Predict Future Transient Fire Heat Release Rates Based on Fire Imagery and Deep Learning

Lei Xu, Jinyuan Dong * and Delei Zou

School of Civil Engineering, Dalian University for Nationalities, Dalian 116000, China; xulei@dlnu.edu.cn (L.X.)
* Correspondence: jinyuandong98@163.com

Abstract: The fire heat release rate (HRR) is a crucial parameter for describing the combustion process and its thermal effects. In recent years, some studies have employed fire scene images and deep learning algorithms to predict real-time fire HRR, which has led to the advancement of HRR prediction in terms of both lightweightness and real-time monitoring. Nevertheless, the development of an early-stage monitoring system for fires and the ability to predict future HRR based on current moment data represents a crucial foundation for evaluating the scale of indoor fires and enhancing the capacity to prevent and control such incidents. This paper proposes a deep learning model based on continuous fire scene images (containing both flame and smoke features) and their time-series information to predict the future transient fire HRR. The model (Att-BiLSTM) comprises three bi-directional long- and short-term memory (Bi-LSTM) layers and one attention layer. The model employs a bidirectional feature extraction approach, followed by the introduction of an attention mechanism to highlight the image features that have a critical impact on the prediction results. In this paper, a large-scale dataset is constructed by collecting 27,231 fire scene images with instantaneous HRR annotations from 40 different fire trials from the NIST database. The experimental results demonstrate that Att-BiLSTM is capable of effectively utilizing fire scene image features and temporal information to accurately predict future transient HRR, including those in high-brightness fire environments and complex fire source situations. The research presented in this paper offers novel insights and methodologies for fire monitoring and emergency response.

Keywords: fire scene images; temporal information; HRR prediction; NIST fire dataset; Att-BiLSTM

1. Introduction

The heat release rate (HRR) is defined as the amount of heat released by a combustion system per unit of time. It reflects the characteristics and risks of a fire and is an important indicator for assessing the danger level of fires. HRR is widely used in the safety design of building fires and firefighting operations [1]. In the laboratory, two common methods for measuring the HRR of a fire scene are the combustion rate method based on fuel mass loss [2] and the calorimetry method based on oxygen consumption [3]. However, these methods require expensive and complex equipment and are unable to predict the HRR at future moments. Consequently, the monitoring of the early stages of an actual fire and the prediction of future HRR based on current data, with the objective of judging the development scale of indoor fires and providing early warnings, has become one of the most pressing scientific problems in the field of fire research.

In numerous fire tests and actual fire scenes, closed-circuit television cameras and mobile device cameras are frequently utilized to obtain fire videos, record alterations in flames and smoke, and assess related fire parameters [4–6]. The extracted fire frame images from these videos contain data about the behavior and characteristics of the fire, including the size, color, brightness, and oscillation frequency of the flames and smoke, as well as their changes over time. A comprehensive analysis of fire scene images can yield crucial insights into the progression of a fire.

The field of artificial intelligence (AI) has witnessed a remarkable advancement in recent years, significantly enhancing the capabilities of image analysis. AI methods have been extensively utilized in diverse domains, including image recognition [7] and object detection [8]. Additionally, AI techniques have been employed to identify implicit information in fire images and predict the evolution of fires and smoke. For instance, Hodges et al. [9] employed a transposed convolutional neural network (TCNN) to predict the spatial resolution of temperature and velocity in compartment fires. Wu et al. [10–12] utilized deep learning methods to predict the development and smoke propagation of tunnel fires, thereby demonstrating the potential of intelligent firefighting systems in laboratory-scale tunnel models. Su et al. [13] employed AI to train smoke images derived from numerical fire simulations to assist performance-based fire engineering design, which is applicable to atrium design. Ghosh et al. [14] proposed a hybrid deep learning model of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) for forest fire detection, which provides new insights into computer vision forest fire detection. Choi et al. [15] employed convolutional neural networks (CNNs) for semantic image segmentation in wildfire scenes. Ban et al. [16] developed a deep learning-based framework to monitor the development of wildfires in real time, including complex conditions such as smoke, clouds, and nighttime. Wang et al. [17] generated a large compartment fire database using CFD models, obtaining numerical simulation smoke images (front and side dual views) produced outside buildings, and used VGG16 to extract smoke features under different building fire scenarios. They established a relationship between smoke features (based on external fire information) and HRR, thereby predicting the HRR of fires inside buildings. Wang et al. [18] also used the NIST database [19,20] to construct a large fire scene image database, extracting continuous fire scene images from experimental videos, and proposed an AI image fire calorimetry method using the VGG16 deep learning model, achieving real-time prediction of fire HRR.

Previous research has concentrated on target detection tasks such as flames or smoke, real-time analysis, or prediction of basic parameters affecting fire development (such as HRR). However, studies on predictive methods for future fire parameters are extremely rare. Moreover, traditional video-based fire detection methods mostly analyze single flames or smoke, ignoring the coexistence features of flames and smoke in fire scenes. Flames are the direct result of combustion, manifesting as glowing and heating gasification phenomena. Smoke is a byproduct of combustion, appearing as a collection of floating particles after oxidation [21], as illustrated in Figure 1. Figure 1 is derived from the NIST database. Therefore, considering the common characteristics of flames and smoke is of significant importance for improving the accuracy and practicality of fire HRR predictions.
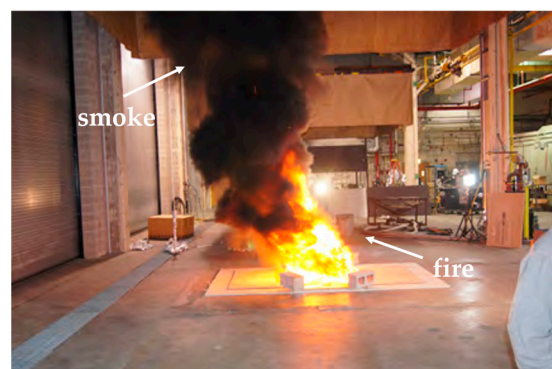

Figure 1. Image of indoor fire scene.

In summary, this paper aims to predict the future transient heat release rate (HRR) of fire scenes at the next moment/frame based on continuous fire scene images of flames and smoke and their temporal information. It integrates deep learning methods such as Bi-LSTM and Attention [22–24] (Att-BiLSTM), comprehensively modeling the temporal

relationships between fire scene image features. To construct a large-scale fire scene image dataset, this paper utilized fire scene videos from the NIST public database. Continuous fire scene images were extracted from these experimental videos in chronological order and annotated for HRR. These images were then preprocessed for the training of the deep learning model. Finally, the proposed HRR prediction method was applied to other fire scene experiments to verify its generalization ability and reliability in predicting future transient fire HRR.

2. Materials

2.1. Fire Scene Image Database

The training of deep learning models requires a substantial number of continuous fire scene images annotated with combustion HRR. This paper adopted the NIST fire calorimetry database for model training, which measures the HRR changes throughout the entire fire process from ignition to burnout using the oxygen consumption calorimetry method [25]. The database encompasses a diverse array of fire scenarios, including single burning items, fully furnished rooms, controlled burners, well-characterized fuels, and fuels of unknown composition. It encompasses HRR measurements of various transient combustibles in industrial environments, with heat release rates ranging from 50 kW to 20,000 kW. During the experiments, a digital camera at a fixed angle was used to film the entire fire test process.

A total of 40 fire experiments were selected from the NIST fire calorimetry database's Transient Combustion Calorimetry (TCC) project [26,27], which included tests with peak heat release rates (PHRRs) ranging from 10.5 to 4174 kW and total heat released ranging from 0 to 5120 MJ. These experiments covered a range of potential scenarios, from small to large fires. Figure 2 depicts a selection of combustion tests conducted with different ignition sources. The burning items represent common items found in our daily lives, including solid fuels. These experiments were conducted at the National Fire Research Laboratory of the National Institute of Standards and Technology (NIST) in Gaithersburg. The selected experimental data were all conducted under a ventilation hood measuring 6.1 m (20 feet) by 6.1 m (20 feet) (the indoor environment of buildings), with a rated capacity of 3 megawatts (MW) [28]. The duration of the fire scene experiments ranged from 4 min to 115 min, providing a variety of fire scene images and their corresponding HRRs for training deep learning models.
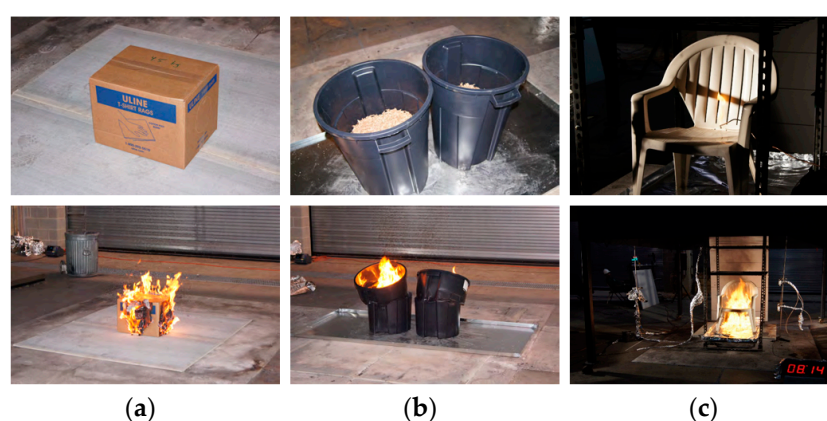

Figure 2. Combustion tests of different ignition sources. (**a**) Cardboard boxes; (**b**) rubber trash bins; (**c**) plastic chairs.

2.2. Dataset Preprocessing

This paper utilizes a series of full-process video frame images from the NIST fire calorimetry database (with a dense sampling strategy [29] at 30 FPS) to construct an image sequence training database annotated with HRR. This database contains a total of 27,231 pairs of fire scene images and HRR data. Prior to inputting the data into the model

for training, it was necessary to perform the requisite preprocessing to ensure the quality and consistency of the data, as illustrated in Figure 3.
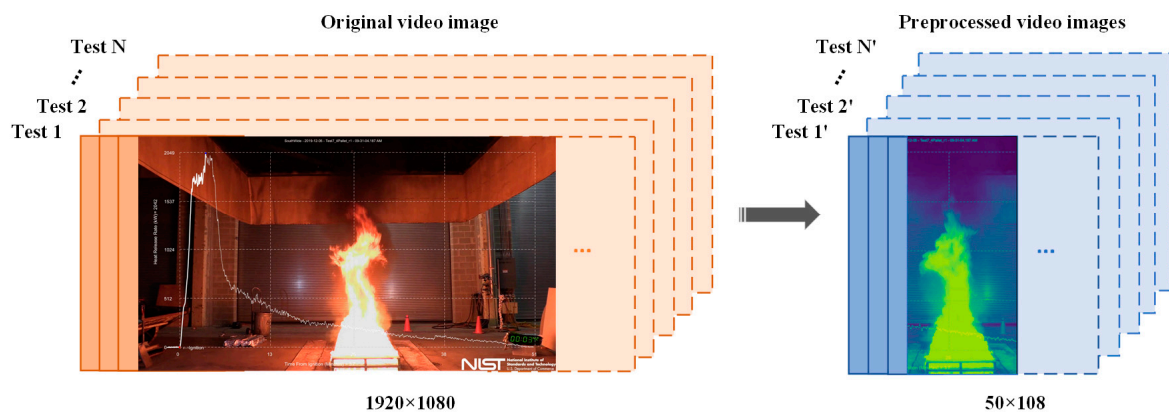

Figure 3. Preprocessing of fire scene video images.

- Image Cropping: Given the high quality of the fire video images in the NIST database (1920 × 1080), using the original images directly for training would increase the computational complexity and cost. Therefore, all fire scene images were first resized to 50 × 108. A resolution of 50 pixels in width was selected to accommodate the spatial characteristics of all the fire images in this dataset. This resolution was chosen to maximize the retention of flame features in the field of view and minimize the interference of background noise. Additionally, a height of 108 pixels was selected to ensure that the flame height and upper smoke features were adequately captured, taking into account the height of the original image, which was 1080 pixels. The height of 108 pixels was selected to account for the original image's height of 1080 pixels, ensuring that the flame height and upper smoke features were adequately captured.
- Random Horizontal Image Flipping: To increase data diversity, reduce redundancy, and improve the model's generalization ability and robustness, this paper employed a data augmentation technique of random horizontal image flipping. This technique flips the images horizontally with a probability of 0.5, altering the content of the images without changing the pixel values. This renders the model indifferent to the orientation of fire scene images, enabling more accurate recognition of images from different angles.
- Image Pixel Value Normalization: To ensure that the input parameters (pixels) of fire scene images exhibit a similar data distribution, reduce data bias, and accelerate model convergence, the pixel values of fire scene images were normalized. This process converts the pixel values from the original range of [0, 255] to [0, 1]. This precludes the occurrence of excessive discrepancies in pixel values, which could potentially result in model instability or overfitting.

Finally, the 27,231 fire scene images were divided into two subsets: a training set (80%) and a validation set (20%). This division was undertaken in order to ensure the broad applicability of the research and to eliminate potential biases. Both subsets included a variety of fire experiments conducted with different fuels and heat release rate (HRR) ranges.

3. Methods

Due to the temporal correlation and non-linear characteristics between fire HRR and the flames and smoke, predicting future transient HRR of fires with high accuracy is challenging. Deep learning technology can better capture the features of fire image sequences through the automatic training of deep neural networks, thus addressing these issues. The selection of an appropriate architecture is of paramount importance when addressing image time series tasks. Conventional convolutional neural networks (CNNs), such as the Visual Geometry Group (VGG) [30] and Residual Networks (ResNets) [31],

are networks that exhibit certain limitations in processing image time-series data. While CNNs are effective in processing static images, they are more challenging to utilize in processing time-series data. CNNs are primarily concerned with spatial feature extraction and thus lack the capacity to model temporal dynamics. Furthermore, CNNs typically necessitate a substantial quantity of data for training, as otherwise, they are susceptible to overfitting [32]. In the context of image sequence tasks, CNNs are unable to effectively capture the temporal dependencies between image frames, which represents a significant limitation for tasks that require temporal contextual information.

The proposed Att-BiLSTM model is capable of processing data in sequences and weighting them simultaneously, effectively resolving issues related to the model's sequence correlation and non-linear relationships. The Bi-LSTM model comprises two independent Long Short-Term Memory networks (LSTMs), enabling the network to consider both forward and backward information and thus facilitating the handling of long-term dependencies in image sequences [33]. Attention enhances the temporal information of the target, allowing the model to learn and determine the areas of focus, thereby enabling the model to concentrate on the most effective information with limited resources, and thus achieving better prediction accuracy [34].

3.1. Bi-LSTM Layer

Long Short-Term Memory (LSTM) [35] is a special type of Recurrent Neural Network (RNN) that introduces a structure known as 'memory cells' to address the vanishing and exploding gradient problems that arise during the training of long sequences. Each LSTM unit comprises an input gate i_t, a forget gate f_t, an output gate o_t, a candidate cell state $\tilde{c}$, a cell state c_t, and a hidden state h_t, as illustrated in Figure 4. The input gate i_t determines whether the current input information is written into the cell state c_t. The forget gate f_t decides if the information in the cell state is to be forgotten. The output gate o_t determines whether the information in the memory cell is outputted. The computation is as follows:

$$
\begin{aligned}
i_t &= \sigma\left(W^{(i)} \cdot (h_{t-1} \oplus x_t) + b^{(i)}\right), \\
f_t &= \sigma\left(W^{(f)} \cdot (h_{t-1} \oplus x_t) + b^{(f)}\right), \\
o_t &= \sigma\left(W^{(o)} \cdot (h_{t-1} \oplus x_t) + b^{(o)}\right), \\
\tilde{c}_t &= \tanh\left(W^{(i)} \cdot (h_{t-1} \oplus x_t) + b^{(c)}\right), \\
c_t &= f_t \times c_{t-1} + i_t \times \tilde{c}_t, \\
h_t &= o_t \times \tanh(c_t)
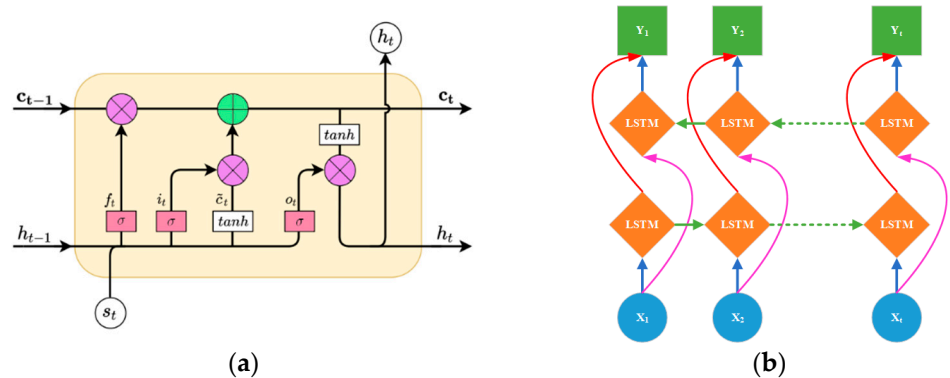\end{aligned} \tag{1}
$$

Figure 4. LSTM and Bi-LSTM. (**a**) Internal structure of an LSTM cell; (**b**) Bi-LSTM architecture diagram.

In this context, σ represents the sigmoid function, ⊕ denotes the concatenation operator, and + and × symbolize element-wise addition and multiplication operations, respectively. $W^{(x)}$ and $b^{(x)}$ are the weight matrix and bias vector for gate x, respectively.

The Bi-LSTM network structure comprises a forward and backward LSTM. It considers both past and future information, enabling the model to better capture the contextual relationships and long-distance dependencies within sequence data. There is evidence that Bi-LSTM performs better than standard LSTM in many domains, including time series prediction [36], phoneme classification [37], and others.

3.2. Attention Layer

In recent years, the attention mechanism [38] has been widely applied in the field of deep learning, inspired by the simulation of human visual attention mechanisms. The core idea of the attention mechanism is to gradually shift focus from all information to key points, that is, to allocate higher weights to important information, reasonably changing the external focus on information, ignoring irrelevant information, and amplifying the required information. In particular, the attention mechanism computes the similarity between the query and key information to obtain a weight. This weight is then normalized to obtain a usable weight. Finally, the weighted summation of this weight with the corresponding value is performed.

$$A(\text{Query, Source}) \sum_{i=1}^{L_x} \text{Similarity}(\text{Query, key}_i) * \text{Value}_i \tag{2}$$

L_x represents the length of the data source. The core idea and basic structure of the attention mechanism are illustrated in Figure 5a. In the model and structure depicted in Figure 5b, x represents the input sequence, and h denotes the hidden state, which contains information from the input sequence. This state can be considered a vector representing the input sequence x, while α represents the weight coefficient, and y is the output.
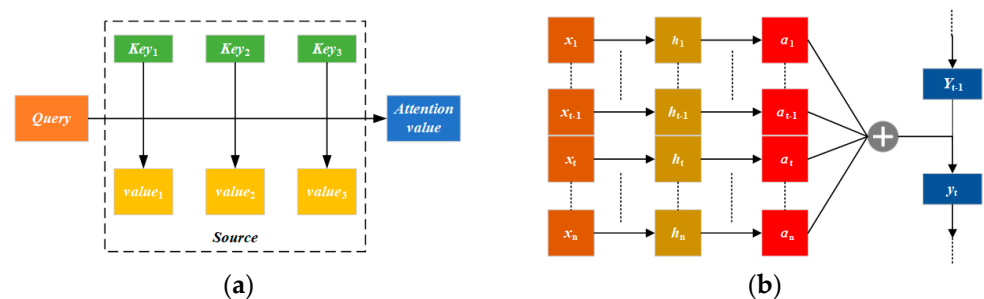

(a)   (b)

Figure 5. The attention mechanism. (**a**) The concept of the attention mechanism; (**b**) the structure of the attention model.

3.3. Model Input

This paper employs a sliding window mechanism with a stride of 1, taking each sequence of t frames of fire scene images $(x_1, x_2, \ldots x_t)$ as a group of inputs, while considering the information of the frames before and after each image, to predict the HRR of the fire scene at x_{t+1} (i.e., the future 1/30 s). In order to fully capture the temporal relationship between image sequences and achieve more accurate prediction performance, this paper employs multiple sets of t-values. However, selecting t-values that are too small (e.g., 1–8) may result in the model being unable to capture sufficient information. This is exemplified by the dynamically changing characteristics of the fire scene, which may result in underfitting. Conversely, selecting too large t-values (e.g., 12 or more) may introduce excessive noise, increase the computational complexity and time cost, and lead to overfitting [39]. Accordingly, in this paper, we select t = 9, 10, 11 for comparison experiments and analyze the main indicators, including goodness of fit (R^2), mean square error (MSE), and root mean square error (RMSE), as well as other parameters, as shown in Table 1. R^2 ranges

from [0, 1], with a closer value to 1 indicating a better fit. The regression line provides a better fit to the observed values, while the remaining four criteria have a range of [0, +∞). When the predicted value is perfectly matched with the true value, this range is equal to 0. Conversely, when the true value is equal to 0, the model is considered to be a perfect fit. As the error increases, the values become larger, indicating a poorer model.

Table 1. Comparison of model predictive performance indicators under different numbers of image sequence inputs.

t	R^2	MSE	RMSE	MAE	MAPE
9	0.96960	0.062650	0.006998	0.006693	0.005965
10	0.99700	0.061348	0.007832	0.005588	0.005463
11	0.97036	0.071012	0.006698	0.004325	0.063250

As illustrated in Table 1, when the number of input image sequences is 10 (i.e., t = 10), the coefficient of determination R^2 value of the model reaches a maximum of 0.99700, which is higher than 0.96960 (i.e., t = 9) and 0.97036 (i.e., t = 11). As illustrated in Figure 6, the comparative analysis of performance indicators, including MSE, RMSE, and MAE, and the comparative analysis of performance metrics, including MSE, RMSE, and MAE, are presented. The colors orange, yellow, and green represent the values taken as 9, 10, and 11, respectively. The vertical axis represents the magnitude of the values represented by each type of metric. It can be observed that all the prediction performance metrics are superior to those of the other two groups, with the exception of the RMSE value of the model at t = 10, which is marginally elevated in comparison to the other two groups. The experimental data indicate that while the results are favorable when t values are selected as 9, 10, and 11, the combined performance of each prediction metric reveals that the performance of inputs from frames 9 to 11 initially increases and then declines, reaching a peak at t = 10, which represents the optimal combined result. The requisite information is captured while avoiding overfitting or underfitting.
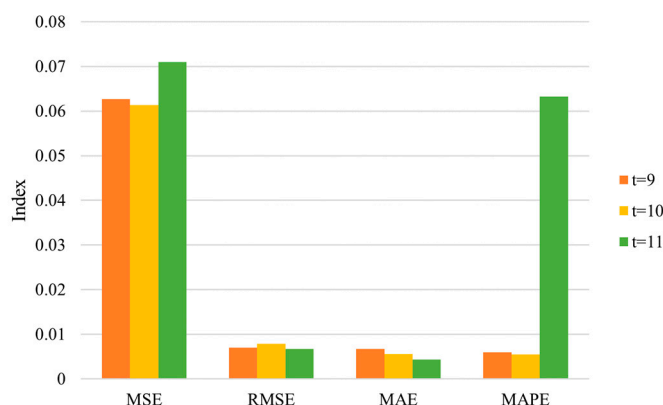

Figure 6. Comparative bar chart of some predictive performance indicators of the model.

3.4. Prediction Process

Initially, the fire scene images undergo transformations and other preprocessing operations (transforms layer) to enhance data quality. Subsequently, the preprocessed image sequences are input into the Bi-LSTM layer to extract spatiotemporal features of the image sequences. Finally, Attention is added at the output of the Bi-LSTM to strengthen the temporal information of the target, thereby identifying key features of fire scene images at different time points. This improves the prediction accuracy of future transient HRR.

Figure 7 presents the network architecture of the Att-BiLSTM. The network comprises two pathways: The upper pathway accepts HRR labels as input, providing the necessary supervisory signal for the model, which is used to correctly identify or predict data dur-

ing training. The network comprises three hidden layers, with 128, 256, and 256 units, respectively. The input dimension is 1×10, and after linear transformation and a Dropout layer, the output dimension is 1×240. The lower pathway is mainly composed of three Bi-LSTM and one Attention layer. The input to the model is a preprocessed sequence of fire scene images with dimensions of $1 \times 10 \times 108 \times 50$. This is reduced in dimensionality to 1×16 through linear activation and a Dropout layer. The two pathways are concatenated in the Connect layer, forming a 256-dimensional vector. This vector is then reduced from 256 dimensions to 1 dimension, i.e., the predicted value of future transient HRR, through a Fully Connected layer (FC) and a Dropout layer.
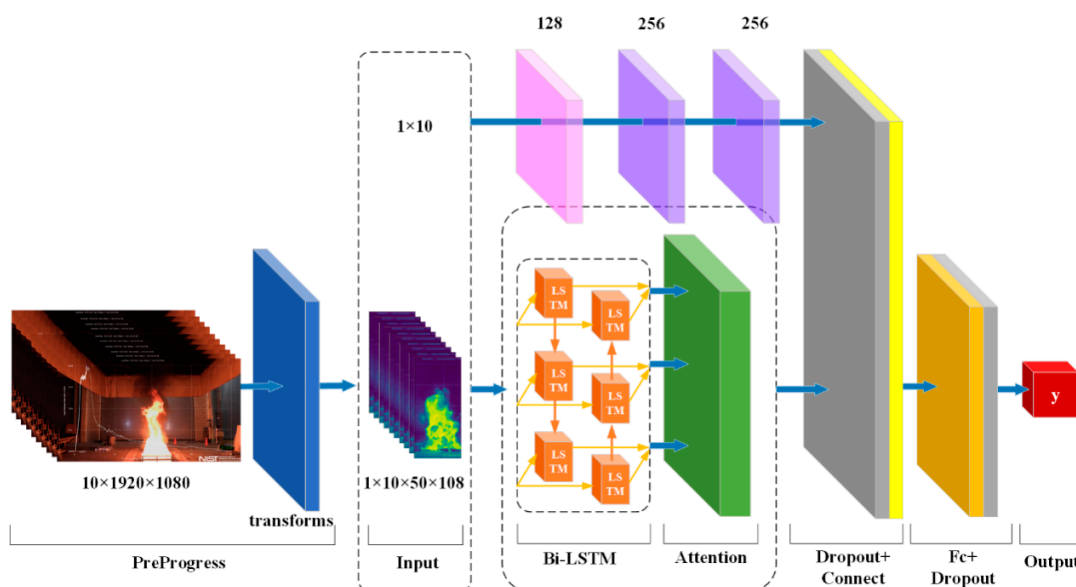

Figure 7. The Att-BiLSTM network architecture for predicting future transient HRR of a fire scene.

The network has approximately 15 million parameters (14,912,385 parameters), effectively modeling the temporal relationships between image sequences and the relationship between image data and heat release rate. The upper pathway employs a ReLU activation function three times, while the lower pathway utilizes Tanh, Softmax, and ReLU as activation functions between the Attention layer and the Connect layer. The paper employs the Mean Squared Error (MSE) and the coefficient of determination (R^2) as loss functions to evaluate the fit between predicted and actual values through residual and control charts. To prevent overfitting, both Dropout layers are set to 0.05. The training was conducted over 20 epochs on a server equipped with an RTX 4090 GPU (24 GB), taking approximately 3 h. The training results indicate that the network is capable of modeling the time dependency of fire scene image sequences and learning the importance of input. It is able to capture the long-term dependencies of image sequences, effectively processing the dynamic changes of flames and smoke in the fire scene. This enables the network to reliably predict the future transient HRR of the fire scene.

4. Results

4.1. Model Training

During the training process, as the number of iterations increased, the performance of the deep learning model continuously improved (Figure 8). After reaching 20 training epochs, the model achieved convergence, with the Mean Squared Error (MSE) reduced to 0.061348 and the coefficient of determination R^2 reaching a maximum of 0.99. Although the original image has been resized in pixels, resulting in the loss of some detail information, the resolution has been able to capture the key features of the model, which has demonstrated a satisfactory training result; this indicates that the model has obtained a satisfactory fit,

demonstrating the capacity to accurately predict the future transient heat release rate (HRR) of fire scenes within the training data.
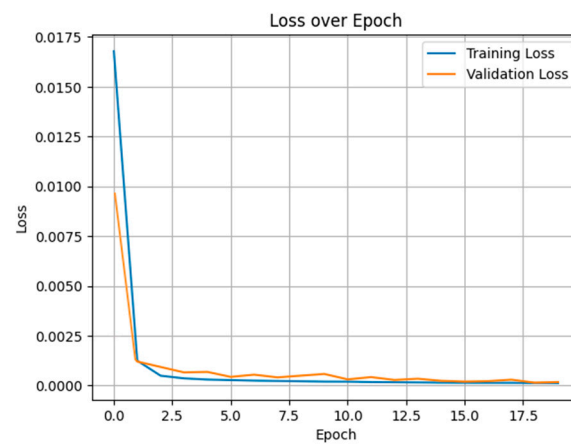

Figure 8. Loss during training and validation of the model.

4.2. Validation Set

Given the training and validation set split ratio of 8:2, the validation set consists of approximately 5000 images of fire scenes. This paper employs scatter plots (left) and line charts (right) to display and evaluate the relationship between the model's predicted values (Predict) and the actual values (Ground Truth). In the chart presented in Figure 9b, the horizontal axis (Index) represents the image sample number, and the vertical axis (Value) represents the corresponding HRR values, as illustrated in Figure 9. The horizontal coordinates of the line graph are arranged in ascending order according to the size of the HRR values. This arrangement facilitates the identification of clear trends and patterns in the data, as well as the characterization and delineation of the distinctive attributes and patterns associated with different HRR value intervals. Additionally, it reduces the noise and fluctuations inherent in time-series data, thereby enhancing the efficacy of the training process and the precision of the model's predictions. These results demonstrate that the Att-BiLSTM model proposed in this paper has excellent predictive capability on the validation set. This provides compelling evidence that applying this model to predict future transient HRR of fire scenes beyond the test set is a viable approach.
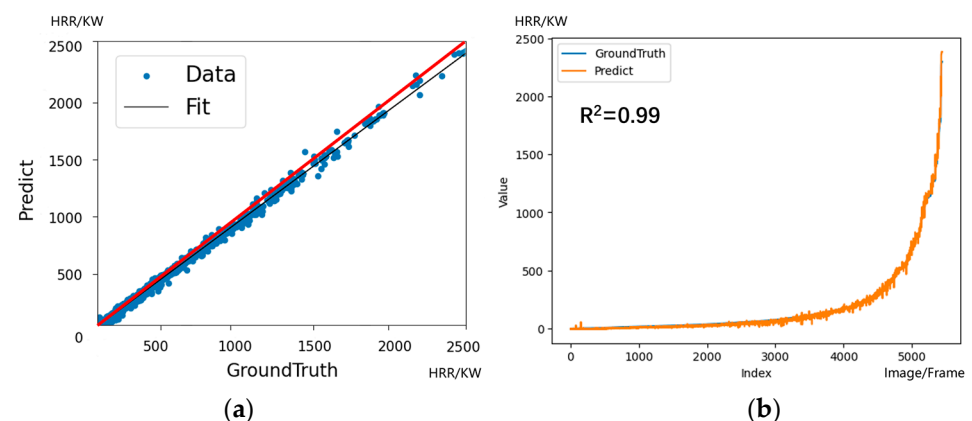

Figure 9. Test results of the validation set. (**a**) Residual plot of validation set results; (**b**) control chart of validation set results.

5. Discussion

In order to assess the model's recognition performance on data outside the training set, this paper selected a series of fire test cases with varying ranges of combustion HRRs

from the NIST fire calorimetry database for prediction. These test cases were not included in the 20% validation set partitioned during the model training process. In other words, these cases represent new, unknown samples for the trained deep learning model. The model must utilize the knowledge acquired during the training phase to predict the future transient HRR values of these unfamiliar fire scenes and unknown combustibles. This approach more accurately reflects the model's ability to generalize to real-world scenarios.

5.1. High-Brightness Fire Scenes

Changes in the brightness of the fire scene environment can result in variations in the brightness and contrast of fire scene images, potentially affecting the model's ability to extract and analyze image features. Figure 10a,b illustrate the impact of images under daylight or strong light exposure and lower brightness conditions, respectively. Therefore, the brightness conditions of the fire scene environment represent a crucial factor influencing the model's capacity for generalization in HRR prediction. To assess the model's predictive performance in high-brightness fire scenes, this paper selected three experiments from the NIST database with higher brightness conditions. These included burning items such as cardboard boxes (Figure 11a), rubber trash bins (Figure 11b), and plastic chairs (Figure 11c). The burning items in question have the same thermal parameters as those in the training dataset, but the experimental conditions differ in brightness. This allows for an examination of the robustness of the model's HRR predictions. Figure 11 illustrates this, with the left column showing a scatter plot of the demonstration results and the right column showing a line chart of the results.
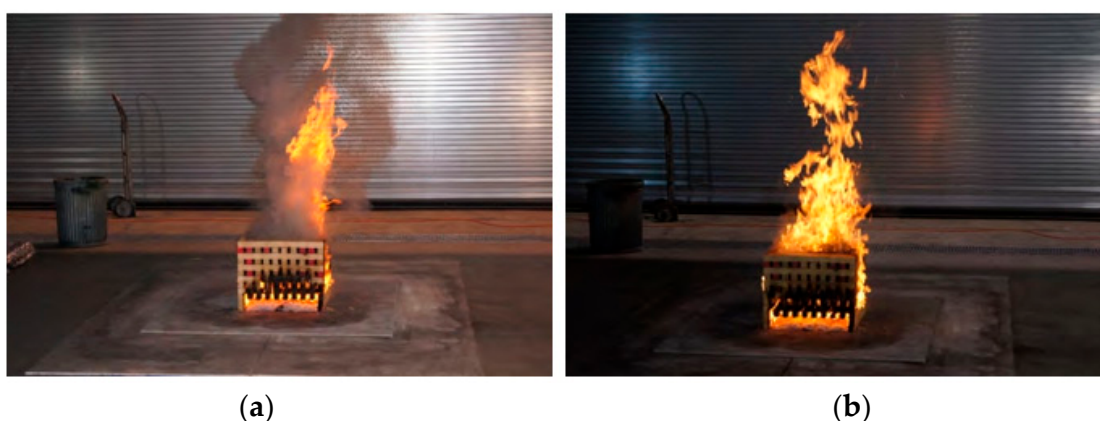

| (**a**) | (**b**) |

Figure 10. Fire scenes under different luminance conditions. (**a**) Higher luminance; (**b**) lower luminance.

The results demonstrate that even in high-brightness fire scenes, the deep learning model can accurately predict the HRR of different combustibles (Figure 11), with all R^2 values exceeding 0.97. The residual plots and result comparison charts indicate a good fit, indicating high prediction accuracy. The residual plot and result comparison chart for the cardboard box experiment (Figure 11a) exhibit a slight degree of inferiority in comparison to the other two experiments. This may be attributed to the relatively limited number of frames in the cardboard fire scene video, which has prevented the full utilization of temporal relationships between images. Overall, the model demonstrates a certain degree of adaptability to changes in the brightness of fire scenes, maintaining a high degree of consistency between predicted results and actual measurements under conditions of increased brightness. This supports the application of the model in various complex brightness environments.

5.2. Complex Combustibles

The presence of complex combustibles increases the difficulty of predicting the actual HRR of fire scenes. Such combustibles may include different states, such as solids, liquids,

and gases, leading to diversity in the characteristics of flames and smoke, such as differences in color and shape. To assess the reliability of the model's HRR predictions in fire scenes with complex combustibles, this paper selected three typical complex combustible fire scene scenarios from the NIST database for validation. These are a "box-type gas burner" (Figure 12a), a "utility cart with a laptop and printer" (Figure 12b), and "propanol liquid" (Figure 12c) [33]. The three sets of experiments simulate the complexity of combustibles in actual fire scenes and are used to test the model's generalization ability. Figure 12 illustrates the results of the demonstration in the left column, while the right column presents the results in the form of a line chart.
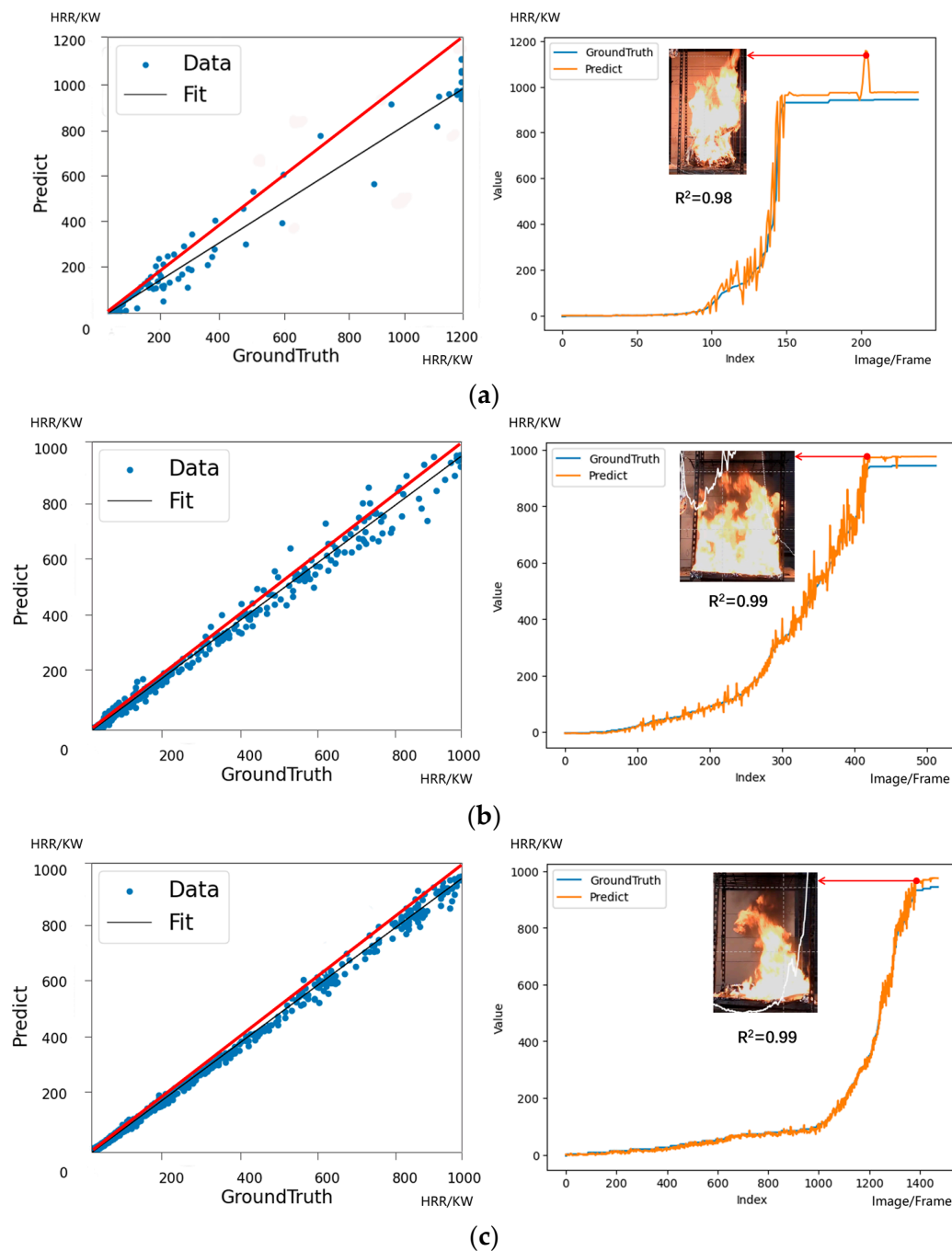

Figure 11. Demonstration result images for high-luminance fire scene environments. (**a**) Cardboard boxes; (**b**) rubber trash bins; (**c**) plastic chairs.
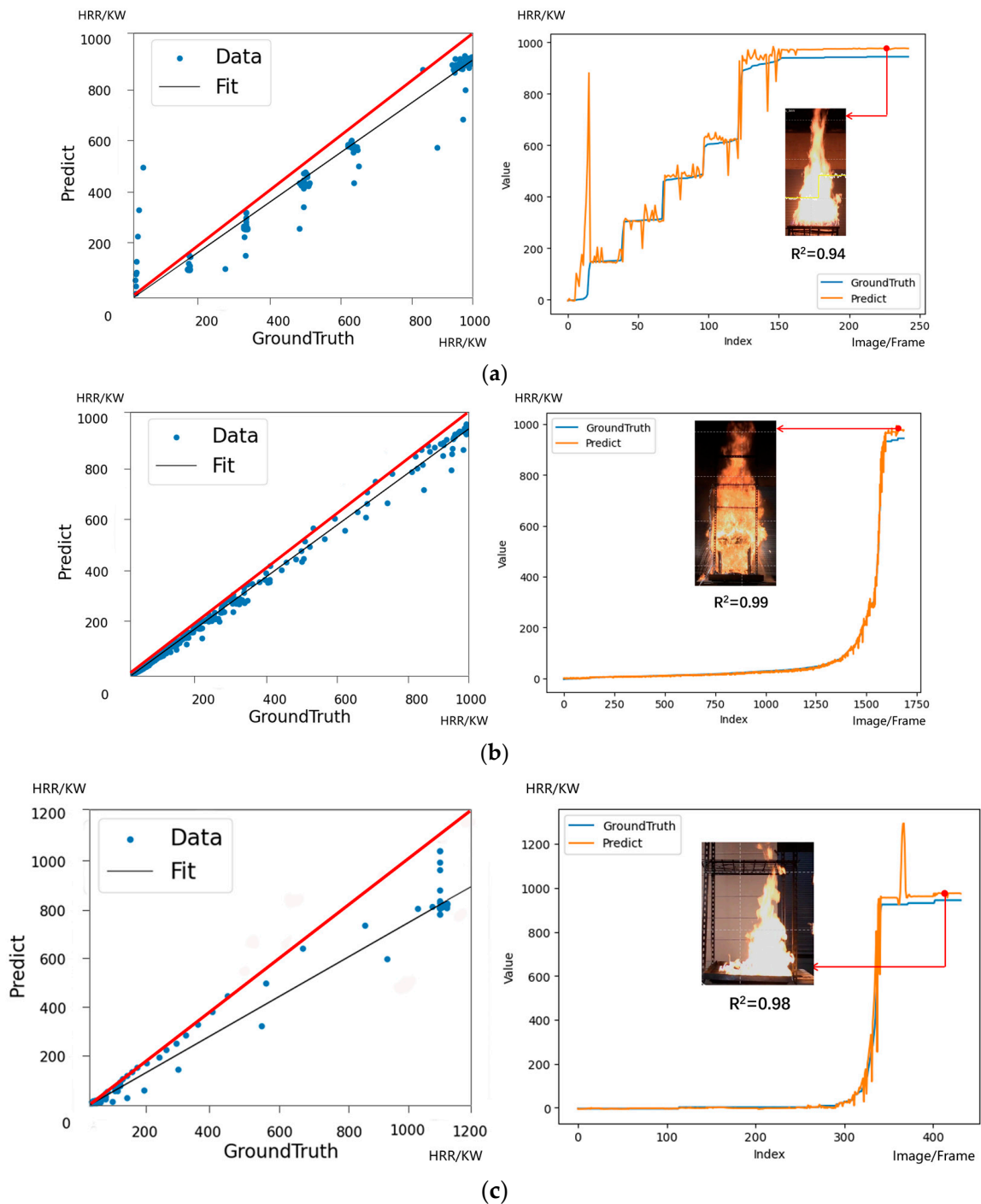
Figure 12. Demonstration result images for other complex combustibles. (**a**) Box-type gas burner; (**b**) utility cart with a laptop and printer; (**c**) propanol liquid.

As illustrated in Figure 12, we compared the actual HRR with the predicted results of the deep learning model for the three groups of fire scenarios with complex combustible characteristics. Despite the complexity of the fuel load and the fire spread process in these experiments, the results demonstrate that the deep learning model can reasonably predict the HRR of fire scenes, with all R^2 values exceeding 0.94, reflecting the changing trend of HRR during combustion. This indicates that the deep learning model has strong adaptability and predictive capability for HRR in complex fire scenarios.

However, when the fire enters the high heat release peak phase, both sets of experiments show a certain overestimation bias in the model's predictions. This may be attributed

to the distribution characteristics of the samples near the high heat release peak in the training dataset. In other words, the fire scene videos in the NIST database are time-lapse, with a greater number of scenes (frames) of small to medium HRR than those at the peak HRR stage. This results in a tendency for the model to overestimate at the peak HRR stage of large fires. In addition, the deep learning model exhibits some inaccuracy in predicting the absolute value of HRR, with a tendency to underestimate. This may be related to the scale and quality of the model's training dataset. The experimental results above demonstrate that the proposed deep learning-based method for predicting future transient HRR of fires can effectively utilize the features and temporal relationships of fire scene images, providing good predictive capability for future fuel combustion in fire scenes without the need for additional equipment or sensors. Although the prediction accuracy and applicability of the model are subject to improvement due to the scale of the training data, there is sufficient evidence to suggest that the model can effectively predict the future transient HRR of fires.

5.3. Comparative Analysis with Similar Studies

The advent of sophisticated deep learning models has facilitated remarkable advancements in the application of image recognition and computer vision techniques in the domains of fire detection and predictive analysis of fire parameters. While existing fire target detection methods [14–16] can identify fire in real time and issue timely warnings, their assessment of the current state and future trend of the fire still relies on empirical judgments and lacks the quantitative analysis of professional fire parameters. Consequently, these methods have limitations. In contrast, the analysis of real-time parameters of flames (e.g., heat release rate) based on video images allows for a more intuitive and reliable assessment of the degree of fire danger. For example, Wang Zilong et al. [18] constructed a large-scale fire image database using the NIST database and successfully predicted the heat release rate of real-time fires by extracting continuous fire images from experimental videos and combining them with the VGG16 deep learning model. Nevertheless, the objective of predicting future fires is to gain a deeper understanding of the fire situation. This approach is still limited. In this study, we propose a future transient heat release rate (HRR) prediction method based on fire video images. This method aims to complement and extend the traditional fire target detection task and real-time fire parameter analysis task. This paper explores the feasibility of futuristic analysis and prediction in the field of fire prevention and control. It presents new ideas and methods for fire monitoring and emergency response. The approach enhances the understanding of fire trends and provides more accurate data support for monitoring and preventing fires.

5.4. Applications in Intelligent Firefighting

The experiments described above have demonstrated the effectiveness of the proposed Att-BiLSTM model in predicting the future transient heat release rate (HRR) of fires. This suggests that deep learning-based technologies are poised to become a key component of intelligent firefighting systems, which could be applied in actual firefighting operations. The development of fires occurring inside buildings is significantly affected by the limited space available. As a result of the incomplete exposure of the burning area, which leads to a lack of oxygen and a slow air flow, these factors work in concert to make the fire more stable in its initial stage and to slow the expansion of the burning area compared to outdoor fires. The particular characteristics of this combustion environment necessitate the development of more sophisticated fire response strategies and safety assessment methodologies [40]. As illustrated in Figure 13, when an indoor fire occurs, the following sequence of events occurs: first, video images of the indoor fire are collected in real-time using cameras such as CCTV and smartphones; next, the fire images are uploaded to a cloud database via the network; finally, the streaming fire images are input into the deep learning model, which outputs the predicted value of the future fire HRR. This method offers the potential to simulate and predict the development of fire situations, thereby providing an earlier warning period and

enabling more effective response and command decisions. It also assists in the optimization of the allocation of firefighting resources, thereby enhancing the protection of personnel and reducing property loss. This method represents a significant advance in the field of intelligent firefighting.
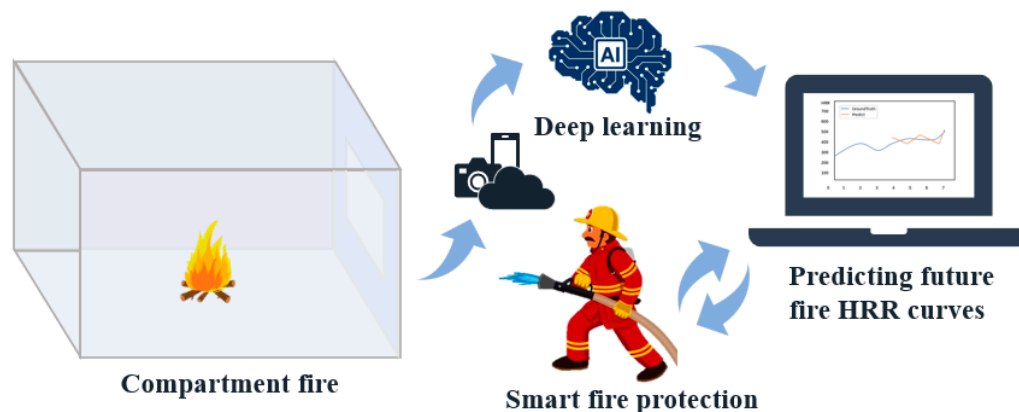

Figure 13. The application of future fire HRR prediction based on fire scene images and deep learning in intelligent firefighting.

Although this method has achieved satisfactory results in laboratory environments, there are still some issues and deficiencies that need further refinement and optimization. Firstly, to enhance the predictive performance and adaptability of this method in various fire scenarios, it is necessary to expand and enrich the database of fire scene images to cover a wider range of real fire situations and scales. Secondly, since flames are three-dimensional and camera images can only capture two-dimensional projections, it is difficult to obtain depth information about flames. Therefore, future depth models should consider using multi-angle camera images to reconstruct the three-dimensional form of the fire scene, combine temporal information, and extract more features. Thirdly, the current method uses AI image calorimetry [41] to obtain the HRR of fire video images in real time. The advantage of this method is that it does not rely on additional instruments and is low-cost, but it may also lead to a certain degree of error.

6. Conclusions

This paper proposes a deep learning model that integrates Bi-LSTM and Attention mechanisms, capable of simultaneously processing and weighting data in sequences. This model effectively addresses the temporal correlation and non-linear relationships between fire HRR and images of flames and smoke. The contributions of this paper include the following aspects:

1. A new end-to-end method for predicting future fire HRR is proposed. By inputting fire scene images and corresponding HRR label data into the Att-BiLSTM model and employing a sliding window mechanism, it is possible to achieve continuous output of future transient fire HRR predictions.
2. In the preprocessing of fire scene images, the quality of images is enhanced while reasonably preserving the information of flames and smoke. This is achieved by fully considering their coexistence characteristics and their impact on fire HRR.
3. The model's generalization ability and reliability were tested in high-brightness environments and fire scenes with complex combustibles. The experimental results demonstrate that the model can accurately predict future transient HRR of fire scenes and can also simulate and predict the development trend of fire situations to a certain extent.

This paper presents a novel method of using deep learning technology to predict the future transient HRR of fires, which has broad application prospects and high potential value for the development of future intelligent firefighting systems. In future work, we intend to further explore and improve the deep learning model. This will involve

introducing more image features and inter-frame information, as well as considering the combined effects of more influencing factors. This will enhance the accuracy and duration of predictions for future fire HRR.

References

1. Johansson, N.; Svensson, S. Review of the Use of Fire Dynamics Theory in Fire Service Activities. *Fire Technol.* **2018**, *55*, 81–103. [CrossRef]
2. Tewarson, A. Heat release rate in fires. *Fire Mater.* **2004**, *4*, 185–191. [CrossRef]
3. Thornton, W.M. XV. The relation of oxygen to the heat of combustion of organic compounds. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **2009**, *33*, 196–203. [CrossRef]
4. Sun, P.; Wu, C.; Zhu, F.; Wang, S.; Huang, X. Microgravity combustion of polyethylene droplet in drop tower. *Combust. Flame* **2020**, *222*, 18–26. [CrossRef]
5. Xiong, C.; Fan, H.; Huang, X.; Fernandez-Pello, C. Evaluation of burning rate in microgravity based on the fuel regression, flame area, and spread rate. *Combust. Flame* **2022**, *237*, 111846. [CrossRef]
6. Sun, X.; Hu, L.; Zhang, X.; Ren, F.; Yang, Y.; Fang, X. Experimental study on flame pulsation behavior of external venting facade fire ejected from opening of a compartment. *Proc. Combust. Inst.* **2021**, *38*, 4485–4493. [CrossRef]
7. Li, C.; Li, X.; Chen, M.; Sun, X. Deep Learning and Image Recognition. In Proceedings of the 2023 IEEE 6th International Conference on Electronic Information and Communication Technology (ICEICT), Qingdao, China, 21–24 July 2023; pp. 557–562.
8. Taşyürek, M. ODRP: A new approach for spatial street sign detection from EXIF using deep learning-based object detection, distance estimation, rotation and projection system. *Vis. Comput.* **2024**, *40*, 983–1003. [CrossRef]
9. Hodges, J.L.; Lattimer, B.Y.; Luxbacher, K.D. Compartment fire predictions using transpose convolutional neural networks. *Fire Saf. J.* **2019**, *108*, 102854. [CrossRef]
10. Wu, X.; Park, Y.; Li, A.; Huang, X.; Xiao, F.; Usmani, A. Smart Detection of Fire Source in Tunnel Based on the Numerical Database and Artificial Intelligence. *Fire Technol.* **2021**, *57*, 657–682. [CrossRef]
11. Wu, X.; Zhang, X.; Huang, X.; Xiao, F.; Usmani, A. A real-time forecast of tunnel fire based on numerical database and artificial intelligence. *Build. Simul.* **2022**, *15*, 511–524. [CrossRef]
12. Xiqiang, W.; Zhang, X.; Jiang, Y.; Huang, X.; Huang, G.Q.; Usmani, A. An intelligent tunnel firefighting system and small-scale demonstration. *Tunn. Undergr. Space Technol.* **2022**, *120*, 104301. [CrossRef]
13. Su, L.-c.; Wu, X.; Zhang, X.; Huang, X. Smart performance-based design for building fire safety: Prediction of smoke motion via AI. *J. Build. Eng.* **2021**, *43*, 102529. [CrossRef]
14. Ghosh, R.; Kumar, A. A hybrid deep learning model by combining convolutional neural network and recurrent neural network to detect forest fire. *Multimed. Tools Appl.* **2022**, *81*, 38643–38660. [CrossRef]
15. Choi, H.-S.; Jeon, M.; Song, K.; Kang, M. Semantic Fire Segmentation Model Based on Convolutional Neural Network for Outdoor Image. *Fire Technol.* **2021**, *57*, 3005–3019. [CrossRef]
16. Ban, Y.; Zhang, P.; Nascetti, A.; Bevington, A.R.; Wulder, M.A. Near Real-Time Wildfire Progression Monitoring with Sentinel-1 SAR Time Series and Deep Learning. *Sci. Rep.* **2020**, *10*, 1322. [CrossRef] [PubMed]
17. Wang, Z.; Zhang, T.; Wu, X.; Huang, X. Predicting transient building fire based on external smoke images and deep learning. *J. Build. Eng.* **2022**, *47*, 103823. [CrossRef]
18. Wang, Z.; Zhang, T.; Huang, X. Predicting real-time fire heat release rate by flame images and deep learning. *Proc. Combust. Inst.* **2023**, *39*, 4115–4123. [CrossRef]
19. Fire Calorimetry Database (FCD). Available online: https://www.nist.gov/el/fcd (accessed on 30 April 2024).
20. The NIST 20 MW Calorimetry Measurement System for Large-Fire Research. Available online: https://www.nist.gov/publications/nist-20-mw-calorimetry-measurement-system-large-fire-research (accessed on 30 April 2024).

21. Jin, C.; Wang, T.; Alhusaini, N.; Zhao, S.; Liu, H.; Xu, K.; Zhang, J. Video Fire Detection Methods Based on Deep Learning: Datasets, Methods, and Future Directions. *Fire* **2023**, *6*, 315. [CrossRef]

22. Wang, Z.; Yang, B. Attention-based Bidirectional Long Short-Term Memory Networks for Relation Classification Using Knowledge Distillation from BERT. In Proceedings of the 2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), Calgary, AB, Canada, 17–22 August 2020; pp. 562–568.

23. Zhang, Q.; Wang, R.; Qi, Y.; Wen, F. A watershed water quality prediction model based on attention mechanism and Bi-LSTM. *Environ. Sci. Pollut. Res.* **2022**, *29*, 75664–75680. [CrossRef]

24. Luo, J.; Zhang, X. Convolutional neural network based on attention mechanism and Bi-LSTM for bearing remaining life prediction. *Appl. Intell.* **2022**, *52*, 1076–1091. [CrossRef]

25. User's Guide for Fire Calorimetry Database (FCD). Available online: https://www.nist.gov/system/files/documents/2020/11/19/FCD_User_Guide_v4a.pdf (accessed on 30 April 2024).

26. Heat Release Rates of Multiple Transient Combustibles. Available online: https://www.nist.gov/publications/heat-release-rates-multiple-transient-combustibles (accessed on 30 April 2024).

27. Heat Release Rate and Fire Characteristics of Fuels Representative of Typical Transient Fire Events in Nuclear Power Plants. Available online: https://www.nrc.gov/docs/ML2009/ML20091L481.pdf (accessed on 30 April 2024).

28. NIST Technical Note 2102 Heat Release Rates of Multiple Transient Combustibles. Available online: https://nvlpubs.nist.gov/nistpubs/TechnicalNotes/NIST.TN.2102.pdf (accessed on 30 April 2024).

29. Wang, L.; Xiong, Y.; Wang, Z.; Qiao, Y.; Lin, D.; Tang, X.; Van Gool, L. Temporal Segment Networks: Towards Good Practices for Deep Action Recognition. In Proceedings of the Computer Vision—ECCV 2016, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Cham, Switzerland, 2016; pp. 20–36.

30. Gu, S.; Ding, L. IEEE A Complex-Valued VGG Network Based Deep Learing Algorithm for Image Recognition. In Proceedings of the 9th International Conference on Intelligent Control and Information Processing (ICICIP), Wanzhou, China, 9–11 November 2018; pp. 340–343.

31. Ho, W.-H.; Huang, T.-H.; Yang, P.-Y.; Chou, J.-H.; Huang, H.-S.; Chi, L.-C.; Chou, F.-I.; Tsai, J.-T. Artificial intelligence classification model for macular degeneration images: A robust optimization framework for residual neural networks. *BMC Bioinform.* **2021**, *22*, 148. [CrossRef] [PubMed]

32. Fu, G.; Wei, Q.; Yang, Y.; Li, C. Bearing fault diagnosis based on CNN-BiLSTM and residual module. *Meas. Sci. Technol.* **2023**, *34*, 12. [CrossRef]

33. Telili, A.; Fezza, S.A.; Hamidouche, W.; Brachemi Meftah, H.F.Z. 2BiVQA: Double Bi-LSTM-based Video Quality Assessment of UGC Videos. *ACM Trans. Multimed. Comput. Commun. Appl.* **2023**, *20*, 1–22. [CrossRef]

34. Soydaner, D. Attention mechanism in neural networks: Where it comes and where it goes. *Neural Comput. Appl.* **2022**, *34*, 13371–13385. [CrossRef]

35. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]

36. Siami-Namini, S.; Tavakoli, N.; Namin, A.S. The Performance of LSTM and BiLSTM in Forecasting Time Series. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 3285–3292.

37. Graves, A.; Schmidhuber, J. Framewise phoneme classification with bidirectional LSTM networks. In Proceedings of the 2005 IEEE International Joint Conference on Neural Networks, Montreal, QC, Canada, 31 July–4 August 2005; pp. 2047–2052.

38. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2017**, *30*. [CrossRef]

39. Sutskever, I.; Vinyals, O.; Le, Q.V. Sequence to Sequence Learning with Neural Networks. *arXiv* **2014**, *2*, 3104–3112. [CrossRef]

40. Pincott, J.; Tien, P.W.; Wei, S.; Calautit, J.K. Indoor fire detection utilizing computer vision-based strategies. *J. Build. Eng.* **2022**, *61*, 105154. [CrossRef]

41. Zilong, W.; Ding, Y.; Zhang, T.; Huang, X. Automatic real-time fire distance, size and power measurement driven by stereo camera and deep learning. *Fire Saf. J.* **2023**, *140*, 103891. [CrossRef]