An Explainable Framework using Deep Attention Models for Sequential Data in Combustion Systems

Tryambak Gangopadhyay
Department of Mechanical Engineering
Iowa State University
Ames, IA 50011
`tryambak@iastate.edu`

Sin Yong Tan
Department of Mechanical Engineering
Iowa State University
Ames, IA 50011
`tsyong98@iastate.edu`

Anthony Locurto
Department of Mechanical Engineering
Iowa State University
Ames, IA 50011
`alocurto@iastate.edu`

James B. Michael
Department of Mechanical Engineering
Iowa State University
Ames, IA 50011
`jmichael@iastate.edu`

Soumik Sarkar
Department of Mechanical Engineering
Iowa State University
Ames, IA 50011
`soumiks@iastate.edu`

Abstract

The explanations provided by a classification framework on sequential data can provide insights to improve scientific understanding in different problems of physical sciences. We propose a model capable of explaining the contribution of individual image frames in the input sequence to generate a single prediction label for the sequence. We achieve this without compromising the accuracy. The performance of our model is demonstrated in a problem involving combustion instability, where explainability has not been explored before. We use a dataset of a combustion system comprising hi-speed flame videos and acoustic time series data. The input sequence of images is encoded using a layer of 2D CNN and 2 layers of stacked LSTMs. We use a temporal attention mechanism to capture the global temporal structure. The attention weights learned by the model highlight the significant image frames that are most relevant for each prediction. We validate the results from a domain knowledge perspective.

1 Introduction

Explanations provided by a model can help in verifying its performance as it not only generates the prediction, but also provides insights behind the prediction. This leads to enhanced understanding beyond the model accuracy. In many physical systems, critical transitions can occur in split seconds especially for combustion-dependent power generating systems like land-based gas turbine engines, jet engines for aviation and rocket engines. Detection of these transitions followed by interpreting the learning of the model can improve understanding of the critical transition which can lead to effective control of these systems.

Second Workshop on Machine Learning and the Physical Sciences (NeurIPS 2019), Vancouver, Canada.

In a combustion system, early detection of the critical transition is crucial [1]. The flow perturbations can cause fluctuations in heat release rate and result in the generation of sound waves. A positive feedback is established if the heat release rate fluctuations are in phase with the fluctuating acoustic pressure [2]. This happens when the combustion is taking place in a confined environment which can cause the sound waves to get reflected back and modify the heat release rate. The oscillations can grow and cause an intense growth of pressure fluctuations with high heat transfer on the combustor surfaces [3, 4]. The engines can develop large levels of vibration due to these oscillations which may lead to its structural damage and catastrophic failure resulting in huge revenue loss [5]. To study combustion instability, full-scale computational-fluid-dynamic (CFD) models, physics-based modeling, dynamic data-driven application methods have been applied [6, 7, 8, 9, 10, 11]. However, these approaches may have several restrictions including simplifying assumptions and hindering effects of combustion noise.

As the field of 3D computer vision is developing, deep sequence modeling tools are becoming a natural choice for perception problems with sequential structure and these methods need little input preprocessing and no hand-designed features [12, 13]. The concepts of deep learning have been applied to various tasks which include extracting meaningful features from the images [14, 15] using convolutional neural networks (CNN). Application of deep learning in studying combustion instability has only started recently. Applications include a deep learning-based framework to extract features from hi-speed flame images [16], a neural-symbolic framework to capture the temporal variation [17], an end-to-end convolutional selective autoencoder [18], a 3D CNN architecture [19] and an instability detection framework using a 2D CNN model [20]. However, the previous works have not concentrated on providing the explanations of the model predictions,

Correlating predictions back to the input data is difficult for deep learning models as these models have a large number of parameters and have a complex approach to extract and combine features. Considering only the accuracy of the model ignores the importance of interpretability in a machine learning model [21]. Most methods cannot provide an answer to why the model predicts a particular label and which part of the data are influential for a particular prediction [22]. Automatic Relevance Determination Regression (ARD) [23] explicitly determines the relevance in the data points. Our approach also explicitly learns and determines the importance of each data sample and uses that as explanations for the prediction. The explanations provided by our proposed model followed by validation from domain knowledge may be a significant step towards improving the scientific understanding of this problem. Using a temporal attention mechanism, this framework also generalizes better without compromising on the accuracy. In this paper, we first describe the dataset used and models proposed, followed by results and discussions with explanations validated using domain knowledge.

2 Experiments

2.1 Dataset

The experimental setup (Fig. 1) is a vertically-oriented open-ended Rijke tube setup with a stabilized flame as the heat source. Our dataset consists of hi-speed flame videos and acoustic time series data, simultaneously recorded for a duration of 12 secs by inducing instability in a combustion system with the variation of the acoustic length. We choose hi-speed videos corresponding to 5 different conditions in the laboratory-scale setup from domain knowledge demonstrating low, moderate and high instability. From the original 5000 frame/sec, we downsample the videos by 5 times (taking every 5 frames) and perform preprocessing (crop, resize) on the extracted image frames. By adopting a moving window approach, we split the entire acoustic time series data into consecutive windows of length 0.1s (100 ms). To classify the dataset into two classes (stable, unstable) we define an instability measure (IM) using the acoustic time series data. This signal-to-noise ratio is computed using two frequency ranges from the Fast Fourier Transform (FFT) plot of each time window using the equation below. The high amplitude zone is mostly in the range of (200 - 500) Hz and to estimate the energy content of the instability, we compute the sum of amplitudes in this range. The noise is estimated by taking average of the amplitudes corresponding to the frequency range of (2000 - 5000) Hz as this range is far away from the (200 - 500) Hz range. Using median as the threshold (1850), the dataset is split into two classes as shown in Fig. 1. Each IM value represents a time window and the sequence of images corresponding to that time window is correlated. Therefore, each point in Fig. 1

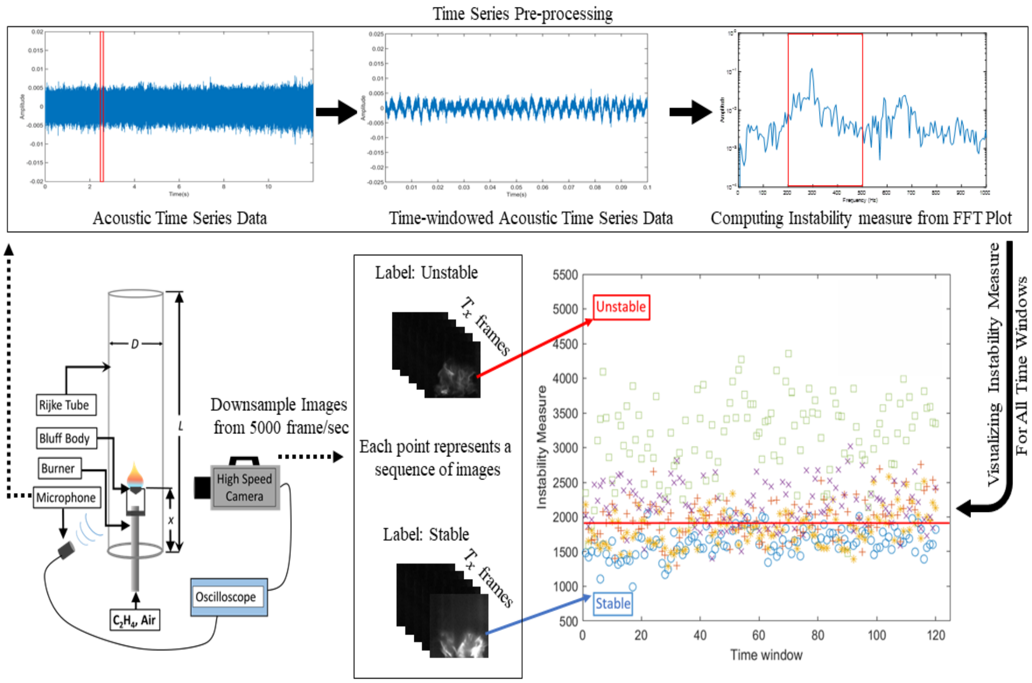

Figure 1: Separation of instability values into stable and unstable. Each point/marker in the plot represents a sequence of input images, and different marker represents the different conditions in the laboratory-scale setup.

corresponds to a class label of a particular image sequence.

$$IM = \frac{\sum \text{Amplitudes in the range 200-500 Hz}}{\text{Mean of the amplitudes in the range 2000-5000 Hz}}$$

2.2 Models

Our proposed model preserve both the spatial and temporal structure of the image sequence by encoding them using 2D-CNN and two stacked layers of LSTMs. The CNN model encodes each image in the input sequence into a 128-dimensional vector which acts as input for the LSTM layer on top of that. For neural machine translation, Bahdanau et al.[24] introduced the concept of soft temporal attention using the attention module fused between the encoder and decoding layers. In our proposed model, we don't use a decoder LSTM as we are performing many-to-one classification. Taking the sequence of hidden states as input, the attention block generates the context vector as shown in Fig. 2. Our baseline model (with only CNN and LSTMs) has the same structure and hyper-parameters as our proposed model except that it doesn't have the attention block. For the baseline model, the LSTM hidden state of the last timestep acts as input to the fully-connected layer.

3 Results

We present the comparison results of our proposed model and the baseline model (CNN-LSTMs) using different input sequence lengths in Table 1. Our approach demonstrates better performance than the CNN-LSTMs model. Taking sequential image frames as input, the model predicts a single class label for the entire input sequence as shown in Fig. 2. The attention weights highlight the most relevant images in generating that particular prediction (Fig. 3). From the results of the dataset, we observe that the model is not focusing only on the first (or last) few frames and some of the higher weighted images are distributed in the middle as well for some sequences. We validate the attention weights from domain knowledge perspective. We illustrate our validation approach in Fig. 3. Considering an image sequence, we divide the corresponding acoustic time series into 25 windowed

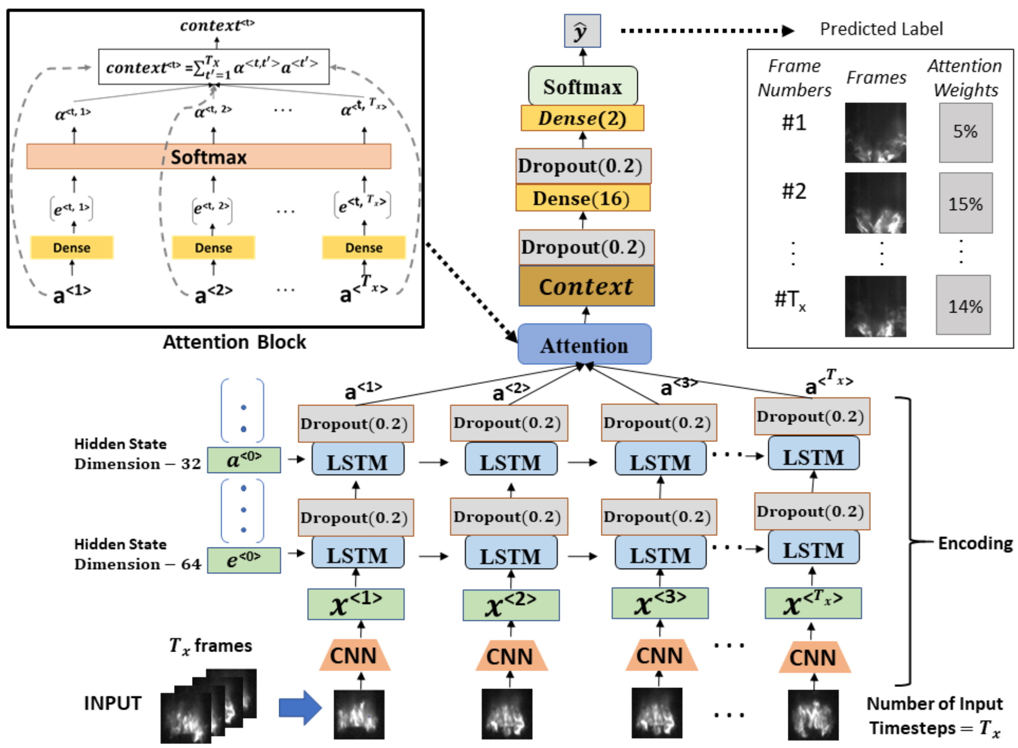

Figure 2: A complete overview of our proposed model. Attention block taking the hidden states of the 2nd LSTM as input to generate the context vector.

Model	Input Sequence Length	Training Accuracy(%)	Validation Accuracy(%)
Proposed Model	25	98.28	82.29
	50	97.08	84.58
CNN-LSTMs	25	88.49	79.79
	50	89.37	79.17

Table 1: Results comparison for our proposed model and baseline model (CNN-LSTMs).

time series surrounding each of the 25 timesteps. And from the FFT plot for each time window, we compute and plot the maximum amplitudes for all the timesteps. We observe that the distribution of attention weights shows similarity with that of the maximum amplitude (an important method to study combustion instability) and it shows that the model highlights most of the timesteps when the combustion system is actually unstable.

4 Conclusion

Explaining the predictions of a model can provide new insights and enhance our understanding of different problems in physical sciences exhibiting complex dynamics. In this paper, we propose a model to successfully predict the stable and unstable states of a combustion system. Along with detection, the model focuses on significant annotations of all the images in a sequence and the explanations provided by the model are also validated.

Acknowledgments

This work has been supported in part by the U.S. Air Force Office of Scientific Research under the YIP grant FA9550-17-1-0220. Any opinions, findings and conclusions or recommendations expressed

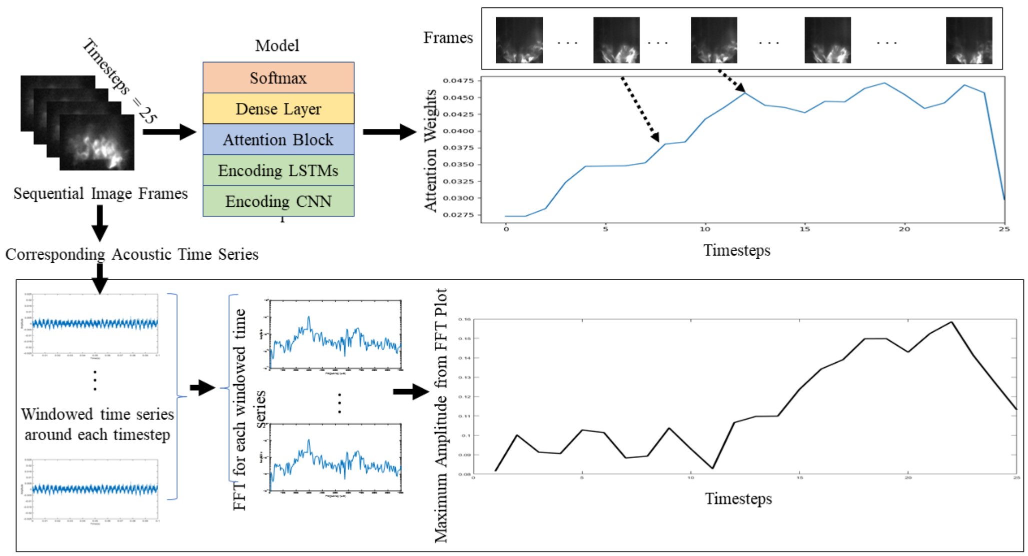

Figure 3: The explanations provided by the proposed framework validated using domain knowledge.

in this publication are those of the authors and do not necessarily reflect the views of the sponsoring agency.

References

[1] EA Gopalakrishnan, Yogita Sharma, Tony John, Partha Sharathi Dutta, and RI Sujith. Early warning signals for critical transitions in a thermoacoustic system. *Scientific reports*, 6:35310, 2016.

[2] John William Strutt Rayleigh. The explanation of certain acoustical phenomena. *Nature*, 18(455):319–321, 1878.

[3] Ann P Dowling. Nonlinear self-excited oscillations of a ducted flame. *Journal of fluid mechanics*, 346:271–290, 1997.

[4] FE Culick and Paul Kuentzmann. Unsteady motions in combustion chambers for propulsion systems. Technical report, NATO RESEARCH AND TECHNOLOGY ORGANIZATION NEUILLY-SUR-SEINE (FRANCE), 2006.

[5] Steven C Fisher and Shamim A Rahman. Remembering the giants: Apollo rocket propulsion development. 2009.

[6] P Palies, T Schuller, D Durox, and S Candel. Modeling of premixed swirling flames transfer functions. *Proceedings of the combustion institute*, 33(2):2967–2974, 2011.

[7] Benjamin D Bellows, Mohan K Bobba, Annalisa Forte, Jerry M Seitzman, and Tim Lieuwen. Flame transfer function saturation mechanisms in a swirl-stabilized combustor. *Proceedings of the combustion institute*, 31(2):3181–3188, 2007.

[8] Peter J Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of fluid mechanics*, 656:5–28, 2010.

[9] Sambuddha Ghosal, Vikram Ramanan, Soumalya Sarkar, Satyanarayanan R Chakravarthy, and Soumik Sarkar. Detection and analysis of combustion instability from hi-speed flame images using dynamic mode decomposition. In *ASME 2016 Dynamic Systems and Control Conference*, pages V001T12A005–V001T12A005. American Society of Mechanical Engineers, 2016.

[10] Vineeth Nair and RI Sujith. Multifractality in combustion noise: predicting an impending combustion instability. *Journal of Fluid Mechanics*, 747:635–655, 2014.

[11] Uddalok Sen, Tryambak Gangopadhyay, Chandrachur Bhattacharya, Achintya Mukhopadhyay, and Swarnendu Sen. Dynamic characterization of a ducted inverse diffusion flame using recurrence analysis. *Combustion Science and Technology*, 190(1):32–56, 2018.

[12] Jeffrey Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell. Long-term recurrent convolutional networks for visual recognition and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2625–2634, 2015.

[13] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507, 2006.

[14] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[15] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[16] Soumalya Sarkar, Kin G Lore, Soumik Sarkar, Vikram Ramanan, Satyanarayanan R Chakravarthy, Shashi Phoha, and Asok Ray. Early detection of combustion instability from hi-speed flame images via deep learning and symbolic time series analysis. In *Annual Conf. of the Prognostics and Health Management*, 2015.

[17] Soumalya Sarkar, Kin Gwn Lore, and Soumik Sarkar. Early detection of combustion instability by neural-symbolic analysis on hi-speed video. In *Proceedings of the 2015th International Conference on Cognitive Computation: Integrating Neural and Symbolic Approaches-Volume 1583*, pages 93–101. CEUR-WS. org, 2015.

[18] Adedotun Akintayo, Kin Gwn Lore, Soumalya Sarkar, and Soumik Sarkar. Prognostics of combustion instabilities from hi-speed flame video using a deep convolutional selective autoencoder. *International Journal of Prognostics and Health Management*, 7(023):1–14, 2016.

[19] Sambuddha Ghosal, Adedotun Akintayo, Paige Boor, and Soumik Sarkar. High speed video-based health monitoring using 3d deep learning.

[20] Tryambak Gangopadhyay, Anthony Locurto, Paige Boor, James B. Michael, and Soumik Sarkar. Characterizing combustion instability using deep convolutional neural network. In *ASME 2018 Dynamic Systems and Control Conference*. American Society of Mechanical Engineers, 2018.

[21] Kush R Varshney, Prashant Khanduri, Pranay Sharma, Shan Zhang, and Pramod K Varshney. Why interpretability in machine learning? an answer using distributed detection and data fusion theory. *arXiv preprint arXiv:1806.09710*, 2018.

[22] David Baehrens, Timon Schroeter, Stefan Harmeling, Motoaki Kawanabe, Katja Hansen, and Klaus-Robert MÃžller. How to explain individual classification decisions. *Journal of Machine Learning Research*, 11(Jun):1803–1831, 2010.

[23] Yi Li, Colin Campbell, and Michael Tipping. Bayesian automatic relevance determination algorithms for classifying gene expression data. *Bioinformatics*, 18(10):1332–1339, 2002.

[24] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.