

## Full Length Article

## A comprehensive investigation of LSTM-CNN deep learning model for fast detection of combustion instability

Zengyi Lyu<sup>a</sup>, Xiaowei Jia<sup>b</sup>, Yao Yang<sup>a</sup>, Keqi Hu<sup>a</sup>, Feifei Zhang<sup>a</sup>, Gaofeng Wang<sup>a,\*</sup><sup>a</sup> School of Aeronautics and Astronautics, Zhejiang University, Hangzhou 310027, China<sup>b</sup> Department of Computer Science, University of Pittsburgh, Pittsburgh, PA 15260, United States

## ARTICLE INFO

## Keywords:

Premixed swirling flame  
Combustion instability  
Deep learning  
Convolutional neural network  
LSTM

## ABSTRACT

In this paper, we propose a deep learning model to detect combustion instability using high-speed flame image sequences. The detection model combines Convolutional Neural Network (CNN) and Long Short-Term Memory network (LSTM) to learn both spatial features and temporal correlations from high-speed images, and then outputs combustion instability detection results. We also visualize the extracted spatial features and their temporal evolution to interpret the detection process of model. In addition, we discuss the effect of different complexity of CNN layers and different amounts of training data on model performance. The proposed method achieves superior performance under various combustion conditions in swirl chamber with high accuracy and a short processing time about 1.23 ms per frame. Hence, we show that the proposed deep learning model is a promising detection tool for combustion instability under various combustion conditions.

## 1. Introduction

Combustion instability (CI) has become one significant issue in the development of propulsion and power systems such as gas turbines and rocket engines. CI arises from a positive coupling between heat release rate fluctuations and pressure oscillations [1–5]. CI often occurs unexpectedly, and can lead to high amplitude noise and vibration levels, sometimes even causes structural damage and catastrophic failure resulting in huge economic losses [6,7]. Due to the great damage of unsteady combustion, CI must be suppressed or controlled. The fast detection of instability in combustion systems is of great importance, and thus CI detection control, including passive control [8–14] and active control [15–20], have been widely studied. Dynamic pressure signal is frequently monitored to characterize a combustion process [21]. However, it remains challenging to effectively distinguish acoustic signals from the whole engine noises, which negatively impacts the early-detection of CI. On the other hand, flame structures show the spatial distribution of heat release and are tightly related to thermoacoustic instabilities [22]. The availability of endoscopic high-speed cameras allows monitoring the flame structure in real-time. Such sequential images of flame structures could be used for CI detection while also providing more physical insights.

With the development of computing power and the enhanced

capabilities for storing large data sets, artificial intelligence (AI) has provided unrealized potential in processing higher-dimensional data. Given the success of AI in a broad range of commercial applications, there is a growing interest in using these AI-based techniques in combustion, e.g. data-driven combustion modeling [23–25], chemistry simulation [26,27], combustion regime identification [28], combustion simulation [29]. Due to the superior performance of the convolutional neural network (CNN) [30] in image processing, several methods have been proposed for CI prediction using flame image sequences. Sarkar et al. [31] proposed a neural-symbolic framework for early detection of instability. The framework includes CNN to extract low-dimensional features and Symbolic Time Series Analysis to capture temporal evolution of the extracted feature. Ghosal et al. [32] formulated CI detection as binary classification and successfully applied 3D CNN architecture for early detection of CI. The 3D model can leverage temporal correlations among consecutive frame images, but the increased dimension of CNN can increase the number of parameters. As the limited receptive field of 3D CNN, image sequence length is limited and the problem of long-term dependence appears. Choi et al. [33] proposed a 2D CNN model with two network fusion layers to improve the performance of combustion states prediction. Instead of modeling complicated temporal behaviors, the model uses fusion layers to predict one combustion state of the last frame of the input sequence. The use of Long Short-Term Memory (LSTM) [34,35] also enables better modeling of long-term dependencies

\* Corresponding author.

E-mail address: [gfwang@zju.edu.cn](mailto:gfwang@zju.edu.cn) (G. Wang).

Nomenclature	
Re	Reynolds number
$\Phi$	equivalence ratio
p	acoustic pressure
$H_F$	convolution kernel feature
W	weight matrices
I	image intensity
b	kernel bias
$h_t$	hidden state of LSTM
$c_t$	internal state of LSTM
$x_t$	input matrices
$y_t$	output matrices
$f_t$	forget gate in LSTM
$i_t$	input gate in LSTM
$o_t$	output gate in LSTM
$\hat{y}_t$	output vector
$\hat{Y}_t$	target state vector for input data
$\odot$	element-wise product
$\sigma$	sigmoid function
S	set of {stable, unstable}
CI	combustion instability
TK	temporal kurtosis of acoustic pressure
DMD	dynamic mode decomposition
PRS	probability of stable regime
PRU	probability of unstable regime
TP	True Positive
FP	False Positive
FN	False Negative
TN	True Negative

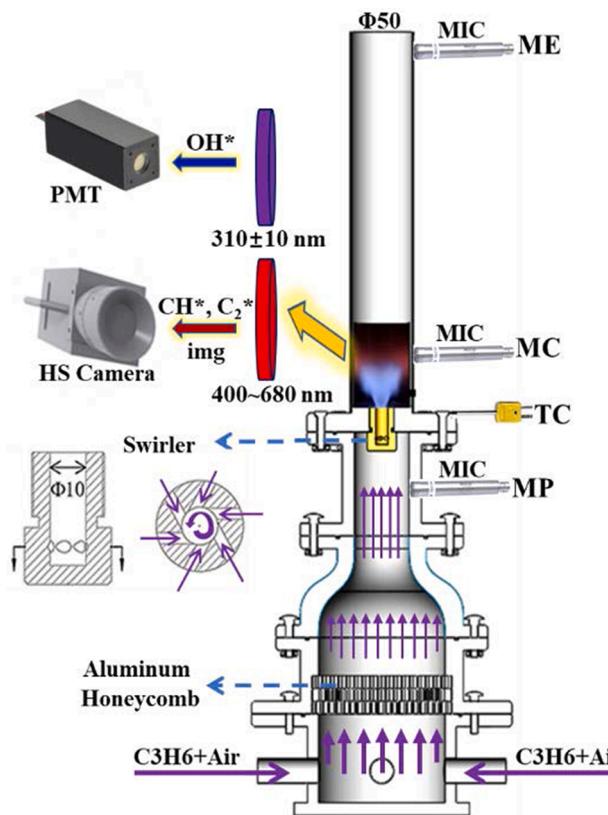


Fig. 1. Schematic of swirling combustor and flame image acquisition system.

in nonlinear dynamic system. Gangopadhyay et al. [7] proposed an efficient detection framework combining 2D CNN and LSTM. They also formulated CI detection as binary classification for stable and unstable image sequences. Their model predicts a single class label for the entire input sequence with an accuracy of 82.9%.

In this paper, we aim to develop a data-driven model which leverages spatial-temporal data patterns to improve the detection of CI. Our model also enables a real-time analysis of high-speed images with a sampling frequency of thousands of FPS. Unlike previous network architectures which only predict one stable or unstable classification detection result for the entire input sequence, our method can generate detection results of each frame in a short processing time. On this basis, research of deep learning on the combustion states change can go further

to develop an early prediction tool for CI which can detect the critical transition at an early stage.

In particular, our detection model takes the input of sequential chemiluminescence flame images that are collected using a high-speed camera on an experimental platform of premixed swirling flame. The model uses a deep learning architecture, which combines CNN and LSTM. The early CNN layers extract spatial features of the input image sequence and then feed them as sequential input to the following LSTM to learn temporal correlation. Finally, the model produces detection results of each image indicating whether the system is stable or unstable. Different model architectures have been tested to better understand the model performance and convergence. In order to interpret the function of each layer in the model, we visualize spatial features extracted by CNN layer and the temporal evolution of these features learned by LSTM. We evaluate the proposed model using high-speed flame image sequences under different combustion conditions in swirl chamber. The model achieves a high accuracy for detecting CI in a short processing time.

This paper is structured as follows. The experimental configuration and condition are described in Section 2. A deep learning-based detection model is proposed in Section 3. Model performance and detection results are discussed in Section 4. Finally, we conclude our work in Section 5.

## 2. Experimental set-up and dataset acquisitions

In this section, we introduce an experimental set-up using a propane-air premixed swirling combustor. We establish a flame image acquisition system to acquire high-speed flame image sequence under different conditions and use an instability system to label these image sequences. A large amount of combustion data was obtained to ensure the reliability of model training.

### 2.1. Experimental configurations

A propane-air premixed swirling combustor is established to acquire image sequence. Fig. 1 provides a schematic of the experimental setup. The laboratory-scale combustor consists of two parts, distribution chamber (plenum) and combustion chamber. Propane and air are fully premixed in a venturi mixer before entering the plenum, uniformly delivering into a swirl chamber through 3 layers of aluminum honeycomb perforated panels and a convergent section, and then ignited at the outlet of the swirl nozzle. The nozzle outlet diameter is 10 mm, and the swirl number is about 0.71 [36]. Above the exit is a cylindrical quartz glass window of which diameter is 50 mm, and length is 200 mm or 400 mm for acoustic feedback and transparent visualization of flame. The

**Table 1**  
Experimental conditions for the flame datasets.

Datasets	Stable swirling flame	Unstable swirling flame
Dataset 1 Priori & Posterior Tests	Case A: $Re=2828, \Phi=0.73$ ;	Case B: $Re=2828, \Phi=0.80$ ;
	Case C: $Re=5000, \Phi=0.75$ ;	Case D: $Re=5000, \Phi=0.85$ ;
Dataset 2	$Re=2828, \Phi=0.75/0.76/0.78$ ;	$Re=2828, \Phi=0.84/0.88/0.92$ ;
Training	$Re=3525, \Phi=0.92/0.95$ ;	$Re=3102, \Phi=0.88$ ;
	$Re=3807, \Phi=1.00$ ;	$Re=3525, \Phi=1.10/1.15/1.20$ ;
Dataset 3	$Re=2828, \Phi=0.78^*$ ;	$Re=3102, \Phi=0.88^*$ ;
Validation Test	$Re=3525, \Phi=0.92^*$ ;	$Re=3525, \Phi=1.20^*$ ;
Dataset 4	$Re=2828, \Phi=0.73^*/0.74$ ;	$Re=2828, \Phi=0.78/0.80^*$ ;
Statistical Test	$Re=3525, \Phi=0.98/1.00$ ;	$Re=3384, \Phi=0.93/0.95$ ;
Dataset 5	$Re=4230, \Phi=1.25$ ;	$Re=4230, \Phi=1.10/1.35$ ;
Fake Sequences for Test	Case F1: one segment consists of 200 repetitions of one single frame extracted from the case B	Case F2: transitional segment by concatenating sequence of 200 frames in Case A and later 200 frames in Case B;

\*: Experiment conditions are selected from Dataset 2, while flame image sequence segments are from different time periods.

# : Data from the same experiment is also used in Dataset 1, more flame image sequence segments included for statistics.

inlet velocity and equivalence ratio are adjusted to obtain enough stable flame and unstable flame in different modes. Table 1 shows the details of our experimental combustion conditions to acquire the flame datasets. The next section will describe how it is organized in detail.

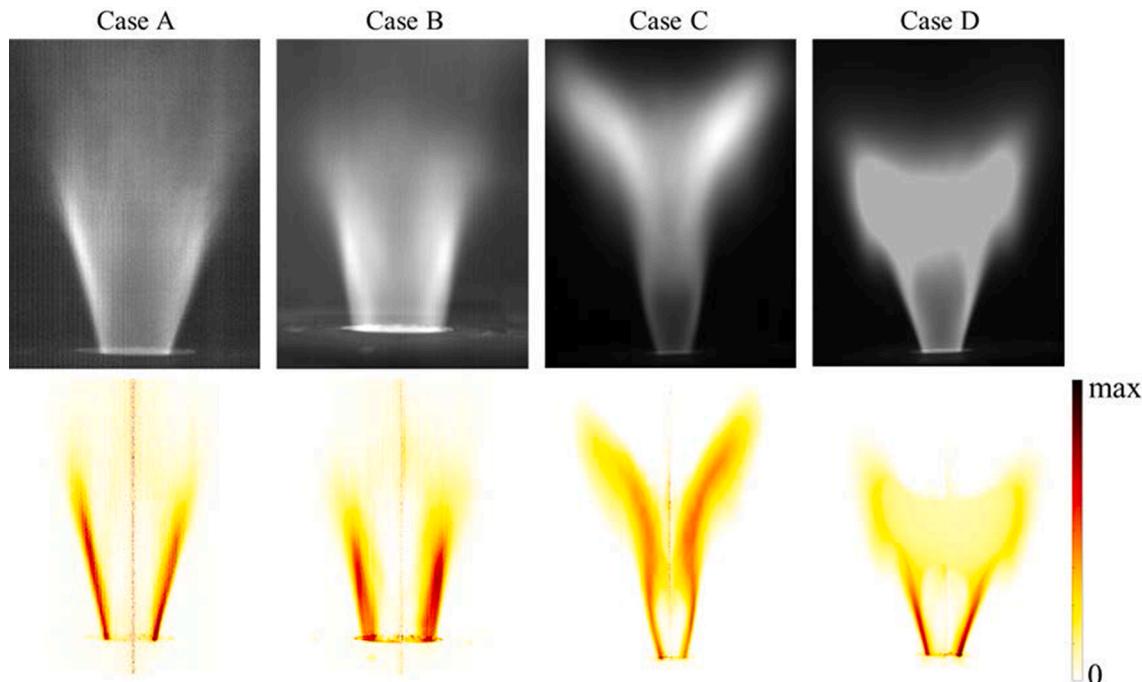
A high-speed video camera (PCO dimax HD, through Lens 100 mm Nikon mounted with a 400–680 nm UV-IR-cut filter) within 2000 fps acquisition frequency and  $1200 \times 976$  pixels image resolution is installed to carry out real-time imaging acquisition of  $\text{CH}^*$  chemiluminescence from flames in the combustion chamber. According to Shannon's sampling theorem, the acquisition frequency of the camera can detect frequency up to 1000 Hz. And the maximum instability frequency corresponding to working condition in this experiment (Tab. 1) is no more than 500 Hz. Therefore, the resolution matches the flame oscillations on the acoustic time scale. Photomultiplier Tube (PMT) with a narrow-pass filter ( $310 \pm 10$  nm) is used to collect  $\text{OH}^*$

chemiluminescence signals from the combustion reactions, which are supposed to be in proportion with the heat releasing rates [37]. Three WM-034CX Microphone (ME, MC, MP) sensors amplified by chip with a 20 dB fixed-gain (Maxim MAX9812) are adopted to record acoustic pressure signals in the exit, chamber and plenum of the combustor respectively.

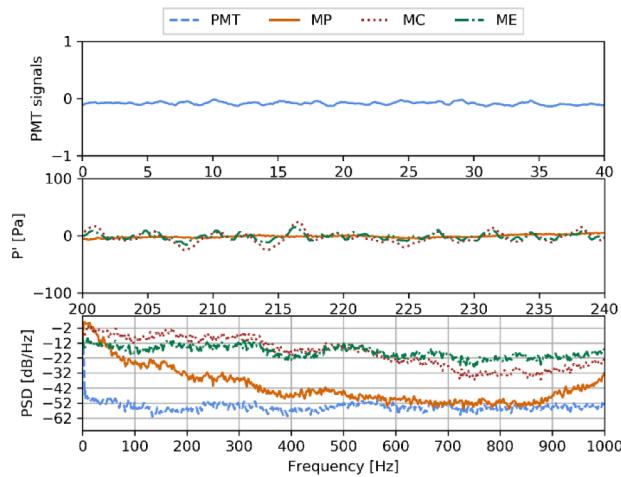
## 2.2. Flame dataset organization

High-speed flame image sequences are needed for training the proposed deep learning model. As shown in Fig. 2, different Reynolds numbers and different equivalence ratios both have influence on the flame structure. Besides, the averaged luminescence structures of premixed swirling flames are related to the experimental conditions. Details of experimental parameters in the four cases can be found in Dataset 1 (Table 1). The top is an ensemble average of two thousand instantaneous images which highlights the mean flame pattern. The bottom is obtained by taking the Abel transform to the average image, which yields a slice through the mean flame. The symmetrical V-type structures are observed for these operating conditions. The flame structures feature different heights and angles under different working parameters.

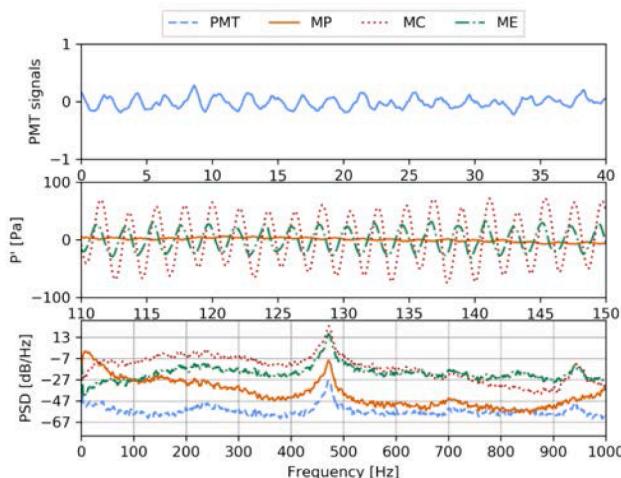
As for the deep learning study, a large amount of flame image sequences was required. They were collected from twenty-six cases under different combustion conditions of Reynolds numbers ( $Re$ ) from 2828 to 5000 and equivalence ratio  $\Phi$  from 0.73 to 1.35. Temporal kurtosis (TK) [21] analysis (detailed in Section 2.3) was adopted here as the indicator of combustion instability. Table 1 shows how we divide the collected data into 5 datasets. The sequences of flame image were divided into a series of elementary segments, each segment containing 200 consecutive frames. Dataset 2 is the training set with 120 segments from thirteen combustion conditions, with a total of 24,000 images. The size of dataset is determined by network depth and computing ability, and the convergence assessment with respect to size is discussed in Section 4. It is used to train the model, i.e., tuning model parameters (e.g., networks weights). Once the model is trained, it is then expected to detect new image sequences. Dataset 3 and Dataset 4 are used here to test the performance of the model. Dataset 3 is the validation set, which is employed to verify that the model can be used for CI detection. There are



**Fig. 2.** Average images (top) and Abel transform of average images (bottom).



(a) Stable regime (Case C)



(b) Unstable regime (Case D)

Fig. 3. Transient signals probed by OH\*-PMT and MICs [38].

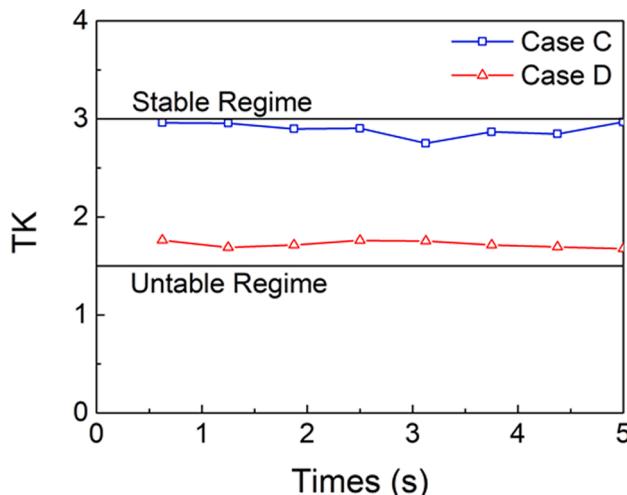


Fig. 4. Stable and unstable regimes defined by TK value corresponding to acoustic pressure.

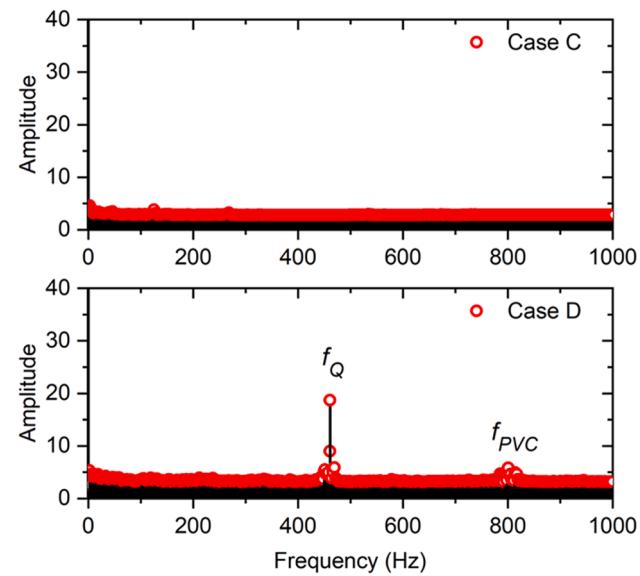


Fig. 5. DMD spectrums of Case C and Case D.

four image sequence segments in Dataset 3. The four corresponding experimental conditions are from Dataset 2, but image sequences of different time periods are selected. Dataset 4 is the statistical test set with 106 segments from eleven experiment conditions, with a total of 21,200 frames. Its experimental conditions are different from those of Dataset 2 and it is used to evaluate the performance of the model. The other two datasets are listed separately. Dataset 1 contains four image sequence segments to illustrate stable and unstable labeling. Case A ( $Re = 2828$ ,  $\Phi = 0.73$ ) and B ( $Re = 2828$ ,  $\Phi = 0.80$ ) are selected for priori and posterior model performance tests. Case C ( $Re = 5000$ ,  $\Phi = 0.75$ ) and D ( $Re = 5000$ ,  $\Phi = 0.85$ ) have larger Reynolds numbers than other cases. They are also employed for testing the generalization of the detection model. Dataset 5 contains two fake cases to verify that the model can leverage both spatial and temporal features to improve the detection of CI. Case F1 is an image sequence segment composed of 200 repetitions of a single frame extract from Case B. Case F2 is a transitional segment combined with a first 200 frames from stable (Case A) and later 200 frames from unstable (Case B).

Flame images have been collected during a 1-second period with a resolution of  $1200 \times 976$  pixels and then resized to  $224 \times 224$  pixels to reduce the computational cost. According to the flame image data collected under different combustion conditions, the flame instability detection problem is proposed. These image data are used to train our detection model and test the performance of our image-based detection model as acoustic feedback levels increase over time.

### 2.3. Stable and unstable labeling

The amplitudes of combustion oscillations would be amplified when the heat release rate and acoustic pressure are in phase with each other. CI is measured with high-speed flame image acquisition, synchronized by transient OH\* and acoustic pressure signals probed by OH\*-PMT and MICs. The corresponding spectral analysis is performed by calculating the power spectral density (PSD) of signals. Fig. 3 describe two typical cases [38]. Fig. 3a presents Case C which is in a stable regime with  $Re = 5000$  and  $\Phi = 0.75$ . There are marginal fluctuations in the acoustic pressure and heat release rate signals. When the equivalence ratio increases to  $\Phi = 0.85$  as Case D, the coupled thermal-acoustic oscillation drastically increases and reaches to the limit cycle regime, which results in an unstable regime of the swirling flame. Both the acoustic pressure (MC, ME) and heat release rate signals present relatively large amplitudes than those in the stable regime, and distinct frequency could be

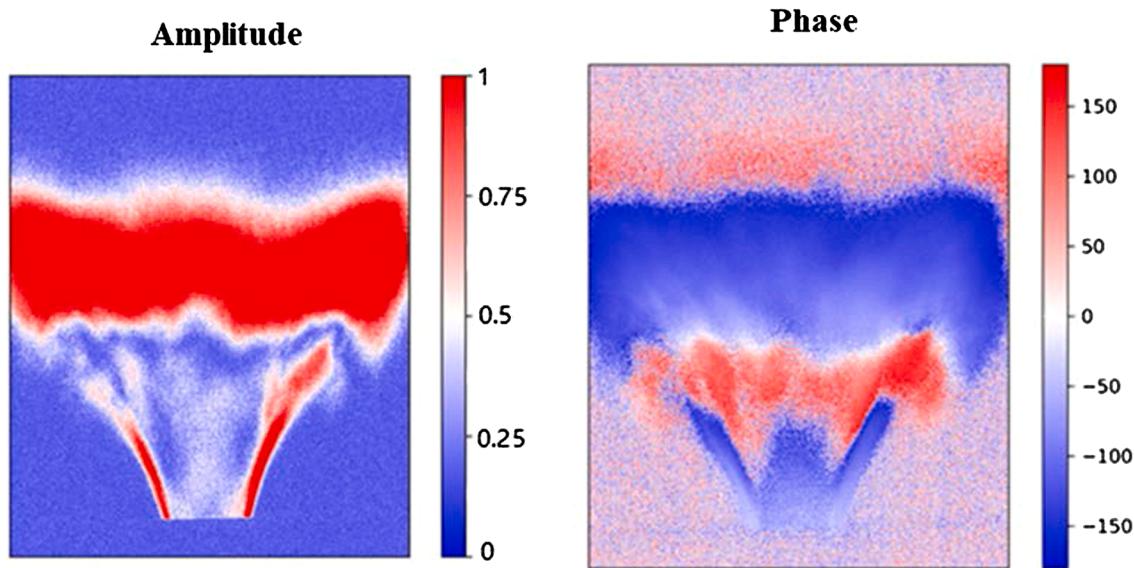
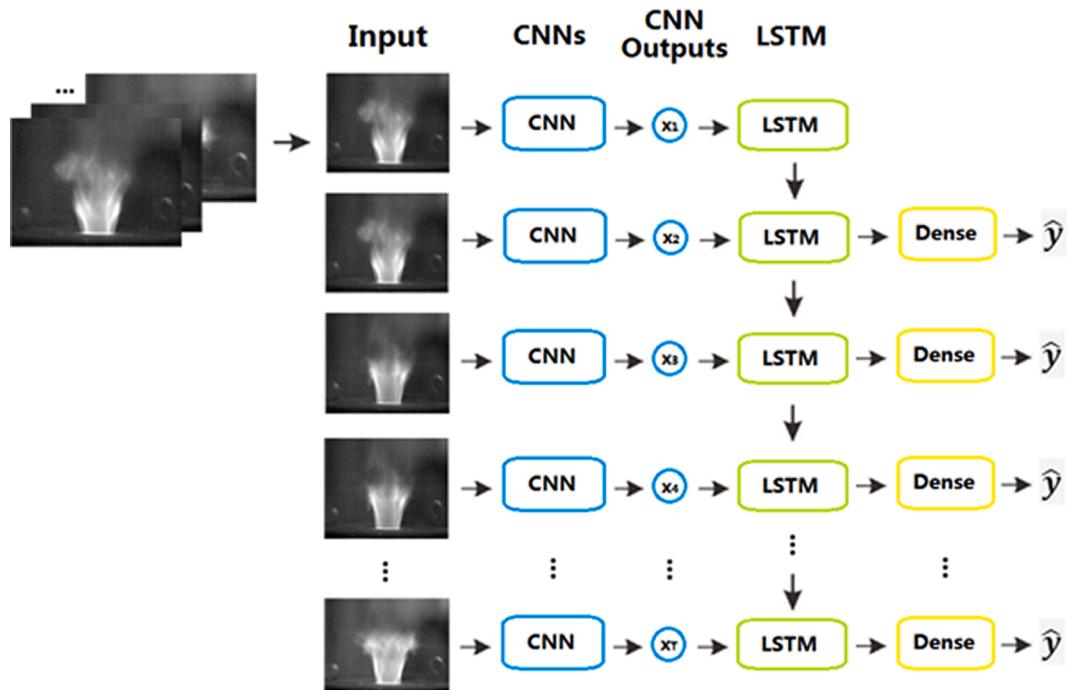
Fig. 6. DMD mode of  $f_Q$  in Case D.

Fig. 7. Detection model architecture, CNNnet-18 + LSTM.

detected.

Temporal kurtosis gives a clearer definition of thermoacoustic instability, which can be used for labeling. TK value mirrors the variation of the temporal characteristic of the acoustic pressure signals. An example of stable and unstable regimes defined by TK value corresponding to acoustic pressure is shown in Fig. 4. For signals in steady combustion, TK hovers around 3.0. If the combustion process enters the highly unstable regime, the limit cycle oscillation leads to the stabilization of TK slightly above 1.5. TK of a discrete time-domain signal  $p(t)$  for a given time duration is described as:

$$TK = \frac{\frac{1}{n} \sum_{i=1}^n (p_i - \bar{p})^4}{\left[ \frac{1}{n} \sum_{i=1}^n (p_i - \bar{p})^2 \right]^2}$$

where  $\bar{p}$  is the mean value of signals in a frame and  $n$  corresponds to the sampling number.

As in typical post-processing treatment for CI analysis, we select 1000 frames of sequential images in each case and analyze the image frequency domain characteristics of flame through dynamic mode decomposition (DMD) [39]. The CPU cost of DMD processing is about 30 min each on a workstation with Intel Core (TM) i7 CPU. Fig. 5 shows the flame spectra obtained by DMD of Case C and D. The spectra of Case D exhibits two distinct modes. The unstable mode at frequency of

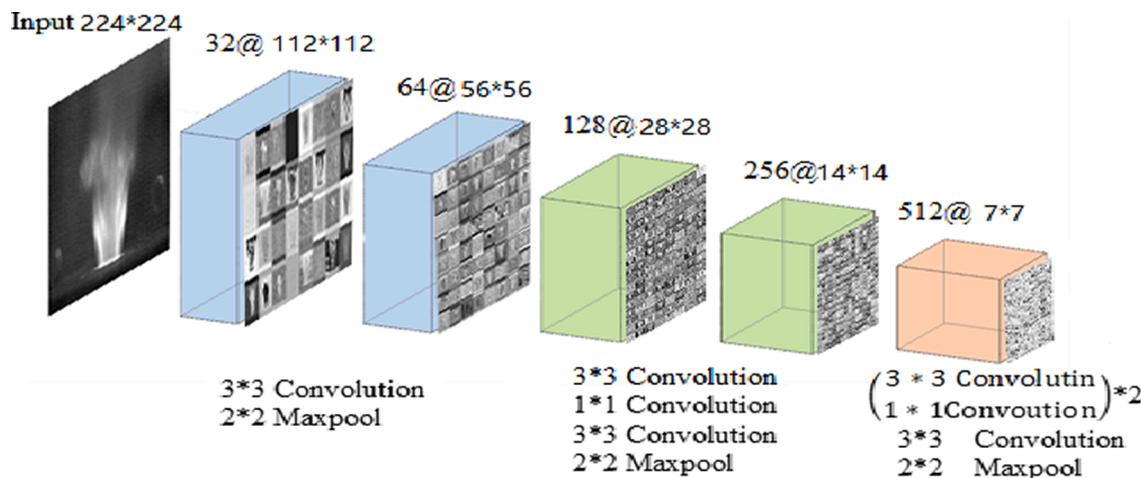


Fig. 8. Structure of convolutional neural network CNNnet-18.

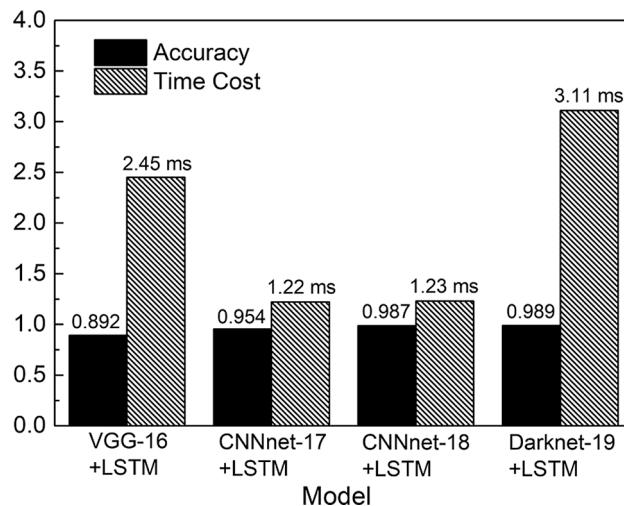


Fig. 9. Performance of different detection models.

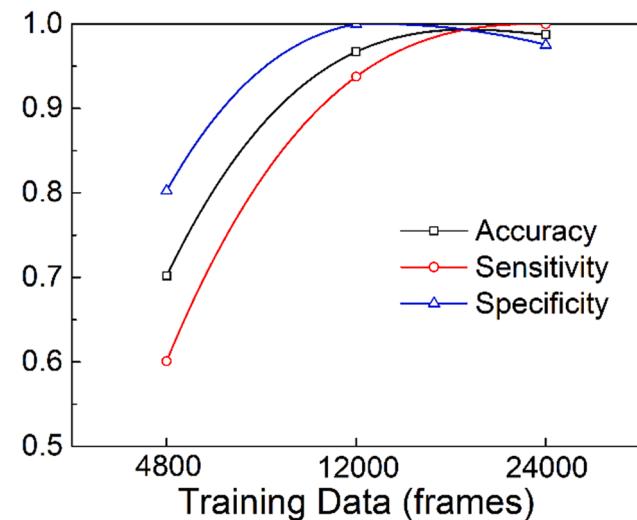


Fig. 10. Detection evaluation of CNNnet-18 + LSTM model with different amounts of flame training data.

**Table 2**  
Results of model detection.

	Detected as stable	Detected as unstable
Stable case	True Positive (TP)	False Negative (FN)
Unstable case	False Positive (FP)	True Negative (TN)

$f_Q = 480$  Hz is the quarter wave mode of combustion chamber [38]. The wavelength of the quarter wave mode is four times the length of the combustion chamber. For the combustor system with an open exit boundary, this mode represents the first resonant mode of the combustor. A Processing Vortex Core (PVC) mode has also been found at frequency of  $f_{\text{PVC}} = 802$  Hz, showing typical flow instability of swirling flows but with much low amplitude. Moreover, the mode phase of Case D is also shown in Fig. 6.

### 3. Deep learning-based detection model

In this section, we introduce the detection model using high-speed image sequence. Fig. 7 shows the model architecture combining CNN and LSTM, including input layer, CNN layer, LSTM layer, dense layer, and output layer. Each image in a continuous flame sequence is used as input into the CNN layer, where convolutional feature extractor is used to extract image spatial features in the flame image. Since CI is a

dynamic process, it is also important to capture the evolution of spatial features. Hence, we build an additional LSTM network on top of the CNN layer, using convolution outputs as sequential inputs. By combining CNN and LSTM, the detection model extracts both spatial features and temporal correlations from the input image sequences. Finally, the proposed model converts CI detection into binary classification (stable or unstable) and outputs the results of stable or unstable state detection. Model details are described in the following sections.

#### 3.1. Convolutional neural network

CNN is a deep learning network for complex nonlinear modeling, especially in handling multi-dimensional data such as images. For an image  $\mathbf{I}$ , 2D convolution operation is defined as follows:

$$\mathbf{H}_F(k, i, j) = \sum_{u=-U}^U \sum_{v=-V}^V \mathbf{W}(k, u, v) \cdot \mathbf{I}(i+u, j+v) + \mathbf{b}(k)$$

where  $(i, j)$  is the pixel coordinate vector.  $\mathbf{I}(i+u, j+v)$  is image intensity at  $(i+u, j+v)$ .  $\mathbf{W}$  is the convolution kernel weight, and  $\mathbf{b}$  is the convolution kernel bias, both of which are learned from training dataset in the training process.  $k$  is the number of convolution kernels and  $\mathbf{H}_F(i, j)$  is the convolution kernel feature at the position  $(i, j)$ .

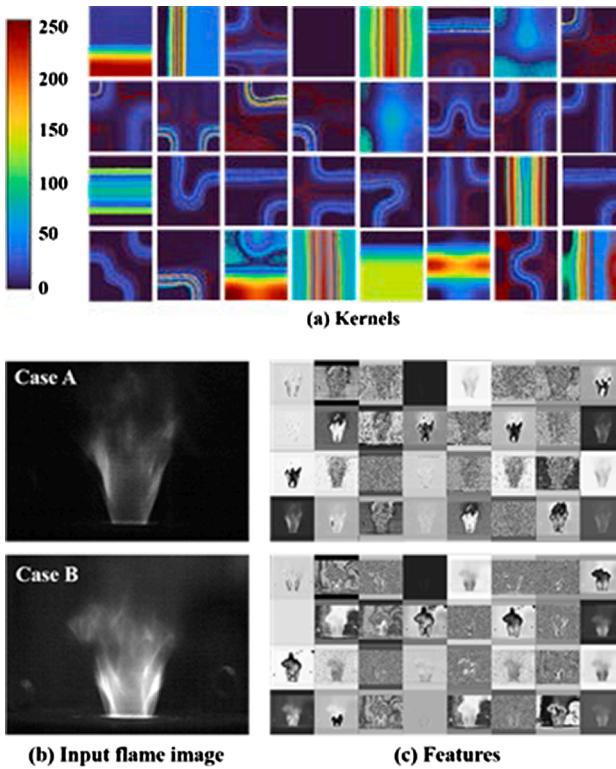


Fig. 11. Convolutional kernels and features of the first layer.

In the training process, the input distribution of each layer is updated with the learning process of back propagation, which can cause the problem called internal covariant transformation. Thus, Batch Normalization (BN) [40] is used to stabilize training processes. Maximum pooling, a nonlinear down-sampling operation is applied to reduce the data dimension while also maintaining the invariability of image rotation, translation and expansion. In this paper, we build CNN using Darknet-19 [41] as the backbone network structure. Darknet-19 is a light-weight neural network with no dependencies, excellent portability, high accuracy and excellent real-time performance.

The detection network model adopts continuous convolution and pooling layers. After the first layer of convolution, 32 convolution kernels of size  $3 \times 3$  extract 32 feature maps from the input flame image. Then, pooling windows with the size of  $2 \times 2$  are used for down sampling to obtain feature maps of size  $112 \times 112$  pixels and input them to the next. The convolution and pooling layer are alternately carried out. To prevent overfitting, a trick called dropout [42] is used, which randomly deactivates some neurons during the training process. At the end of CNN, we add an extra convolution layer with 1024 kernels to convert the feature map back to the hidden representation. In each layer, the number of feature maps increases with the decrease of spatial resolution.

### 3.2. Long Short-Term memory neural network

LSTM introduces a new internal state  $c_t \in R^D$ , also called memory cell state, for linear recurrent information transfer, and nonlinear hidden state  $h_t \in R^D$  of hidden layer at the same time. The unit of current moment receives both feature information from previous hidden state  $h_{t-1}$  and the output feature from underlying CNN to generate a candidate state  $\tilde{c}_t \in R^D$ :

$$\tilde{c}_t = \tanh(W_x^c x_t + W_h^c h_{t-1} + b_c)$$

where  $W_x^c$  and  $W_h^c$  represent the weight matrices that connect  $x_t$  and  $h_{t-1}$ ,

respectively. And the term  $b_c$  represents bias. The weight matrices  $W$  and biases  $b$  are the parameters obtained by training.

LSTM introduces gating mechanism to control information transmission. The gate operation includes sigmoid function  $\sigma(\cdot)$  and convolution. The first forget gate  $f_t$  will control how much information needs to be forgotten, thus, invalid or outdated information will not interfere with the detection:

$$f_t = \sigma(W_x^f x_t + W_h^f h_{t-1} + b_f)$$

Input gate  $i_t$  is used to decide how much information from candidate status of current time needs to be saved:

$$i_t = \sigma(W_x^i x_t + W_h^i h_{t-1} + b_i)$$

Output gate  $o_t$  determines how much information of internal state needs to output to external state at the time:

$$o_t = \sigma(W_x^o x_t + W_h^o h_{t-1} + b_o)$$

Then we compute internal state  $c_t$  and hidden state  $h_t$ :

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$$

$$h_t = o_t * \tanh(c_t)$$

where  $\odot$  represents the element-wise product with the gate value.

Long-distance timing dependent relationship is established for the whole network. Given an input image sequence, LSTM computes a sequence of internal states  $(c_1, \dots, c_n)$  and hidden states  $(h_1, \dots, h_n)$ .

To calculate the probability of each state  $S = \{\text{stable, unstable}\}$ , the softmax function is applied to analyze the output of the dense layer. Finally, the model outputs the detection result by comparing  $\hat{y}_t$ , a vector contains the probability of stable (PRS) and the probability of unstable (PRU), as follows:

$$y_t = W_o h_t$$

$$\text{PRS} = \text{softmax}(y_t(s)) = \frac{\exp(y_t(s))}{\sum_{r \in S} \exp(y_t(r))}$$

$$\text{PRU} = \text{softmax}(y_t(u)) = \frac{\exp(y_t(u))}{\sum_{r \in S} \exp(y_t(r))}$$

$$\hat{y}_t = [\text{PRS}, \text{PRU}]$$

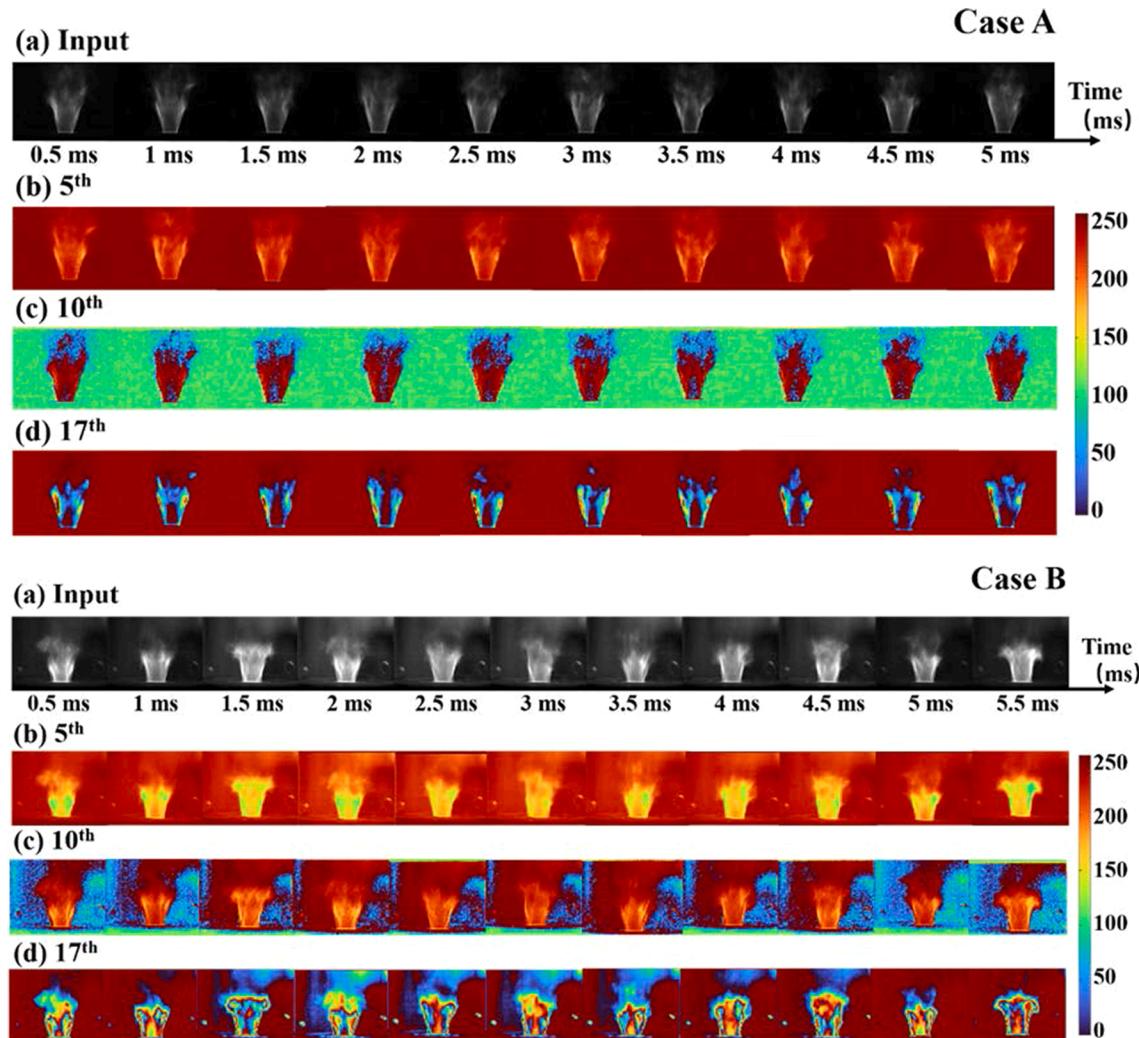
We implement our CNNnet-18 + LSTM model using Python 3.6 and TensorFlow1.12. The model is trained on the NVIDIA GPU 1050Ti platform. The training set expresses the category of each item using One-Hot, which is a target vector  $\hat{Y}_t$  used to represent the identity of the corresponding item. For stable items,  $\hat{Y}_t = [1, 0]$ . For unstable items,  $\hat{Y}_t = [0, 1]$ . Cross entropy [43] is applied as the loss function in training. In a binary problem, the loss function can be defined as follows:

$$\text{Loss} = -[\hat{Y}_t(s) \cdot \log(\text{PRS}) + \hat{Y}_t(u) \cdot \log(\text{PRU})]$$

During the training process, Adam optimizer of stochastic optimization method [44] is used to minimize cross entropy. Training is conducted on the entire training dataset, Dataset 2. At the same time, we monitor the real-time loss and accuracy. In this paper, the training process repeated three times with different random seeds, and the optimal model parameters with the minimal loss and highest accuracy on Dataset 2 were stored.

## 4. Result and discussion

In this section, we first discuss the rationality of model architecture for CI detection and visualize features in the detection process. Then we test and discuss model performance under various combustion conditions.



**Fig. 12.** Visualization of the first layer features evolutions, from left to right are the input image sequence, the 5th, the 10th, and the 17th convolution kernel.

#### 4.1. Discussions on model architecture

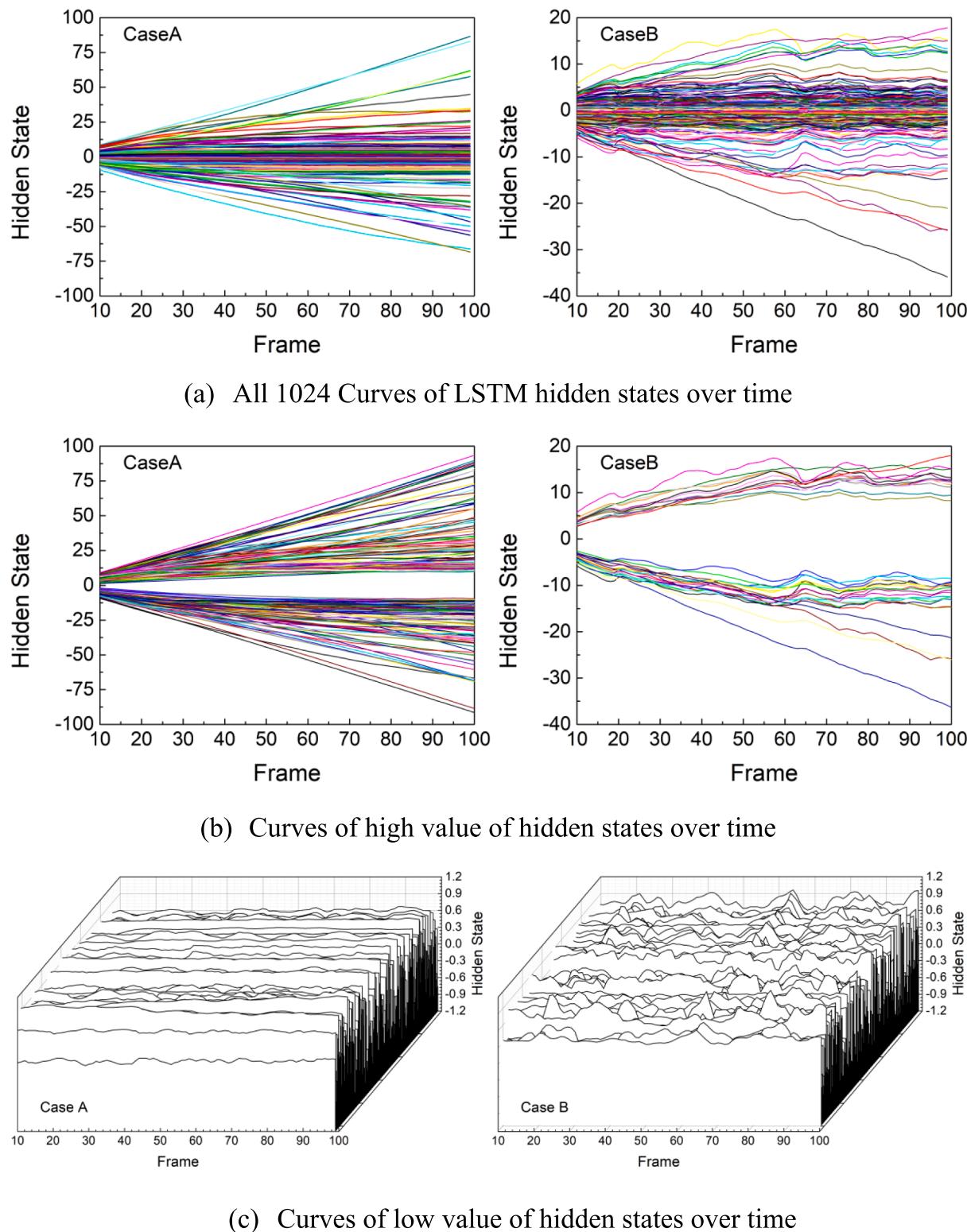
##### 4.1.1. Effects on complexity of CNN

The ability of CI detection is highly dependent on the model architecture. The complexity of CNN layers, including depth and layer arrangement, will have a direct impact on detection performance. To study such impact, CNN networks of different complexity have been built and compared in Fig. 9. The VGG-16 has a more complex layer arrangement than the CNNnet-16 with the same depth. Each model is trained on Dataset 2 and has repeated training processes three times. The optimal model parameters with the highest accuracy on training Dataset 2 are stored, and then test these models on Dataset 4. A correctly detected frame means that the stable or unstable classification detection result from the model is the same as the corresponding label. Accuracy is then defined as the proportion of correctly detected frames over the total number of images. The results show that increasing the complexity can help improve the CI detection accuracy, but at the same time, cause the increase of parameters, thus reducing the efficiency and slowing down the detection speed. As the depth of CNN increases, particularly from CNNnet-18 to CNNnet-19, the accuracy does not increase significantly, but the time cost increases enormously. Meanwhile, CNNnet-18 + LSTM achieves a high accuracy while maintaining a detection speed close to CNNnet-17 + LSTM. Thus, CNNnet-18 + LSTM shows a good balance between detection accuracy and time cost.

##### 4.1.2. Convergence assessment of the size of training data

The size of training data is also related to CI detection capability. According to the model structure and computing ability, the amount of flame training data should not be too small or too large. Here we evaluate model detection performance using various amounts of training data from the training dataset (Dataset 2) to train the model CNNnet-18 + LSTM and then test the model on Dataset 4. In order to illustrate the performance of the dataset in CI detection more clearly, two evaluation parameters, sensitivity and specificity, are introduced. Table 2 demonstrates the classification of detection results of the instability detection model. Sensitivity is defined as the proportion of the image whose actual detection result is stable to all stable input image flame, that is,  $TP/(TP + FN)$ , which measures the model's recognition ability of stable combustion flame. On the other hand, specificity was defined as the proportion of the unstable image with the actual detection result in all unstable input flame images, that is,  $TN/(TN + FP)$ , which measures the model's ability to recognize the unstable flame image.

Test result of convergence is shown in Fig. 10. The model trained with half the data of Dataset 2 has achieved good results in the test results of Dataset 4. Considering the different working conditions of training dataset and test dataset, it is proved that there is no overfitting problem. The result confirms that the amount of training dataset has reached the required need for CI detection while using full Dataset 2.



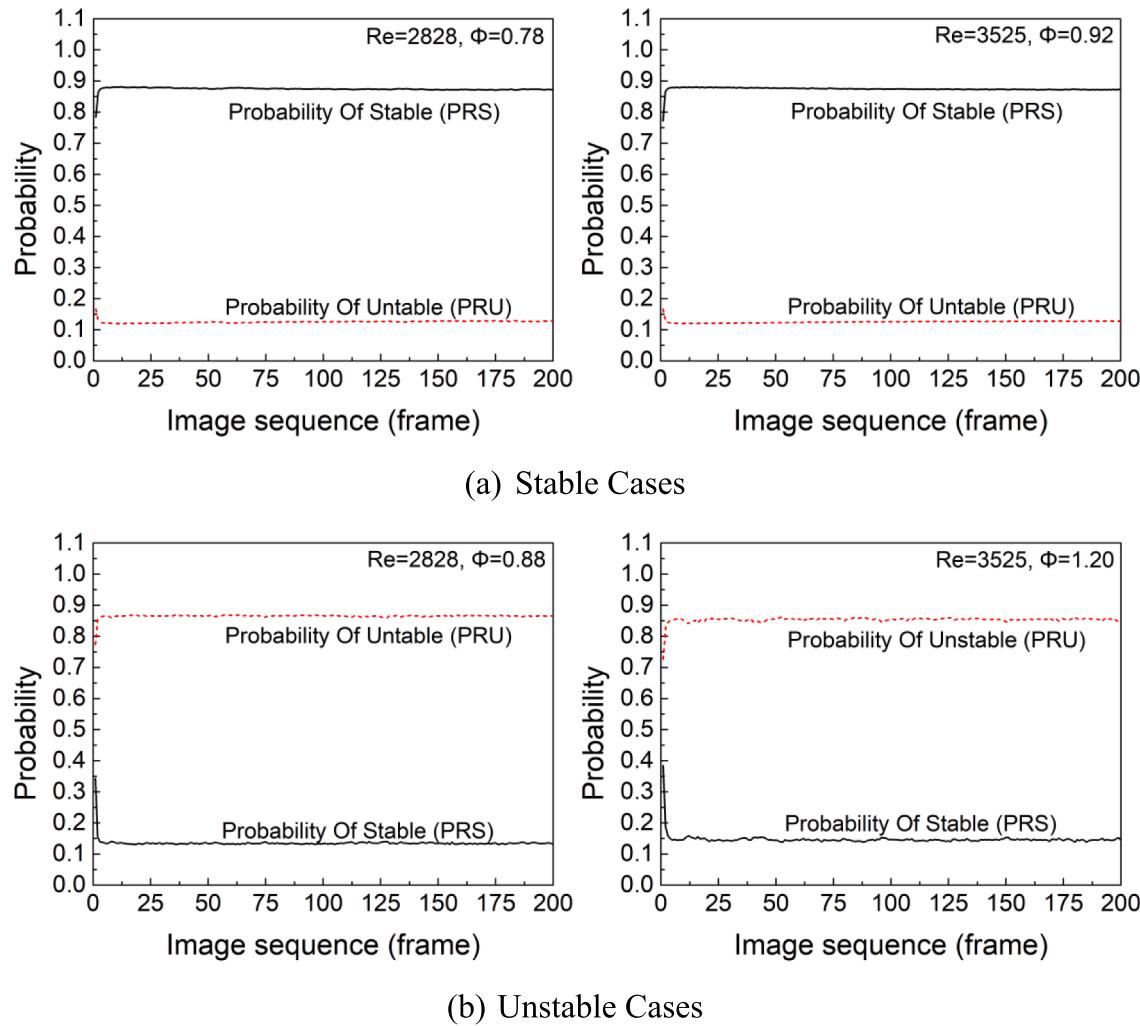
**Fig. 13.** Temporal evolutions of LSTM hidden states.

#### 4.2. Visualization of layer features

##### 4.2.1. Visualization of CNN layer

Given the input of flame image sequence, the deep learning model extracts spatial features using convolution layers and temporal features using LSTM layer. Here we will visualize the obtained spatial and temporal features in CI detection. Using the model described in Section 3,

different convolution kernels automatically extract different features, and features of each pooling layer are also given in Fig. 8. The first layer acts as a collection of various edge detectors, and the parameters of kernels are self-learned through training. Fig. 11 shows convolutional kernels and the features of the first layer corresponding to the input flame image. A color bar is added to these kernels corresponding to the value of convolution weight for ease of understanding. At that stage, the



**Fig. 14.** Detection processes of PRS & PRU on Validation Dataset 3.

activations respond to edge conjunctions, while retaining most of the information in the initial image. Significant differences can be found in the two cases. The kernels extracted not only the outer edges, but also the inner contours. It is worth noting that some features are blank. This means that the pattern encoded by the current kernel is not found in the input image. This also shows that CNN layer does not overfit training data, which indicates that the amount of training datasets is sufficient. Fig. 12 shows a part of visual outputs from the first layer. From left to right are the input image sequence, the 5th, the 10th, and the 17th convolution kernel. Some convolutional kernels, for example the 5th and the 10th, tend to delineate the outer contour, and a symmetrical V-type can be observed. In the flame image sequences, the stable and unstable flame structures are different. For example, CNN can observe that the flame angle and height of the two are different. Some convolutional kernels, for example the 17th, give a clearer view of the interior with distinct divisions of the input flame image. The sensitivity of different areas in flame image is different. Kernels sensitive areas in both cases focus on the wall region. In the meanwhile, they are more sensitive to the inner zone in an unstable flame image, and more detailed features can be observed.

CNN can also play a role in background noise reduction, which makes CI detection more robust. In unstable combustion, the noise level is large. The 5th convolution kernel extracts flame from the background. The 10th and 17th convolutional kernels extract flame while retaining the noise. The noise caused by the reflection of the glass wall can be clearly seen on the 10th. The combination of these kernels can

distinguish flame structures and background noise. It should be noted that CNN does not have the concept of flame, noise, etc. It learns and denoises by extracting and comparing spatial features in input images.

The first layer of CNN has a relatively shallow understanding of stable and unstable flames. Depth of neural networks plays an important role in performance. As CNN layers go deeper, they start encoding higher-level concepts. Higher-up presentations are very abstract, they carry increasingly less information about the visual contents of the image, and increasingly more information related to the instability detection of the image. By extracting the features of stable and unstable flame images, the CNN layer learns instability at the image level and provides spatial feature input to the later LSTM to capture evolution.

#### 4.2.2. Visualization of LSTM layer

Fig. 12 also shows the evolution of the first layer features in the input image sequence. The flame structure changes with time. As mentioned before, LSTM's function is to capture temporal coherence. At the end of CNN, one convolution layer with 1024 kernels generates 1024 feature maps, and feeds them into the LSTM cell. Unlike single image processing in the CNN layer, LSTM layer calculates information of each image with previous information being considered. Finally, each image in the sequence gets a hidden state vector  $h_t$  containing 1024 values. This vector provides the quantifiable parameters of the features in the previous section. Each value in the current state is plotted according to the sequence of image inputs, 1024 hidden state change curves have been obtained, shown in Fig. 13.

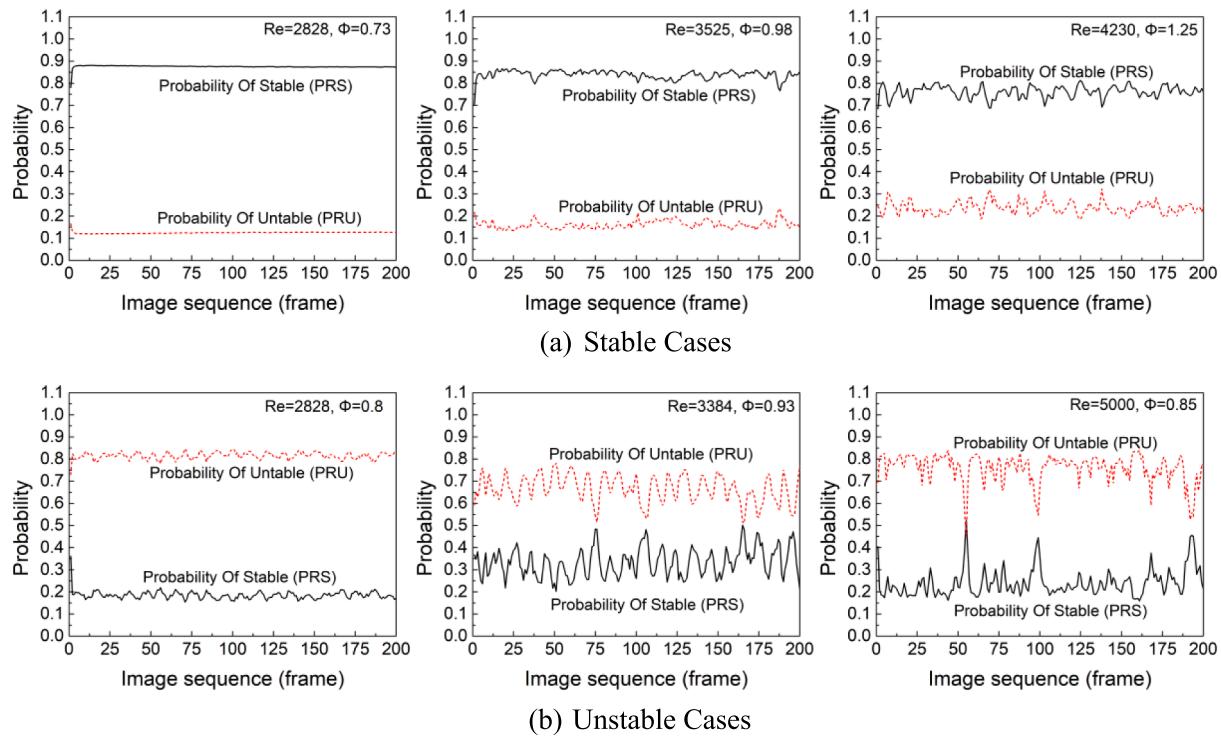


Fig. 15. Detection processes of PRS &amp; PRU on Datasets 1 and 4.

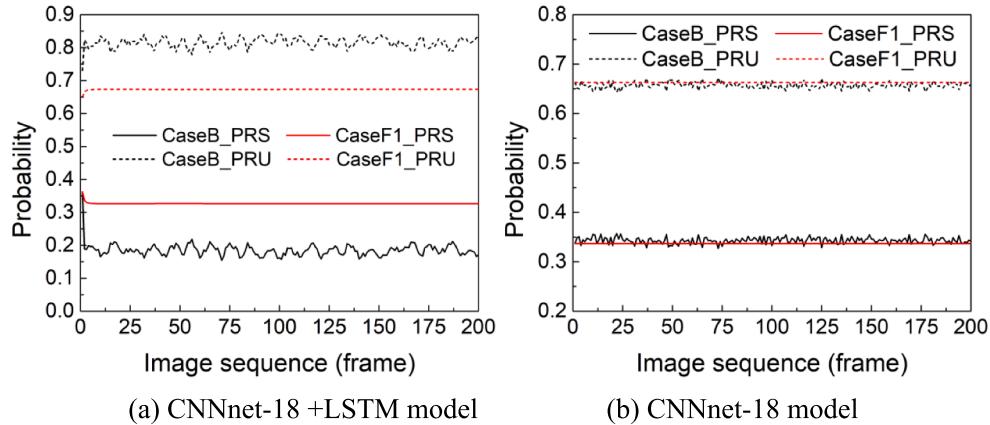


Fig. 16. Detection processes of Cases B and F1 on model CNN and CNN + LSTM.

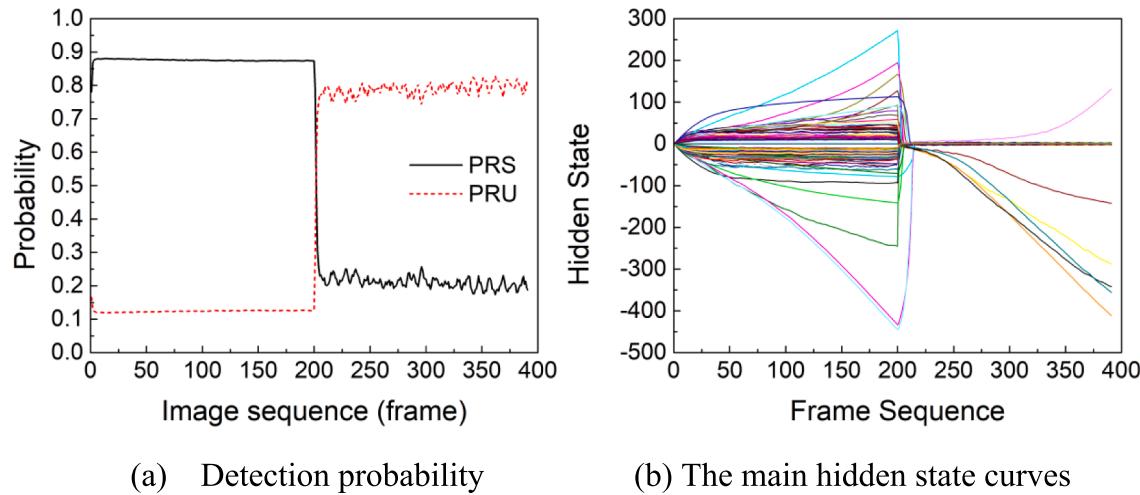
By comparing stable and unstable curves in Fig. 13a, we can observe that curves of stable flame are smoother and more uniform distribution while more fluctuation of unstable flame. The absolute value of feature states increases because information accumulates over time. The number of curves was also counted in different ranges. Fig. 13b shows curves of high value states over time, where the high value threshold is set as 10. It is related to the background of flame image, so the information is always cumulative. The stable flame has 230 curves, while unstable flame only has 34 curves. This indicates that more hidden state values in unstable cases tend to discard previous memories during unstable detection processing. Fig. 13c shows more obvious fluctuating behavior in unstable Case B. The coordinate axes are the frame number of the flame image sequence, the serial number of the state change curve and the corresponding state value of the current frame. In unstable combustion, the fluctuation of state curve is related to the change of flame structure, such as the change of flame angle. The curve of unstable case has a period along the direction of information accumulation, which

indicates certain periods during the evolution of feature states. By learning temporal evolution of stable and unstable flame image sequences, the LSTM layer learns the combustion instability detection at the image sequence level.

#### 4.3. Detection results

##### 4.3.1. Validation of detection process

After training the proposed model using flame image sequences, we can obtain the detection results, i.e., stable or unstable, by comparing the probabilities of being stable and unstable, as described in Section 3. We validate detection process using image sequences in Dataset 3. During the detection process, we will keep track of the detection probability of stable and unstable combustion cases as shown in Fig. 14. For stable combustion, the probability of stable from the model detection output is much higher than the probability of unstable. As for unstable combustion, the detection results are on the contrary, the probability of



**Fig. 17.** Temporal results in detection of a handcrafted transition case – Case F2.

unstable is greater than probability of stable. The model can effectively distinguish between stable and unstable combustion cases after a few frames when the experiment condition of flame image sequences is the same as that in training.

Practically, we need to consider the cases at Reynolds numbers and equivalence ratios different from the training data. This also raises concern about the generalization of the model. To this end, we conduct another test using image sequences under different combustion conditions, namely Datasets 1 and 4, with a total of 13 untrained conditions. Similarly, in the case of stable combustion, the stable probability is higher than the unstable probability. In the case of unstable combustion, the detection result is the opposite. The test result shows that it can still detect combustion instability, which shows the generalization ability of the model. However, the test result in Fig. 15 shows that the model performs better when conditions of test data are more similar to training data, for example the same Reynolds number with different equivalence ratios. When tested on scenarios that are very different with training data, for example a relatively high Reynolds number ( $Re = 5000$ ), the model can still produce correct results at most time steps, but the CI detection confidence is degraded and the detection probability curve has more fluctuation.

#### 4.3.2. Importance of LSTM

The model takes only a few frames to get the right detection results, as shown in Fig. 14 and Fig. 15, which represents the great role of CNN layer. In order to verify the importance of LSTM in learning flame dynamic evolution, a fake data (Case F1) is made. Case F1 is an image sequence segment obtained by selecting a single frame in Case B and repeating it 200 times. As shown in the detection process of Case B and F1 in Fig. 16a, the detection curve of case F1 maintains the detection probability of the initial frame, while the instability probability of case B becomes much larger than the stability probability after a few frames. It demonstrates that the proposed model detects CI by learning the dynamic changes in the image sequence, rather than just learning the spatial features of an image. To further verify the role played by the LSTM, a standard CNN model CNNnet-18 without LSTM was constructed and trained on Dataset 2. Its detection processes for Cases B and F1 in Fig. 16b are basically the same, although the detection curve of case B has inconspicuous fluctuation. This is because a standard 2D CNN model is more sensitive to spatial features of the flame image. Fig. 16 shows that LSTM does greatly improve detection confidence by leveraging the temporal data dependencies. This confirms that LSTM captures the dynamics of the flame, which helps distinguish between stable and unstable combustion.

#### 4.3.3. Transient detection from stable to unstable regime

Now that we have demonstrated the ability of LSTM in CI detection, we wish to further verify its capacity to detect and track transitions from stable to unstable regime as well. Case F2 is then made to test. It is a joint sequence to simulate the transition by combining a first 200 frames randomly chosen from stable (Case A) and later 200 frames from unstable (Case B). The result is shown in Fig. 17a. After the 200th frame, the probability of stable decrease sharply and the probability of unstable increases immediately. This shows that it can detect the transition in a very short time. Fig. 17b shows the curves of activated new states and deactivated ones which exhibit drastic changes during the transition. Some new status values were activated, while the previous ones deactivated. This indicates that LSTM has delayed response to clear the preceding memory and record new information. Although the Case F2 is fake, it also shows that the model can play a role in detecting transitions.

#### 4.3.4. Statistical results

Model accuracy was evaluated in the image sequences on Dataset 4. The proposed model CNNnet-18 + LSTM got a high accuracy of 98.72%. Meanwhile, the detection was conducted in real-time on NVIDIA GPU 1050Ti, and the average computing time per frame was just 1.23 ms per frame. It is worth mentioning that processing time could be further reduced on GPU with higher capacity. Sensitivity and specificity are used to measure the model's recognition ability of stable and unstable flame image respectively. The sensitivity of the model reached 99.99%, which indicates less omission errors. The specificity of the model reached 97.50%. The results show that the model can effectively detect stable and unstable combustion.

## 5. Conclusion

In this paper, we propose a deep learning model for the detection of thermoacoustic instability based on high-speed flame image sequences. The flame chemiluminescence images provide more information about flame structures than the acoustic signals. The datasets are collected from the propane-air premixed swirling flames under different Reynolds number and equivalent ratio, labeling with stable and unstable regimes indicated by Temporal Kurtosis analysis of the measured transient OH\* and acoustic signals. We have discussed the rationality for building model by studying how the performance and convergence are affected by the complexity of CNN layer and the size of training data. Our proposed method eventually achieves high accuracy (98.72%), sensitivity (99.99%) and specificity (97.50%), with a short processing time about 1.23 ms per frame on a commercial GPU adapter which can be considered for real-time identification.

The hidden features of CNN and LSTM are extracted and visualized to explain the ability of accurate detection of combustion instability. Deep CNN layers used in our deep learning model extract various spatial features from flame images by 1024 convolution kernels. The evolutions of spatial features obtained in stable and unstable images are different, which helps distinguish between the stable and unstable flame structures at the level of image sequences. The jointly learned CNN and LSTM model is more efficient than a standard CNN model, as the former could learn the temporal evolution of the flame image sequences. The generalization ability of the model is tested under 13 untrained experimental conditions with a satisfactory outcome. The higher the similarity with the training experimental conditions, the better the model detection performance. Furthermore, the segment of simulating transition proves that the model has the ability to detect transition from stable to unstable regime.

In future research, more experimental data and more advanced neural networks will be introduced to improve the performance and versatility of combustion instability detection. The proposed model will be further extended to other combustors through transfer learning to test its generalizability.

## 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.

## Acknowledgements

This study is financially supported by the National Natural Science Foundation of China (91841302 , 51976184). The authors gratefully thank Prof. Givi Peyman (from University of Pittsburgh) for fruitfully discussions and kindest helps in this work.

## References

- [1] Rayleigh. The explanation of certain acoustical phenomena. *Nature* 1878;18: 319–21.
- [2] Durox D, Schuller T, Noiray N, Birbaud AL, Candel S. Rayleigh criterion and acoustic energy balance in unconfined self-sustained oscillating flames. *Combust Flame* 2009;156(1):106–19. <https://doi.org/10.1016/j.combustflame.2008.07.016>.
- [3] Kim KT. Combustion instability feedback mechanisms in a lean-premixed swirl-stabilized combustor. *Combust Flame* 2016;171:137–51. <https://doi.org/10.1016/j.combustflame.2016.06.003>.
- [4] Huang Y, Yang V. Dynamics and stability of lean-premixed swirl-stabilized combustion. *Prog Energy Combust Sci* 2009;35(4):293–364. <https://doi.org/10.1016/j.pecs.2009.01.002>.
- [5] Poinsot T. Prediction and control of combustion instabilities in real engines. *Proc Combust Inst* 2017;36(1):1–28. <https://doi.org/10.1016/j.proci.2016.05.007>.
- [6] Schuller T, Durox D, Candel S. Self-induced combustion oscillations of laminar premixed flames stabilized on annular burners. *Combust Flame* 2003;135:525–37. <https://doi.org/10.1016/j.combustflame.2003.08.007>.
- [7] Mukhopadhyay A, Sen S, Basu DN, Mondal S. Dynamics and Control of Energy Systems. 2020.
- [8] Richards GA, Straub DL, Robey EH. Passive control of combustion dynamics in stationary gas turbines. *J Propuls Power* 2003;19(5):795–810. <https://doi.org/10.2514/2.6195>.
- [9] Noiray N, Durox D, Schuller T, Candel S. Dynamic phase converter for passive control of combustion instabilities. *Proc Combust Inst* 2009;32 II:3163–70. doi: 10.1016/j.proci.2008.05.051.
- [10] Noiray N, Durox D, Schuller T, Candel S. Passive control of combustion instabilities involving premixed flames anchored on perforated plates. *Proc Combust Inst* 2007; 31(1):1283–90. <https://doi.org/10.1016/j.proci.2006.07.096>.
- [11] Kobayashi T, Murayama S, Hachijo T, Gotoda H. Early detection of thermoacoustic combustion instability using a methodology combining complex networks and machine learning. *Phys Rev Appl* 2019;11:1. <https://doi.org/10.1103/PhysRevApplied.11.064034>.
- [12] Schadow KC, Gutmark E. Combustion instability related to vortex shedding in dump combustors and their passive control. *Prog Energy Combust Sci* 1992;18(2): 117–32. [https://doi.org/10.1016/0360-1285\(92\)90020-2](https://doi.org/10.1016/0360-1285(92)90020-2).
- [13] Zhao D, Morgans AS. Tuned passive control of combustion instabilities using multiple Helmholtz resonators. *J Sound Vib* 2009;320(4-5):744–57. <https://doi.org/10.1016/j.jsv.2008.09.006>.
- [14] Oztarlik G, Selle L, Poinsot T, Schuller T. Suppression of instabilities of swirled premixed flames with minimal secondary hydrogen injection. *Combust Flame* 2020;214:266–76. <https://doi.org/10.1016/j.combustflame.2019.12.032>.
- [15] McManus KR, Poinsot T, Candel SM. A review of active control of combustion instabilities. *Prog Energy Combust Sci* 1993;19(1):1–29. [https://doi.org/10.1016/0360-1285\(93\)90020-F](https://doi.org/10.1016/0360-1285(93)90020-F).
- [16] Roy A, Mondal S, Pawar SA, Sujith RI. On the mechanism of open-loop control of thermoacoustic instability in a laminar premixed combustor. *J Fluid Mech* 2019. <https://doi.org/10.1017/jfm.2019.884>.
- [17] Morgans AS, Stow SR. Model-based control of combustion instabilities in annular combustors. *Combust Flame* 2007;150(4):380–99. <https://doi.org/10.1016/j.combustflame.2007.06.002>.
- [18] Morgans AS, Dowling AP. Model-based control of combustion instabilities. *J Sound Vib* 2007;299(1-2):261–82. <https://doi.org/10.1016/j.jsv.2006.07.014>.
- [19] Docquier N, Candel Sébastien. Combustion control and sensors: a review. *Prog Energy Combust Sci* 2002;28(2):107–50. [https://doi.org/10.1016/S0360-1285\(01\)00009-0](https://doi.org/10.1016/S0360-1285(01)00009-0).
- [20] Dowling AP, Morgans AS. Feedback control of combustion oscillations. *Annu Rev Fluid Mech* 2005;37(1):151–82. <https://doi.org/10.1146/annurev.fluid.36.050802.122038>.
- [21] Song WJ, Che DJ. Temporal kurtosis of dynamic pressure signal as a quantitative measure of combustion instability. *Appl Therm Eng* 2016;104:577–86. <https://doi.org/10.1016/j.applthermeng.2016.05.094>.
- [22] Shanbhogue SJ, Sanusi YS, Taamallah S, Habib MA, Mokheimer EMA, Ghoniem AF. Flame macrostructures, combustion instability and extinction strain scaling in swirl-stabilized premixed CH<sub>4</sub> /H<sub>2</sub> combustion. *Combust Flame* 2016; 163:494–507. <https://doi.org/10.1016/j.combustflame.2015.10.026>.
- [23] Lapeyre CJ, Misdariis A, Cazard N, Veynante D, Poinsot T. Training convolutional neural networks to estimate turbulent sub-grid scale reaction rates. *Combust Flame* 2019;203:255–64. <https://doi.org/10.1016/j.combustflame.2019.02.019>.
- [24] Yellapantula S, Perry BA, Grout RW. Deep learning-based model for progress variable dissipation rate in turbulent premixed flames. *Proc Combust Inst* 2021;38 (2):2929–38. <https://doi.org/10.1016/j.proci.2020.06.205>.
- [25] Nikolaou ZM, Chrysostomou C, Minamoto Y, Vervisch L. Evaluation of a neural network-based closure for the unresolved stresses in turbulent premixed V-flames. *Flow, Turbul Combust* 2021;106(2):331–56. <https://doi.org/10.1007/s10494-020-00170-w>.
- [26] Wan K, Barnaud C, Vervisch L, Domingo P. Chemistry reduction using machine learning trained from non-premixed micro-mixing modeling: application to DNS of a syngas turbulent oxy-flame with side-wall effects. *Combust Flame* 2020;220: 119–29. <https://doi.org/10.1016/j.combustflame.2020.06.008>.
- [27] An J, He G, Luo K, Qin F, Liu B. Artificial neural network based chemical mechanisms for computationally efficient modeling of hydrogen/carbon monoxide/kerosene combustion. *Int J Hydrogen Energy* 2020;45(53):29594–605. <https://doi.org/10.1016/j.ijhydene.2020.08.081>.
- [28] Wan K, Hartl S, Vervisch L, Domingo P, Barlow RS, Hasse C. Combustion regime identification from machine learning trained by Raman/Rayleigh line measurements. *Combust Flame* 2020;219:268–74. <https://doi.org/10.1016/j.combustflame.2020.05.024>.
- [29] An J, Wang H, Liu B, Luo KH, Qin F, He GQ. A deep learning framework for hydrogen-fueled turbulent combustion simulation. *Int J Hydrogen Energy* 2020;45 (35):17992–8000. <https://doi.org/10.1016/j.ijhydene.2020.04.286>.
- [30] Yam L, Yoshua B. Convolutional Networks for Images, Speech, and Time-Series. In: Arbib MA, editor. Handb. brain theory neural networks, MIT Press; 1995, p. 255–8.
- [31] Sarkar S, Lore KG, Sarkar S. Early detection of combustion instability by neural-symbolic analysis on hi-speed video. *CEUR Workshop Proc* 2015;1583.
- [32] Sarkar SGAB. High Speed Video-based health monitoring using 3D Deep Learning. *Dyn Data-Driven Appl Syst* 2017:3–6.
- [33] Choi O, Choi J, Kim N, Lee MC. Combustion instability monitoring through deep-learning-based classification of sequential high-speed flame images. *Electron* 2020; 9(5):848. <https://doi.org/10.3390/electronics9050848>.
- [34] DiPietro R, Hager GD. Deep learning: RNNs and LSTM. Elsevier Inc.; 2019. DOI: 10.1016/B978-0-12-816176-0-00026-0.
- [35] Hochreiter S, Schmidhuber Jürgen. Long short-term memory. *Neural Comput* 1997;9(8):1735–80. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- [36] Durox D, Prieur K, Schuller T, Candel S. Different flame patterns linked with swirling injector interactions in an annular combustor. *J Eng Gas Turbines Power* 2016;138:1–8. <https://doi.org/10.1115/1.4033330>.
- [37] Schuller T, Durox D, Candel S. Dynamics of and noise radiated by a perturbed impinging premixed jet flame. *Combust Flame* 2002;128(1-2):88–110. [https://doi.org/10.1016/S0010-2180\(01\)00334-0](https://doi.org/10.1016/S0010-2180(01)00334-0).
- [38] Yang Y, Wang G, Fang Y, Jin T, Li J. Experimental study of the effect of outlet boundary on combustion instabilities in premixed swirling flames. *Phys Fluids* 2021;33(2):027106. <https://doi.org/10.1063/5.0038984>.
- [39] antares Development Team. Antares Documentation Release 1.17.0 2020. <https://cerfacs.fr/antares/>.
- [40] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. 32nd Int Conf Mach Learn ICML 2015 2015;1: 448–56.
- [41] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger. Proc - 30th IEEE Conf Comput Vis Pattern Recognition, CVPR 2017 2017;2017-Janua:6517–25. DOI: 10.1109/CVPR.2017.690.

- [42] Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR. Improving neural networks by preventing co-adaptation of feature detectors. *Comput Sci* 2012;3:212–23.
- [43] de Boer P-T, Kroese DP, Mannor S, Rubinstein RY. A tutorial on the cross-entropy method. *Ann Oper Res* 2005;134(1):19–67. <https://doi.org/10.1007/s10479-005-5724-z>.
- [44] Kingma DP, Ba JL. Adam: A method for stochastic optimization. 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., 2015, p. 1–15.