (model A) reduces the FAR by 10.19 % compared to using only a local-scale convolution kernel (model B). The visual analysis for Liangshan fire is generally consistent with the quantitative analysis. With the integration of multi-scale spatial–temporal information, the MSSTF model can effectively capture the spatial–temporal dynamic of each pixel and demonstrate high robustness across various fire scenes.

5. Discussion

Compared to MODIS and VIIRS data, Himawari data is characterized by its remarkable 10-minute temporal resolution. The short intervals provide rich temporal dependencies, which are beneficial for monitoring the near real-time dynamics of wildfires. However, the high temporal resolution also poses challenges to the effective modeling of temporal information. Based on these facts, we designed the LSTT module, which integrates the learning of temporal representations of fire characteristics within both short- and long-term window lengths. Moreover, the temporal features are fused with multi-scale spatial features for the discrimination between fire and non-fire pixels. Compared with existing solutions for fire detection that use single-scale information, the MSSTF model can effectively capture subtle changes in pixel brightness temperature and accurately identify early fire pixels. We observe that the proposed MSSTF model achieves promising results in wildfire detection. Compared with various commonly used deep learning models (e.g., LSTM, ConvLSTM, and Transformers), the FA scores of the proposed method show an improvement of 14.47 %-27.49 %. Moreover, the mapping results demonstrate that the MSSTF model can effectively illustrate the overall progress of fire events with different spatial scales and temporal changing patterns. To present the model performance under various conditions, more experimental results can be accessed on our GitHub repository (https://github.com/eagle-void/MSSTF).

Although Himawari-8/9 data boasts high temporal resolution, the spatial resolution in the infrared band is limited to 2 km. This limitation impedes the creation of wildfire maps with higher spatial resolution. To address this challenge, applying data fusion techniques for reconstructing data at higher spatial resolutions is a plausible approach. Furthermore, special attention should be paid to the impact of clouds on

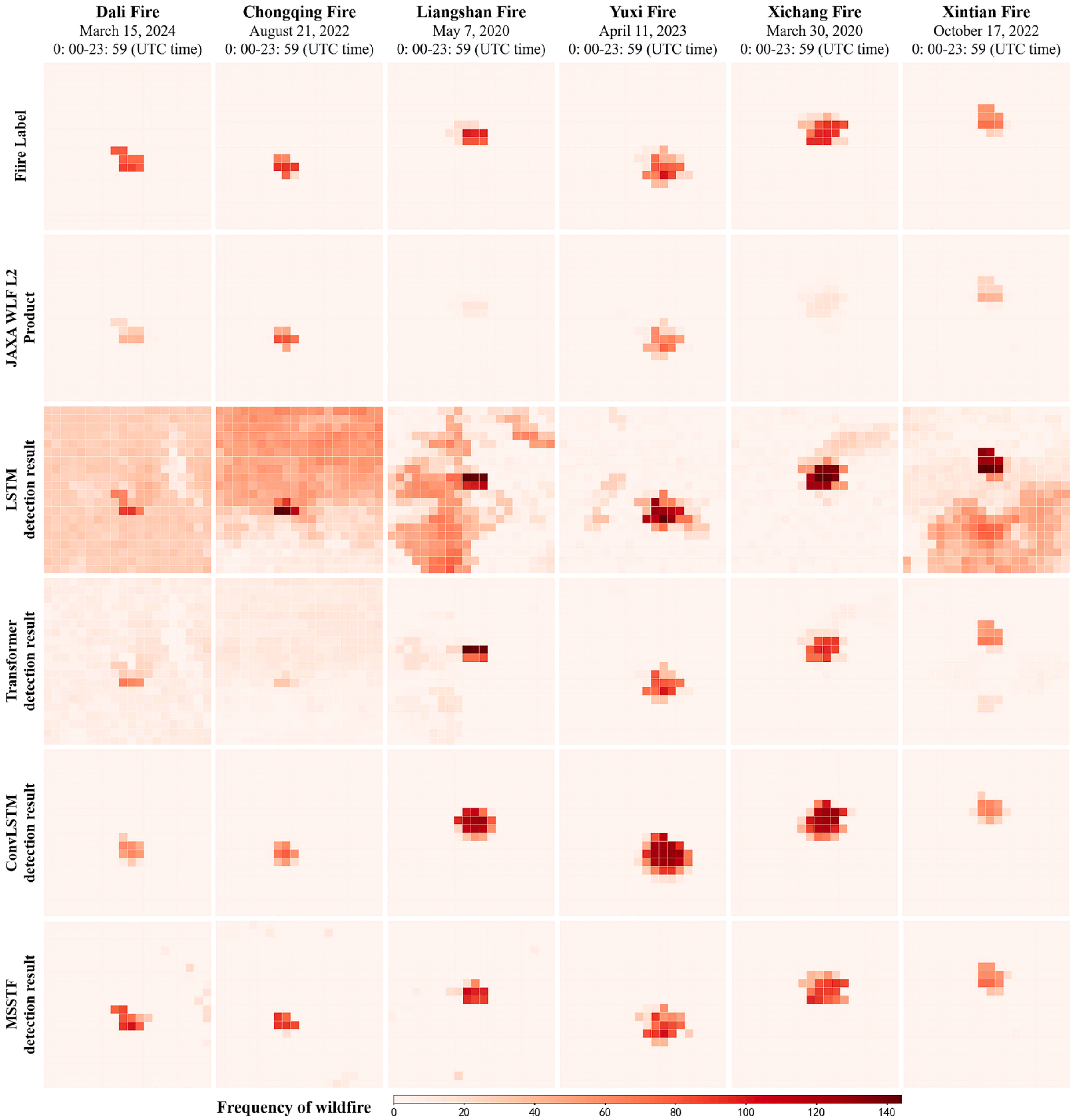


Fig. 8. Frequency map of the detected fire pixels in the six study areas.

model applications. In this work, cloud masks were employed to exclude cloud pixels from the fire detection process. High-precision cloud masks are crucial for reducing false alarm rates in wildfire detection. However, existing cloud mask algorithms based on Himawari data suffer from unstable accuracies across different regions, and the spatial resolution of cloud products provided by JAXA is limited to 5 km. The uncertainties of cloud masks makes it difficult to effectively eliminate the interference caused by clouds, resulting in potential false alarms in fire detection results.

In addition, the Himawari-8/9 AHI sensor captures data from 16 spectral bands within the wavelength range of 0.46–13.3 μm, and the data in these bands contain substantial information about interference factors such as fire or clouds. The distinct characteristics of BT differences between fire and non-fire pixels from both spatial and temporal perspectives are key aspects for effective wildfire detection. Therefore, we used the difference between BT_{07} and BT_{14} as the model input, which aligns with extensive existing work based on Himawari AHI data. However, we believe that enhancing model performance can be achieved by combining multiple feature indices and incorporating 'spectral' information. To accomplish this goal, the model network should be designed to fully integrate the high-dimensional information required for the detection task.

6. Conclusion

In this paper, we propose a novel deep learning model integrating multi-scale spatial–temporal features (MSSTF) to efficiently capture the dynamics of wildfires. The core idea of the proposed approach is to integrate both global and local-scale spatial information and long short-term temporal information for near-real-time wildfire detection based on Himawari-8/9 AHI data. To adapt to wildfire characteristics, the multi-kernel attention-based convolution module and long short-term Transformer module are carefully designed for learning the spatial and temporal patterns for complex wildfire scenes, respectively.

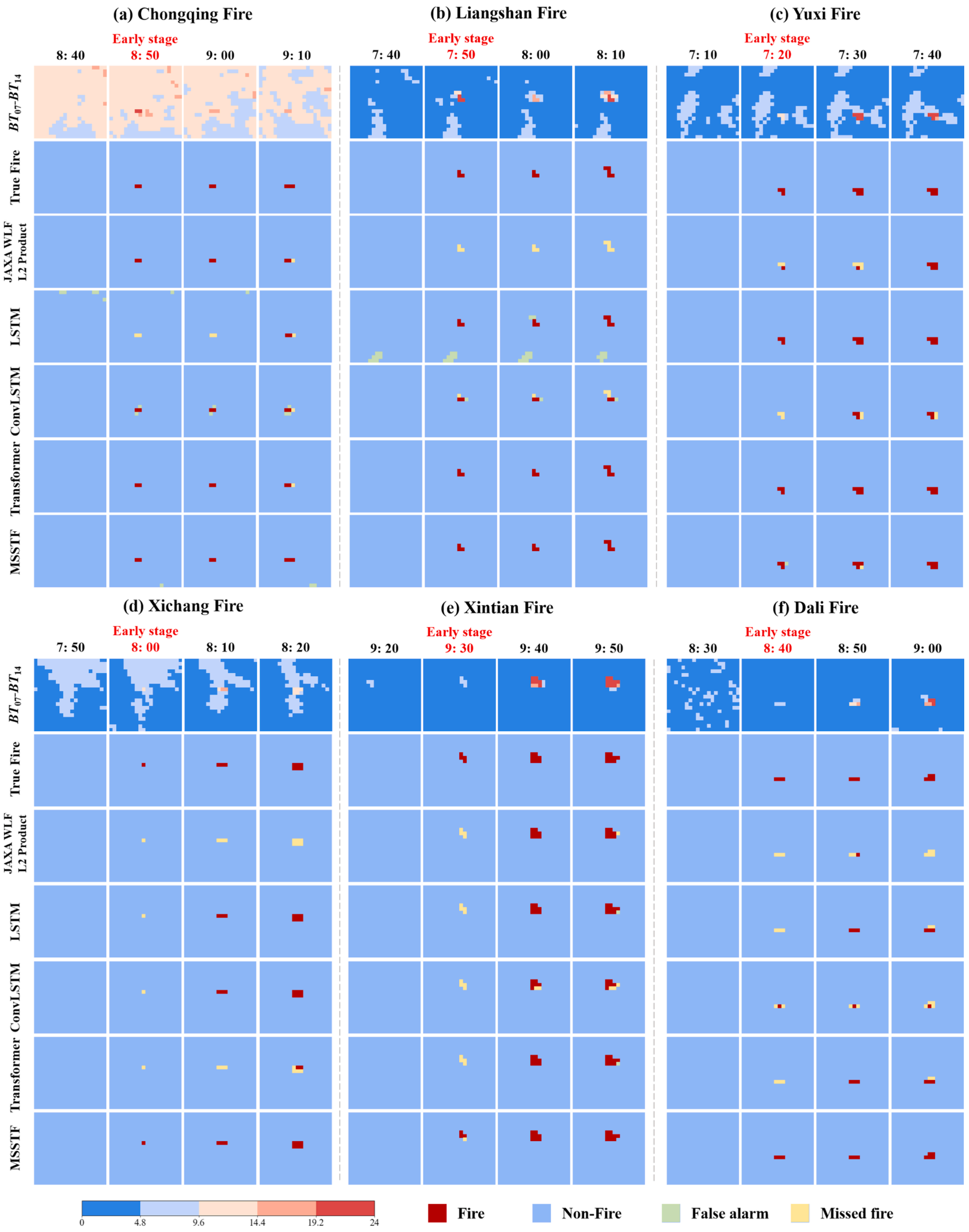


Fig. 9. Visual maps of the fire detection results obtained by MSSTF model in the early stage of the six events.

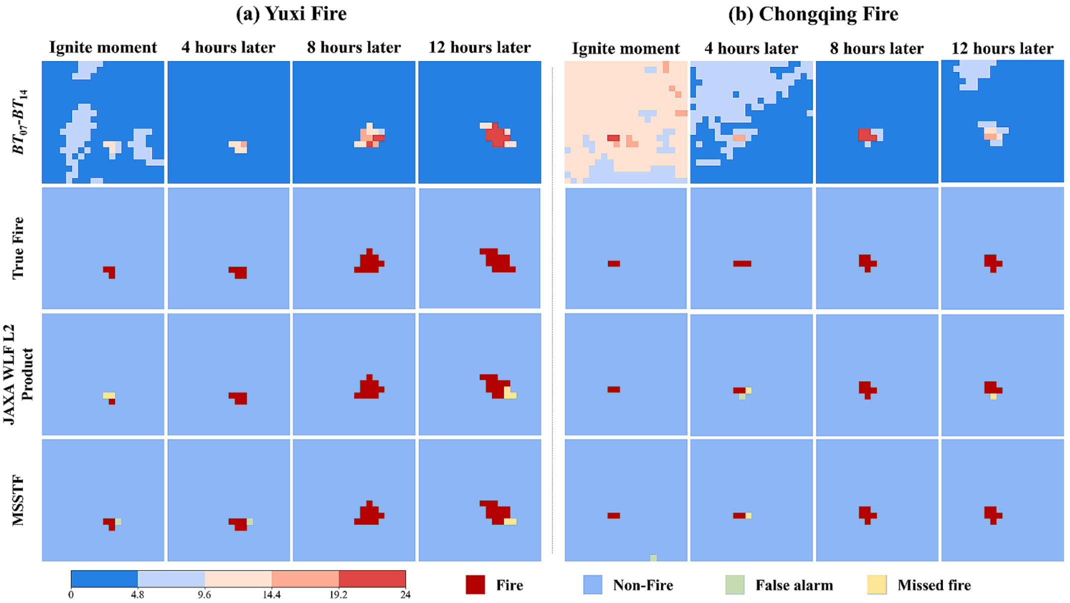


Fig. 10. Visual analysis of the detection results at different stages of Yuxi and Chongqing fires. Four moments are selected to analyze the progress of (a) Yuxi fire and (b) Chongqing fire.

Table 4
Quantitative analysis of the influence of different spatial and temporal modules on detection results in the Liangshan fire.

Model	Temporal module	Spatial module	FA	OA	FAR	OFR
A	**Long short-term**	**Global local-scale**	**0.8884**	**0.9979**	**0.0809**	**0.1116**
B	Long short-term	Local-scale	0.8764	0.9952	0.1828	0.1236
C	Long-term	Global local-scale	0.8134	0.9942	0.2975	0.1866
D	Long-term	Local-scale	0.8182	0.9937	0.3268	0.1819
E	Short-term	Global local-scale	0.7352	0.9802	0.6787	0.2648

Extensive tests have been conducted on various wildfire events to test the model performance. Compared with JAXA WLF L2 products and several widely used baseline models, the proposed MSSTF model has been verified to achieve superior performance and demonstrates robustness across fire events with complex dynamics. Particularly, the MSSTF model is able to capture early wildfires sensitively, and thus can offer early warnings for potential wildfires. To enhance the practicality of the proposed method for real-world applications, developing an unsupervised fire detection framework is one of the major targets for future work.

CRediT authorship contribution statement

Lizhi Zhang: Writing – original draft, Visualization, Validation, Software, Methodology, Conceptualization. **Qiang Zhang:** Conceptualization. **Qianqian Yang:** Conceptualization. **Linwei Yue:** Writing – review & editing, Conceptualization. **Jiang He:** Conceptualization.

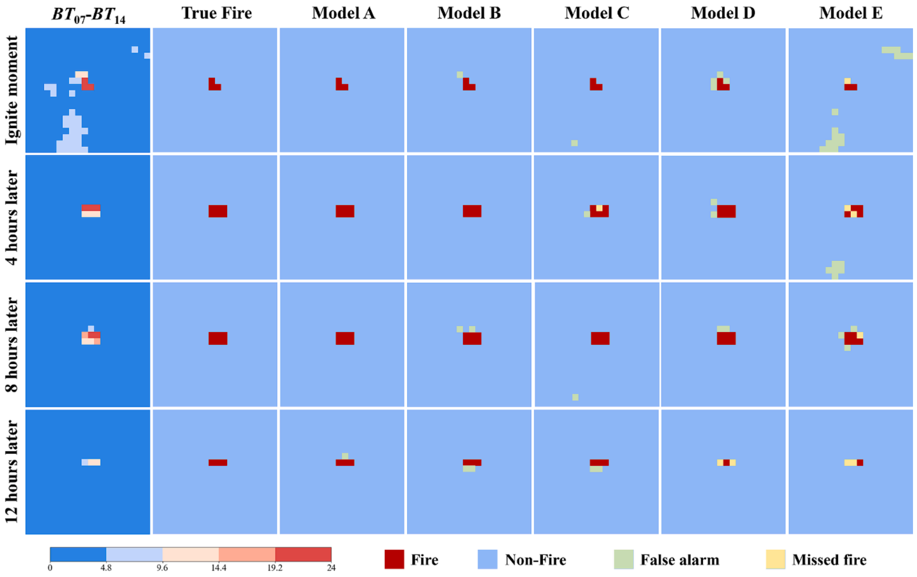


Fig. 11. Visual analysis of the influence of different spatial and temporal modules on detection results in the Liangshan fire.

Xianyu Jin: Conceptualization. **Qiangqiang Yuan:** Writing – review & editing, Software, Conceptualization.

Declaration of competing interest

The authors 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 was supported in part by the Science and Technology Program of Hubei Province [grant number 2023BCA003], and the National Natural Science Foundation of China [grant number 42230108, 42471414].

Data availability

The code package and partial datasets can be accessed via https://github.com/eagle-void/MSSTF.

References

Ashish, V., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30, I.

Barmpoutis, P., Papaioannou, P., Dimitropoulos, K., Grammalidis, N., 2020. A review on early forest fire detection systems using optical remote sensing. Sensors 20, 6442.

Bessho, K., Date, K., Hayashi, M., Ikeda, A., Imai, T., Inoue, H., Kumagai, Y., Miyakawa, T., Murata, H., Ohno, T., 2016. An introduction to Himawari-8/9—Japan's new-generation geostationary meteorological satellites. J. Meteorological Soc. Japan Ser. II 94, 151–183.

Chen, J., Zheng, W., Liu, C., 2017. Application of grassland fire monitoring based on Himawari-8 geostationary meteorological satellite data. J. Nat. Disasters 26, 197–204.

Chen, J., Zheng, W., Liu, C., Tang, S., 2021. Temporal sequence method for fire spot detection using Himawari-8 geostationary meteorological satellite. Nation. Remote Sens. Bull. 25, 2095–2102.

Chen, J., Zheng, W., Wu, S., Liu, C., Yan, H., 2022. Fire monitoring algorithm and its application on the geo-kompsat-2A geostationary meteorological satellite. Remote Sens. 14, 2655.

Dennison, P.E., Charoensiri, K., Roberts, D.A., Peterson, S.H., Green, R.O., 2006. Wildfire temperature and land cover modeling using hyperspectral data. Remote Sens. Environ. 100, 212–222.

Dey, R., Salem, F.M., 2017. Gate-variants of gated recurrent unit (GRU) neural networks,. In: 2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS). IEEE, pp. 1597–1600.

Dozier, J., 1981. A method for satellite identification of surface temperature fields of subpixel resolution. Remote Sens. Environ. 11, 221–229.

Duchi, J., Hazan, E., Singer, Y., 2011. Adaptive subgradient methods for online learning and stochastic optimization. J. Mach. Learn. Res. 12.

Elman, J.L., 1990. Finding structure in time. Cogn. Sci. 14, 179–211.

Fan, Q., Huang, H., Guan, J., He, R., 2023. Rethinking local perception in lightweight vision transformer. arXiv preprint arXiv:2303.17803.

Filizzola, C., Corrado, R., Marchese, F., Mazzeo, G., Paciello, R., Pergola, N., Tramutoli, V., 2016. RST-FIRES, an exportable algorithm for early-fire detection and monitoring: description, implementation, and field validation in the case of the MSG-SEVIRI sensor. Remote Sens. Environ. 186, 196–216.

Giglio, L., Descloitres, J., Justice, C.O., Kaufman, Y.J., 2003. An enhanced contextual fire detection algorithm for MODIS. Remote Sens. Environ. 87, 273–282.

Giglio, L., Schroeder, W., Justice, C.O., 2016. The collection 6 MODIS active fire detection algorithm and fire products. Remote Sens. Environ. 178, 31–41.

Gong, A., Li, J., Chen, Y., 2021. A spatio-temporal brightness temperature prediction method for forest fire detection with modis data: a case study in san diego. Remote Sens. 13, 2900.

Guo, L., Ma, Y., Tigabu, M., Guo, X., Zheng, W., Guo, F., 2020. Emission of atmospheric pollutants during forest fire in boreal region of China. Environ. Pollut. 264, 114709.

Hally, B., Wallace, L., Reinke, K., Jones, S., Engel, C., Skidmore, A., 2018. Estimating fire background temperature at a geostationary scale—an evaluation of contextual methods for AHI-8. Remote Sens. 10, 1368.

Hally, B., Wallace, L., Reinke, K., Jones, S., Skidmore, A., 2019. Advances in active fire detection using a multi-temporal method for next-generation geostationary satellite data. Int. J. Digit. Earth 12 (9), 1030–1045.

Hochreiter, S., 1998. The vanishing gradient problem during learning recurrent neural nets and problem solutions. Int. J. Uncertainty, Fuzziness Knowl.-Based Syst. 6, 107–116.

Hong, Z., Tang, Z., Pan, H., Zhang, Y., Zheng, Z., Zhou, R., Ma, Z., Zhang, Y., Han, Y., Wang, J., 2022. Active fire detection using a novel convolutional neural network based on Himawari-8 satellite images. Front. Environ. Sci. 10, 794028.

Hong, Z., Tang, Z., Pan, H., Zhang, Y., Zheng, Z., Zhou, R., et al., 2024. Near real-time monitoring of fire spots using a novel SBT-FireNet based on Himawari-8 satellite images. IEEE J. Select. Top. Appl. Earth Obs. Remote Sens. 17, 1719–1733.

JAXA Himawari Monitor (P-Tree System), 2024. User's Guide. https://www.eorc.jaxa.jp/ptree/userguide.html (accessed 22 December 2024).

Jethva, H., Torres, O., Field, R.D., Lyapustin, A., Gautam, R., Kayetha, V., 2019. Connecting crop productivity, residue fires, and air quality over northern India. Sci. Rep. 9, 16594.

Jones, M.W., Abatzoglou, J.T., Veraverbeke, S., Andela, N., Lasslop, G., Forkel, M., Smith, A.J., Burton, C., Betts, R.A., van der Werf, G.R., 2022. Global and regional trends and drivers of fire under climate change. Rev. Geophys. 60 e2020RG000726.

Kang, Y., Jang, E., Im, J., Kwon, C., 2022. A deep learning model using geostationary satellite data for forest fire detection with reduced detection latency. Gisci. Remote Sens. 59, 2019–2035.

Kaufman, Y.J., Justice, C.O., Flynn, L.P., Kendall, J.D., Prins, E.M., Giglio, L., Ward, D.E., Menzel, W.P., Setzer, A.W., 1998. Potential global fire monitoring from EOS-MODIS. J. Geophys. Res.-Atmos. 103, 32215–32238.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Adv. Neural Informat. Process. Syst. 25.

Laneve, G., Castronuovo, M.M., Cadau, E.G., 2006. Continuous monitoring of forest fires in the Mediterranean area using MSG. IEEE Trans. Geosci. Remote Sens. 44, 2761–2768.

Li, M., Qi, J., Tian, X., Guo, H., Liu, L., Fathollahi-Fard, A.M., Tian, G., 2024. Smartphone-based straw incorporation: an improved convolutional neural network. Comput. Electron. Agric. 221, 109010.

Liew, S.C., 2019. Detecting active fires with Himawari-8 geostationary satellite data. IEEE Int. Geosci. Remote Sens. Symp. 9984–9987.

Liu, X., He, B., Quan, X., Yebra, M., Qiu, S., Yin, C., Liao, Z., Zhang, H., 2018. Near real-time extracting wildfire spread rate from Himawari-8 satellite data. Remote Sens. 10, 1654.

Majid, S., Alenezi, F., Masood, S., Ahmad, M., Gündüz, E.S., Polat, K., 2022. Attention based CNN model for fire detection and localization in real-world images. Expert Syst. Appl. 189, 116114.

Parto, F., Saradjian, M., Homayouni, S., 2020. MODIS brightness temperature change-based forest fire monitoring. J. Indian Soc. Remote Sens. 48, 163–169.

Phan, T.C., Nguyen, T.T., Hoang, T.D., Nguyen, Q.V.H., Jo, J., 2020. Multi-scale bushfire detection from multi-modal streams of remote sensing data. IEEE Access 8, 228496–228513.

Rogers, B.M., Soja, A.J., Goulden, M.L., Randerson, J.T., 2015. Influence of tree species on continental differences in boreal fires and climate feedbacks. Nat. Geosci. 8, 228–234.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., Woo, W.-C., 2015. Convolutional LSTM network: a machine learning approach for precipitation nowcasting. Adv. Neural. Inf. Process Syst. 28.

Sullivan, A., Ellis, P., Knight, I., 2003. A review of radiant heat flux models used in bushfire applications. Int. J. Wildland Fire 12, 101–110.

van den Bergh, F., Udahemuka, G., van Wyk, B.J., 2009. Potential fire detection based on Kalman-driven change detection. Proc. IEEE Int. Geosci. Remote Sens. Symp. pp. IV-77-IV-80.

Wickramasinghe, C.H., Jones, S., Reinke, K., Wallace, L., 2016. Development of a multi-spatial resolution approach to the surveillance of active fire lines using Himawari-8. Remote Sens. 8, 932.

Wooster, M.J., Xu, W., Nightingale, T., 2012. Sentinel-3 SLSTR active fire detection and FRP product: pre-launch algorithm development and performance evaluation using MODIS and ASTER datasets. Remote Sens. Environ. 120, 236–254.

Wooster, M.J., Roberts, G.J., Giglio, L., Roy, D.P., Freeborn, P.H., Boschetti, L., Justice, C., Ichoku, C., Schroeder, W., Davies, D., 2021. Satellite remote sensing of active fires: history and current status, applications and future requirements. Remote Sens. Environ. 267, 112694.

Xiao, J., Chevallier, F., Gomez, C., Guanter, L., Hicke, J.A., Huete, A.R., Ichii, K., Ni, W., Pang, Y., Rahman, A.F., 2019. Remote sensing of the terrestrial carbon cycle: a review of advances over 50 years. Remote Sens. Environ. 233, 111383.

Xiao, Z., Liang, S., Wang, J., Xiang, Y., Zhao, X., Song, J., 2016. Long-time-series global land surface satellite leaf area index product derived from MODIS and AVHRR surface reflectance. IEEE Trans. Geosci. Remote Sens. 54, 5301–5318.

Xie, Z., Song, W., Ba, R., Li, X., Xia, L., 2018. A spatiotemporal contextual model for forest fire detection using Himawari-8 satellite data. Remote Sens. 10, 1992.

Xu, G., Zhong, X., 2017. Real-time wildfire detection and tracking in Australia using geostationary satellite: Himawari-8. Rem. Sens. Lett. 8, 1052–1061.

Yan, J., Qu, J., Ran, M., Zhang, F., 2020. Himawari-8 AHI fire detection in clear sky based on time-phase change. Nation. Remote Sens. Bull. 4 (5), 571–577.

Yang, S., Huang, Q., Yu, M., 2024. Advancements in remote sensing for active fire detection: a review of datasets and methods. Sci. Total Environ. 173273.

Zhang, L., Lu, C., Xu, H., Chen, A., Li, L., Zhou, G., 2023b. MMFNet: forest fire smoke detection using multiscale convergence coordinated pyramid network with mixed attention and fast-robust NMS. IEEE Internet Things J. 10, 18168–18180.

Zhang, H., Sun, L., Zheng, C., Ge, S., Chen, J., Li, J., 2023a. A weighted contextual active fire detection algorithm based on Himawari-8 data. Int. J. Remote Sens. 44 (7), 2400–2427.

Zhang, Q., Zhu, J., Huang, Y., Yuan, Q., Zhang, L., 2023c. 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, 103506.

Zhao, Y., Ban, Y., Sullivan, J., 2023. Tokenized time-series in satellite image segmentation with transformer network for active fire detection. IEEE Trans. Geosci. Remote Sens. 61, 1–13.

Zheng, S., Zou, X., Gao, P., Zhang, Q., Hu, F., Zhou, Y., Wu, Z., Wang, W., Chen, S., 2024. A forest fire recognition method based on modified deep CNN model. Forests 15, 111.