Cross-Modal Virtual Sensing for Combustion Instability Monitoring

Tryambak Gangopadhyay
Iowa State University
Ames, IA, USA
tryambak@iastate.edu

Vikram Ramanan
Indian Institute of Technology Madras
Chennai, India
vikrambest@yahoo.co.in

Satyanarayanan R Chakravarthy
Indian Institute of Technology Madras
Chennai, India
satyachakra@gmail.com

Soumik Sarkar
Iowa State University
Ames, IA, USA
soumiks@iastate.edu

Abstract

In many cyber-physical systems, imaging can be an important but expensive or 'difficult to deploy' sensing modality. One such example is detecting combustion instability using flame images, where deep learning frameworks have demonstrated state-of-the-art performance. The proposed frameworks are also shown to be quite trustworthy such that domain experts can have sufficient confidence to use these models in real systems to prevent unwanted incidents. However, flame imaging is not a common sensing modality in engine combustors today. Therefore, the current roadblock exists on the hardware side regarding the acquisition and processing of high-volume flame images. On the other hand, the acoustic pressure time series is a more feasible modality for data collection in real combustors. To utilize acoustic time series as a sensing modality, we propose a novel cross-modal encoder-decoder architecture that can reconstruct cross-modal visual features from acoustic pressure time series in combustion systems. With the "distillation" of cross-modal features, the results demonstrate that the detection accuracy can be enhanced using the virtual visual sensing modality. By providing the benefit of cross-modal reconstruction, our framework can prove to be useful in different domains well beyond the power generation and transportation industries.

1 Introduction

In aerospace and energy industries, ultra-lean premixed combustion is preferred to make gas turbine engines more fuel-efficient with lower cost and low NOx (nitrogen oxides) emissions. With this attempt to make engines efficient and environment-friendly, such operating regimes can make engines more prone to an undesirable phenomenon called combustion instability, which is caused by the establishment of a positive feedback loop between heat release rate fluctuations and fluctuating acoustic pressure [1]. Combustion instability can lead to large levels of vibration in an engine [2, 3] resulting in huge revenue loss due to poor performance, reduced life, or catastrophic failure of engine [4].

Previously, researchers have studied combustion instability using full-scale computational fluid dynamics [5], physics-based [6] and reduced order [7] modeling approaches. An alternative is to implement data-driven methods utilizing acoustic pressure time-series [8, 9, 10], based only on acoustic pressure time series, which can sometimes be inaccurate due to interference from broadband background noise. With the rapid development in the field of computer vision, the application of deep

Fourth Workshop on Machine Learning and the Physical Sciences (NeurIPS 2021).

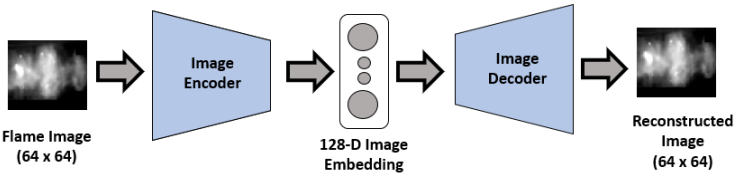

Figure 1: The encoder and decoder of the convolutional autoencoder model used for pre-training.

learning models has started in this domain to detect instability from flame images [11, 12, 13, 14, 15]. The image-based deep learning frameworks have proved to be accurate, trustworthy and can build the confidence of domain experts to implement these models in real systems to detect combustion instability. However, the current roadblock exists on the hardware side.

Acquisition and processing of high-volume flame image data may not be feasible to perform fast enough using existing commercial hardware. Also, flame imaging is not a common sensing modality in engines today. Therefore, the image-based deep learning frameworks can only become feasible with rapid improvement in the hardware sector. Acoustic pressure time series is a more feasible modality for data collection in real combustors. To circumvent the hardware roadblock and simultaneously ensure high detection accuracy, the optimal solution can be to utilize acoustic time series as a sensing modality and implement image-based deep learning models. In this work, we attempt to think in that direction by proposing a novel virtual sensing model (VSenseNet) to reconstruct cross-modal visual features from acoustic pressure time series in combustion systems. While researchers have proposed cross-modal reconstruction models for text-to-image [16, 17, 18], and speech-to-face [19, 20], there has been no work on the reconstruction of visual features from time series in any application domain.

Contributions. We summarize the contributions of this work as follows:

1. To the best of our knowledge, this is the first work on cross-modal reconstruction of visual features from time series in any domain. The proposed cross-modal encoder-decoder model VSenseNet is novel in the context of combustion systems to reconstruct flame images from acoustic pressure time series.
2. In VSenseNet, visual reconstruction from time series is achieved by training the encoder-decoder with “distillation" of cross-modal features from models pre-trained on images. During testing, the classification performance of synthetic images is compared against that of actual images.
3. With acoustic time series as the sensing modality, we demonstrate that instability detection accuracy can be enhanced using our proposed virtual sensing modeling approach. VSenseNet can prove to be a great resource in different sectors where imaging is an important but ‘difficult to deploy’ sensing modality.

2 Virtual Sensing Model

To address the hardware roadblock of flame image acquisition, to use acoustic time series as sensing modality, and to simultaneously utilize an image-based detection framework for better accuracy, we propose a novel cross-modal encoder-decoder virtual sensing model VSenseNet.

Convolutional Autoencoder Pre-Training: The first step is to utilize the training dataset of flame images to pre-train a convolutional autoencoder (Fig. 1), which comprises an image encoder and an image decoder. From the 128-dimensional embedding computed by the encoder (using 2D convolutional and 2D max-pooling layers), the decoder attempts to reconstruct the original flame image as closely as possible using a series of 2D up-sampling and 2D convolutional transpose layers.

Time Series Encoder: The training framework of VSenseNet consists of the time series encoder to compute an embedding from the time series. Long Short Term Memory (LSTM) networks can effectively capture long-term temporal dependencies [21], and LSTM networks are effective in different applications involving time series data [22]. We develop the time series encoder model consisting of two LSTM layers with dropout added after each layer to prevent over-fitting. The last

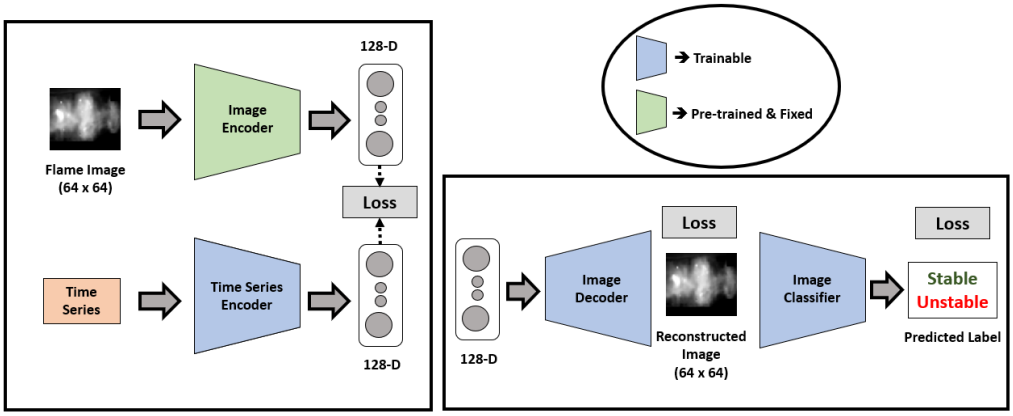

Figure 2: Training framework of VSenseNet. It comprises two steps. The first step is to train the time series encoder. The next step is to train the image decoder and image classifier utilizing reconstruction loss and classification loss.

hidden state of the second LSTM layer is considered the compressed information for the time series. After that, a fully connected layer is used to get a 128-dimensional time series embedding.

VSenseNet: VSenseNet has an image classifier model in the training framework (Fig. 2) apart from the time series encoder and the image decoder. The training framework consists of two steps. In the first step, the time series encoder is trained to regress to the image embedding computed from the pre-trained image encoder. With the time series as input, the time series encoder computes a 128-dimensional time series embedding. The learning objective is to minimize the embedding loss (MSE) between the time series and image embedding. In the next step, the image decoder (same architecture as used in Fig. 1) and the image classifier model are trained. The image classifier model comprises 2D convolutional, 2D max-pooling, and fully connected layers. It is a binary classification model to predict a flame image as stable or unstable. The generated time series embeddings are utilized for training this part of the framework. From an embedding, the image decoder learns to reconstruct the corresponding flame image as closely as possible. With the reconstructed flame image as input, the image classifier model predicts it as stable or unstable. Therefore, the overall test framework for VSenseNet consists of three steps - time series encoder, image decoder, and image classifier.

3 Experiments

Dataset: For dataset collection, we induce combustion instability in a laboratory-scale swirl combustor. The fuel is injected co-axially with air at selected upstream distances. The chosen upstream distances are 90 mm and 120 mm. For the upstream distance of 90 mm, partial premixing of the fuel with air occurs, while the distance of 120 mm facilitates full premixing of the fuel and air. The ground truth labels (stable, unstable) for the hi-speed flame image sequences are provided by the domain experts. We identify the conditions by upstream distance (premixing length), airflow rate (AFR), and fuel flow rate (FFR). Both AFR and FFR are expressed in lpm (liters per minute). The hi-speed images are captured at 3000 Hz (with a resolution of 1024 x 1024) for 3 seconds at each condition. Simultaneously, the pressure data is recorded at 4 locations of the experimental setup with a frequency of 9000 Hz. Therefore, for each condition, we have 9000 frames and 27000 time steps of multivariate pressure data. From a total of six conditions, we use four conditions for training our proposed models and keep two conditions for testing the performance of the models.

The stable and unstable conditions in the training set are:

1. $\mathbf{Stable_{120/60/600}}$: Condition has Premixing Length = 120 mm, FFR = 60 and AFR = 600.
2. $\mathbf{Stable_{90/45/450}}$: Condition has Premixing Length = 90 mm, FFR = 45 and AFR = 450.
3. $\mathbf{Unstable_{120/45/900}}$: Condition has Premixing Length = 120 mm, FFR = 45 and AFR = 900.

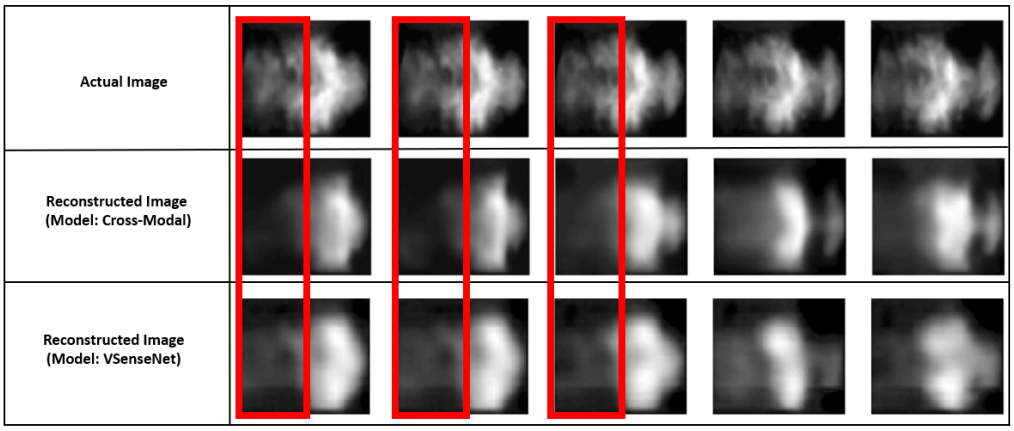

Figure 3: Reconstructions obtained from VSenseNet are compared against reconstructions obtained from Cross-Modal baseline model, and actual flame images for $\mathbf{Unstable_{90/45/900}}$ test condition.

4. $\mathbf{Unstable_{90/28/600}}$: Condition has Premixing Length = 90 mm, FFR = 28 and AFR = 600.

The stable and unstable conditions in the test set are:

1. $\mathbf{Stable_{120/45/450}}$: Condition has Premixing Length = 120 mm, FFR = 45 and AFR = 450.

2. $\mathbf{Unstable_{90/45/900}}$: Condition has Premixing Length = 90 mm, FFR = 45 and AFR = 900.

Results: For comparison of results, we use three baseline models: *Image Classifier* (image-based), *Time Series Classifier* (entirely time series based with no cross-modal reconstruction) and *Cross-Modal Model* (inspired from the Speech2Face [19] model). We use two evaluation metrics for classification performance: Accuracy and False Negative Rate (FNR). For reconstruction performance, we use Mean Squared Error (MSE) and Structural Similarity Index Measure (SSIM) as the evaluation metrics. MSE computed between the actual and reconstructed images may not always be highly indicative of the structural similarity. SSIM [23, 24] addresses this issue by considering texture and also including luminance masking and contrast masking terms. We use Adam optimizer with a learning rate of 0.001 and a batch size of 32. The models are trained using NVIDIA Titan RTX GPU.

Table 1 presents the empirical results for the test set conditions. Using time series as the input modality with virtual sensing approach, we can enhance the average classification accuracy from 98.50% (Time Series Classifier) to 99.01% and approach closer towards the 99.38% (Image Classifier) accuracy achieved by the imaging modality-based model. In terms of reconstruction performance, VSenseNet outperforms the Cross-Modal baseline model in terms of SSIM and MSE. For the test set condition $\mathbf{Unstable_{90/45/900}}$, the VSenseNet model can reconstruct the flame structures better than the baseline Cross-Modal model as highlighted by red boxes in Fig. 3.

Model	Reconstruction Performance		Classification Performance	
	SSIM	Mean Squared Error	Accuracy	FNR
Image Classifier	NA	NA	**0.9938 ± 0.0064**	**0.0122 ± 0.0129**
Time Series Classifier	NA	NA	**0.9850 ± 0.0021**	**0.0300 ± 0.0044**
Cross-Modal	0.6655	0.0072	0.9888 ± 0.0005	0.0222 ± 0.0010
VSenseNet	**0.6876**	**0.0070**	**0.9901 ± 0.0003**	**0.0197 ± 0.0007**

Table 1: Empirical Results for the test set. For classification performance, average and standard deviation of the evaluation metrics (Accuracy, FNR) are reported after training each model five times. For reconstruction performance, average values of SSIM and MSE are reported.

4 Conclusion

Our proposed VSenseNet approach demonstrates effectiveness in reconstructing the flame images. We demonstrate that by cross-modal reconstruction of synthetic images, the classification performance is better than that from time series alone. Therefore we are enhancing the accuracy of combustion instability detection using time series data with our virtual sensing modeling approach. Domain experts can also gain valuable insights during an offline analysis of the reconstructed flame images. In the future, we plan to extend VSenseNet to capture transitions in a combustion system.

References

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

[2] 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.

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

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

[5] 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.

[6] 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.

[7] Simon R Stow and Ann P Dowling. Low-order modelling of thermoacoustic limit cycles. In *ASME turbo expo 2004: power for land, sea, and air*, pages 775–786. American Society of Mechanical Engineers Digital Collection, 2004.

[8] Hiroshi Gotoda, Hiroyuki Nikimoto, Takaya Miyano, and Shigeru Tachibana. Dynamic properties of combustion instability in a lean premixed gas-turbine combustor. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 21(1):013124, 2011.

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

[10] 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.

[11] Soumalya Sarkar, Kin Gwn Lore, and Soumik Sarkar. Early detection of combustion instability by neural-symbolic analysis on hi-speed video. In *CoCo@ NIPS*, 2015.

[12] 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.

[13] Tryambak Gangopadhyay, Anthony Locurto, James B Michael, and Soumik Sarkar. Deep learning algorithms for detecting combustion instabilities. In *Dynamics and Control of Energy Systems*, pages 283–300. Springer, 2020.

[14] Tryambak Gangopadhyay, Sin Yong Tan, Anthony LoCurto, James B Michael, and Soumik Sarkar. Interpretable deep learning for monitoring combustion instability. *IFAC-PapersOnLine*, 53(2):832–837, 2020.

[15] Tryambak Gangopadhyay, Vikram Ramanan, Adedotun Akintayo, Paige K Boor, Soumalya Sarkar, Satyanarayanan R Chakravarthy, and Soumik Sarkar. 3d convolutional selective autoencoder for instability detection in combustion systems. *Energy and AI*, 4:100067, 2021.

[16] Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran, Bernt Schiele, and Honglak Lee. Generative adversarial text to image synthesis. In *International Conference on Machine Learning*, pages 1060–1069. PMLR, 2016.

[17] Tao Xu, Pengchuan Zhang, Qiuyuan Huang, Han Zhang, Zhe Gan, Xiaolei Huang, and Xiaodong He. Attngan: Fine-grained text to image generation with attentional generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1316–1324, 2018.

[18] Wenbo Li, Pengchuan Zhang, Lei Zhang, Qiuyuan Huang, Xiaodong He, Siwei Lyu, and Jianfeng Gao. Object-driven text-to-image synthesis via adversarial training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12174–12182, 2019.

[19] Tae-Hyun Oh, Tali Dekel, Changil Kim, Inbar Mosseri, William T Freeman, Michael Rubinstein, and Wojciech Matusik. Speech2face: Learning the face behind a voice. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7539–7548, 2019.

[20] Amanda Duarte, Francisco Roldan, Miquel Tubau, Janna Escur, Santiago Pascual, Amaia Salvador, Eva Mohedano, Kevin McGuinness, Jordi Torres, and Xavier Giro-i Nieto. Wav2pix: Speech-conditioned face generation using generative adversarial networks. In *ICASSP*, pages 8633–8637, 2019.

[21] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[22] Tryambak Gangopadhyay, Sin Yong Tan, Zhanhong Jiang, Rui Meng, and Soumik Sarkar. Spatiotemporal attention for multivariate time series prediction and interpretation. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3560–3564. IEEE, 2021.

[23] Zhou Wang and Alan C Bovik. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE signal processing magazine*, 26(1):98–117, 2009.

[24] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] Please check Section 3.
   (b) Did you describe the limitations of your work? [Yes] In the future, we plan to extend VSenseNet to capture transitions in a combustion system.
   (c) Did you discuss any potential negative societal impacts of your work? [N/A]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] More details will be made publicly available during the final submission.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] Please check Section 3.

(c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] Please check Table 1.

(d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] Please check Section 3.

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...

(a) If your work uses existing assets, did you cite the creators? [N/A]

(b) Did you mention the license of the assets? [N/A]

(c) Did you include any new assets either in the supplemental material or as a URL? [N/A]

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]

5. If you used crowdsourcing or conducted research with human subjects...

(a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]

(b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]