Temporal Learning in Video Data Using Deep Learning and Gaussian Processes

Devesh K. Jha[1], Abhishek Srivastav[2], and Asok Ray[3]

[1,3] *Mechanical and Nuclear Engineering Department, Pennsylvania State University, PA, USA*
dkj5042@psu.edu
axr2@psu.edu

[2] *Machine Learning Lab, GE Global Research, San Ramon, CA, USA*
srivastav@ge.com

ABSTRACT

This paper presents an approach for data-driven modeling of hidden, stationary temporal dynamics in sequential images or videos using deep learning and Bayesian non-parametric techniques. In particular, a deep Convolutional Neural Network (CNN) is used to extract spatial features in an unsupervised fashion from individual images and then, a Gaussian process is used to model the temporal dynamics of the spatial features extracted by the deep CNN. By decomposing the spatial and temporal components and utilizing the strengths of deep learning and Gaussian processes for the respective sub-problems, we are able to construct a model that is able to capture complex spatio-temporal phenomena while using relatively small number of free parameters. The proposed approach is tested on high-speed grey-scale video data obtained of combustion flames in a swirl-stabilized combustor, where certain protocols are used to induce instability in combustion process. The proposed approach is then used to detect and predict the transition of the combustion process from stable to unstable regime. It is demonstrated that the proposed approach is able to detect unstable flame conditions using very few frames from high-speed video. This is useful as early detection of unstable combustion can lead to better control strategies to mitigate instability. Results from the proposed approach are compared and contrasted with several baselines and recent work in this area. The performance of the proposed approach is found to be significantly better in terms of detection accuracy, model complexity and lead-time to detection.

1. MOTIVATION AND INTRODUCTION

Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstractions (LeCun, Bengio, & Hinton, 2015; Bengio, Courville, & Vincent, 2013; Hinton & Salakhutdinov, 2006). Deep learning methods are representation-learning techniques obtained by composition of non-linear modules or layers that each transform the representation at the previous level into a higher and slightly more abstract level in a hierarchical manner. The main idea is that by cascading a large number of such transformations, very complex functions can be learned in a data-driven manner. Convolutional deep neural nets (Krizhevsky, Sutskever, & Hinton, 2012; Lee, Grosse, Ranganath, & Ng, 2009) are designed to process data that come in the form of multiple arrays, for example images. In this technique, a discrete spatial convolution filter is used for detecting highly correlated distinctive motifs across an image. The same process can be extended to find temporal features by using a convolution filter over time, and using a stack of images as input. However, this will lead to an increase in the number of the hyperparameters used to train such a spatio-temporal convolutional network. Instead, we propose a framework, where a convolutional network is used to extract features from images and a Bayesian nonparametric model like the Gaussian process (Rasmussen & Williams, 2006) is used to model the temporal dynamics of the features; the objective is to reduce the number of parameters used to train such a network while modeling the spatio-temporal dynamics in an image sequence or videos. The proposed algorithm is used to detect stable and unstable combustion flame dynamics in a swirl-stabilized combustor. Another motivation is that adding a Gaussian process-based filter might allow to reduce over-fitting of the deep CNN to some extent as it allows to relate to the causal dynamics present in the data in a much lower dimensional space.

Combustion instability is a highly nonlinear coupled thermoacoustic phenomenon that results in self-sustained oscillations in a combustor. These oscillations may result in severe structural degradation in gas turbine engines. Some good surveys on the current understanding of the mechanisms for combustion instability phenomena can be found in (O'Connor, Acharya, & Lieuwen, 2015; Sé et al., 2003; Candel, Durox, Schuller, Bourgouin, & Moeck, 2014; Huang & Yang, 2009; Moeck, Bourgouin, Durox, Schuller, & Candel, 2012). The current state-of-the-art techniques rely heavily on model-based approaches for analysis of the process. The current technical literature lacks rigorous statistical analysis of the combustion instability phenomenon so that the predictive power of the data has been largely overlooked.

Active combustion instability control (ACIC) with fuel modulation has proven to be an effective approach for reducing pressure oscillations in combustors (Banaszuk, Mehta, Jacobson, & Khibnik, 2006; Banaszuk, Mehta, & Hagen, 2007). Based on the work available in literature, one can conclude that the performance of ACIC is primarily limited by large delays in the feedback loop and limited actuator bandwidth (Banaszuk et al., 2006, 2007). Model-based approaches for active control are infeasible as complexity of the models and uncertain measurements make real-time estimates difficult. On the other hand, use of machine learning techniques for information extraction remain unexplored for this problem. From the perspective of active control of the unstable phenomena, it is necessary to accurately detect and, desirably, predict the future states of the combustion process. The goal of this paper is to present a statistical model for the instability phenomenon during combustion which could be used to design a statistical filter to accurately predict the system states. This can potentially alleviate the problems with delay in the ACIC feedback loop and thus possibly improve the performance.

Some recent work on statistical analysis of combustion instability using pressure time-series data could be found in (Jha, Virani, & Ray, 2016; Virani, Jha, & Ray, 2016). The work presented in (Jha et al., 2016) shows the change in the underlying Markov model for pressure data as the system approaches the unstable regime resulting in self-sustained oscillations of the flame. Some other popular methods for detection of coherent structures include proper orthogonal decomposition (POD) (Berkooz, Holmes, & Lumley, 1993) and dynamic mode decomposition (DMD) (Schmid, 2010), which use tools of spectral theory to derive spatial coherent structure modes. Specifically, DMD has been used to estimate growth rates and frequencies from experimental data and also for stability analysis of experimental data. Recently, some analyses were also presented using deep learning for detection of combustion instabilities (Sarkar, Lore, Sarkar, Ramanan, et al., 2015; Sarkar, Jha, Lore, Sarkar, & Ray, 2016). More recently the work done in (Sarkar, Lore, & Sarkar, 2015) presents a

neuro-symbolic approach where the output a deep convolutional network are analyzed by a Markov modeling module to construct an anomaly measure. However, as we demonstrate later, the advantages of using a deep neural network is unclear. Another analysis is presented in (Hauser, Li, Li, & Ray, 2016) using various image analysis techniques like histogram of oriented gradients (HOG) and Wavelets; however, the final decision is made by making a Markov model out of the features and thus requires long sequences to arrive at a decision. Moreover, perfect separability between stable and unstable classes is not achieved.

This paper presents further insights into the combustion instability phenomena from a data-driven perspective (Darema, 2005) and presents a framework which allows temporal learning in video data in a lower dimensional subspace of features produced by a deep learning module. We show that flames have different spatial structures at unstable behavior when compared to stable behavior and with an appropriate architecture of neural network, we can perfectly capture the associated distinguishing features.

Contributions: This paper presents a framework for modeling of spatio-temporal dynamics in sequential images using deep learning and Gaussian processes. We use the proposed algorithm to present a statistical analysis of the complex combustion instability phenomena and show that we are able to capture the change in the model as the system moves from a stable regime to an unstable regime. We show that the deep CNN is able to achieve fairly good classification performance; however the use of a Gaussian process filter to model temporality of extracted features results in perfect detection with very low false alarm rates and much shorter lead time to detection. The advantage of the proposed method is that it can be used for making early predictions of the transition from a stable regime to quasi-periodic unstable oscillations in combustion system.

2. PRELIMINARIES AND PROBLEM FORMULATION

In this section, we will very briefly describe the convolutional neural nets and Gaussian processes. Interested readers are referred to (Lee et al., 2009; Rasmussen & Williams, 2006; Bengio et al., 2013) for an in-depth discussion on these topics. After the brief review, we will state the problem considered in this paper. The idea to use Gaussian processes to model the features obtained from deep CNN is to be able to add memory to the estimation filter without going to the architecture of a recurrent neural network.

2.1. Convolutional Neural Networks

In this section, we briefly introduce the deep convolutional neural networks (CNN) for the completeness of the paper and motivation for use in this work. Deep convolutional neural networks have a long history (Le Cun Y. et al., 1990) but
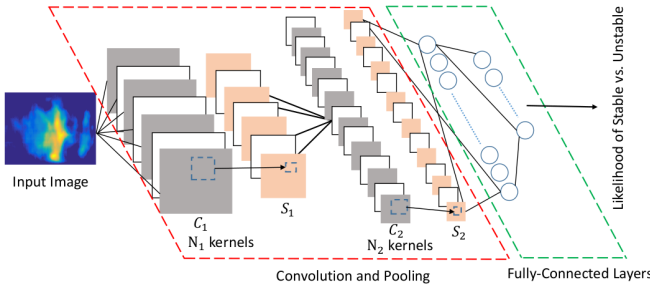
Figure 1. This figure shows the deep convolutional neural network used with two convolutional layers and two hidden layers. The number of kernels and number of units in the hidden layer could be varied to study the affect of size of hyperparameter set on the results. The convolution layers are denoted by C_i and the pooling layers are denoted by S_i. Each rectangular image is a feature map.

they caught the attention of vision community when they surpassed the state-of-the-art algorithms for image recognition problem by large margins (Krizhevsky et al., 2012). Since then, it has been very widely used for almost all software applications for recognition in images and videos. These networks are widely used to process data that come in form of multiple arrays, for example a color image composed of $2D$ arrays containing pixel intensities. In practice, the deep convolutional networks have shown to outperform most of the other hand-tuned feature extraction algorithms and thus, have become very popular for learning with image data. This is also the motivation for the use of CNN in this work. The general architecture of a deep convolutional neural network is shown in Figure 1 and is structured as a series of stages. In general, the first few stages (left) contain a lot of local information which is aggregated as we go deeper (right) in the network. The first stages are composed of two types of layers: convolutional and pooling layers (shown as C_i and S_i respectively in Figure 1). Units in a convolutional layer are organized in feature maps, within which each unit is connected to local patches in the feature maps of the layer through a set of weights called filter banks (or kernels). The result of this local weighted sum is then passed through a non-linearity. Mathematically, the filtering operation performed by a feature map (or kernel) is equivalent to discrete convolution (and hence the name). The motivation for use of discrete convolution is to be able to find local correlated, distinctive patches (or motifs) in the images.

Contrary to the convolution layer, the role of the pooling layer is to merge semantically similar features into one. A typical pooling unit computes the maximum of a local patch of units in one feature map (or in a few feature maps). Neighboring pooling units take input from patches that are shifted by more than one row or column, thereby reducing the dimension of the representation and creating invariance to small shifts

and distortions. This also helps with the overfitting problem by reducing the dimension of the representation. Two or three layers of convolution, non-linearity and max-pooling are stacked, followed by the fully connected layers. The network is then trained using backpropagation using stochastic gradient descent and it allows to train the weights of the filter banks (kernels).

2.2. Gaussian Processes

In this section, we very briefly explain the Gaussian process model which is used to model the features extracted by CNN.

Definition 2.1 (Gaussian Processes) *A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution.*

A Gaussian process (GP) is completely specified by its mean function and covariance function. We define the mean function $m(\mathbf{x})$ and covariance function $k(\mathbf{x}, \mathbf{x}')$ of a real process $f(\mathbf{x})$ as follows.

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$$
$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]$$

and the Gaussian process is written as follows.

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \tag{1}$$

A Gaussian process is defined as a collection of random variables such that the if the GP specifies $(x_1, x_2) \sim \mathcal{N}(\mu, \Sigma)$, then it must also specify $x_1 \sim \mathcal{N}(\mu_1, \Sigma_{11})$ where Σ_{11} is the relevant submatrix of Σ. Some of the common choice for GP covariance function are rational quadratic (RQ), squared exponential (SE) and Matérn covariance functions. Some of the functional forms of the covariance function and the associated free parameters are listed in Table 1. The inference problem associated with GP is to infer the related hyperparameters (or the free parameters). The associated hyperparameters with a GP are the parameters corresponding to the mean and covariance function for the GP. The inference methods compute an approximate posterior, an approximate negative log marginal likelihood and its partial derivatives w.r.t. the hyperparameters, given the choice of mean, covariance and likelihood functions. Some of the common inference methods are expectation propagation, Laplace's approximation and Kullback-Leibler divergence minimisation (Rasmussen & Williams, 2006).

2.3. Problem Statement

Consider a sequence of images, $\{\boldsymbol{X}_t\}_{t \in \mathbb{N}}$, $\boldsymbol{X}_t \in \mathbb{R}^{d \times d}$, $d \in \mathbb{N}$, where $\boldsymbol{X}_t$ represents observations from a dynamical system with finite memory s.t. $\boldsymbol{X}_t = f(\boldsymbol{X}_{t-1}, \dots, \boldsymbol{X}_{t-m})$. Then, the statistical learning task is to find a representation of the dynamical system f in a lower dimensional subspace such that $\tilde{\boldsymbol{X}}_t = g(\tilde{\boldsymbol{X}}_{t-1}, \dots, \tilde{\boldsymbol{X}}_{t-p})$, where $\tilde{\boldsymbol{X}}_t = \varphi(\boldsymbol{X}_t)$,

Table 1. Some examples of covariance functions used for Gaussian Process. The first and second are Squared Exponential (SE) and the third function is a rational quadratic (RQ).

Mathematical form	Free Parameters
$\sigma^2 \exp(-(x-x')^T \Lambda^{-2}(x-x')/2)$	$\{\lambda_1, \ldots, \lambda_D, \sigma\}$
$\sigma^2 \exp(-(x-x')^T (x-x')/2\ell^2)$	$\{\ell, \sigma\}$
$\sigma^2(1 + \frac{1}{2\alpha}(x-x')^T \Lambda^{-2}(x-x'))^{-\alpha}$	$\{\lambda_1, \ldots, \lambda_D, \sigma, \alpha\}$

$\tilde{\boldsymbol{X}}_t \in \mathbb{R}^{k \times k}$, $k << d$. It should be noted that the feature extraction transformation (φ) may not retain the original system memory (i.e., $m \neq p$). The idea of the underlying problem is also shown as a schematic in Figure 2. In the proposed problem, the deep neural network could be just used as a feature extractor and we allow the decisions to be made by modeling the features using a Gaussian process. The motivation for using such a composite approach is twofold. Firstly, it can help relate to the causal dynamics of the underlying physical system, and second, it can possibly help with the overfitting of the neural network. However, we mainly consider the first part in this paper.

3. COMBUSTION EXPERIMENT DETAILS

The experimental apparatus consists of a swirl combustion chamber where the inlet conditions can be varied to achieve combustion in stable and unstable modes. In these experiments the inlet conditions are controlled via a combination of premixing condition of air and fuel, fuel-flow rate (FFR) and inlet Reynolds number (Re), each combination defining a test protocol. Reynolds number is a dimensionless quantity and is defined as the ratio of inertial and viscous forces in a flow; typically high Reynold number flows are turbulent and vice-versa. More information on the experimental apparatus is provided in the Appendix.

In one test protocol, two inlet Reynolds numbers (Re) were chosen $-=7,971$ and $=15,942$ for a fixed fuel flow rate of 0.495 g/s, where the lower Re lead to stable combustion behavior and higher Re exhibited unstable behavior. In another other protocol, the inlet Re is held constant at $10,628$ for two different fuel flow rates (FFR). The details of the protocols along with their ground truths (e.g., stable, relatively stable and unstable) are presented in Table 2.

Apart from this, an intermediate stage is also tested to capture the transition between stable and unstable combustion. This is done by (1) increasing the air flow rate (AFR) while keeping the FFR constant, and, (2) by decreasing the fuel flow rate while keeping AFR constant so that the system state is varied from stable to unstable. The transition protocols are listed in Table 3. (It is noted that we only consider the protocol (2) in this paper, i.e., instability is induced by decreasing the fuel flow rate). All units in liters per minute

Table 2. Experiment set 1: Protocols with respective ground truth conditions for data collection. $3\,\text{s}$ of greyscale image sequence at $3\,\text{kHz}$.

Premixing	FFR (g/s)	Re	Ground truth
Partial	0.495	7,971	Stable
	0.495	15,942	Unstable
	0.308	10,628	Unstable
	0.660	10,628	Stable
Full	0.495	7,971	Stable
	0.495	15,942	Unstable
	0.308	10,628	Unstable
	0.660	10,628	Stable

Table 3. Experiment set 2: Protocols to produce transition from stable to unstable combustion

AFR		FFR (g/s)		Ground truth
start	end	start	end	
500	500	40	30	Stable→Unstable
600	600	50	35	Stable→Unstable

The swirl combustor has an Inlet Optical Access Module (IOAM) which is used to collect High-speed images of the combustion process at a resolution of 1024×1024 and a frequency of 3kHz. Data was collected for a period of 3 seconds yielding a sequence of 9000 images for every operating condition. In total 72000 images are generated from the experiment with $36,000$ images each for the stable and unstable class that are analysed using the baselines and the proposed method. Since the flame is limited mainly along the horizontal axis, the images are cropped along the vertical dimension to remove background. The resolution is then scaled down to get a final image size of 51×128, that is analyzed in this paper.

In Figure 3a, we show sequence of images at different time instants to show the changes in the spatial structure of flames at stable condition. Figure 3b presents sequences of images of dimension 392×1000 pixels for unstable (Re $= 15,942$, FFR$= 0.495\,\text{g s}^{-1}$ and full premixing) state. The flame inlet is on the right side of each image and the flame flows downstream to the left. Figure 3b shows formation of mushroom-shaped vortex at $t = 0, 0.001s$ and the shedding of that towards downstream from $t = 0.002\,\text{s}$ to $t = 0.004\,\text{s}$. If looked carefully, one can observe the temporal changes in the flames during unstable combustion; it shows the periodic blowing off of images at times (e.g., $t = 0.001\,\text{s}$). This is sometimes studied as appearance of *coherent structures* which can be loosely defined as organized spatial features which repeatedly appear and undergo a characteristic temporal life cycle. The temporal life cycle can be associated to the limit cycle that the system gets locked on to during an unstable condition which can occur at a lot of operating conditions.
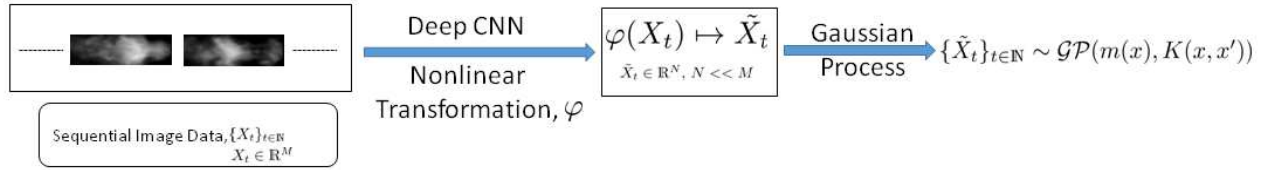
Figure 2. Concept of the proposed modeling scheme using the deep convolution net and Gaussian processes which is used to model the hidden dynamics in the sequential image data.
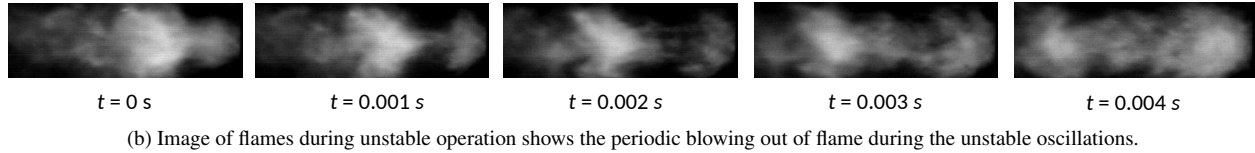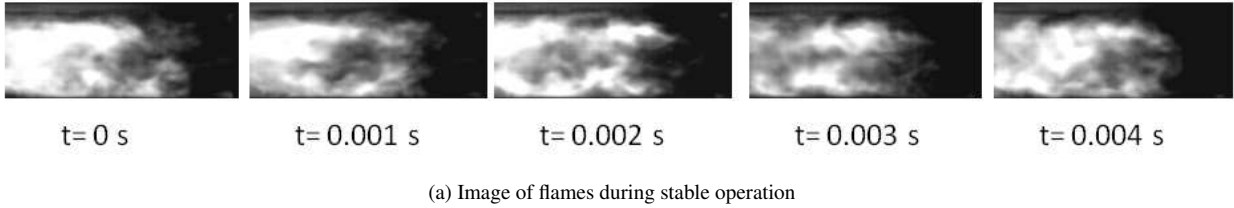

(a) Image of flames during stable operation

(b) Image of flames during unstable operation shows the periodic blowing out of flame during the unstable oscillations.

Figure 3. High-speed image data from stable and unstable regimes of combustion. Spatial and temporal changes in the flame structure during the unstable process are visible. The flame enters on the right end and moves towards the left.

4. RESULTS AND DISCUSSION

We will first show that the unstable pressure fluctuations occur when the system gets locked onto a quasi-periodic, limit cyclic behavior due to the synergy between heat release rate fluctuations and the acoustic properties of the combustor. In classical combustion literature, these limit cycles are associated with phase lag between heat release rate frequency and velocity (also known as Rayleigh's criterion). It is in general difficult to quantify this phase lag from image data, we are able to capture this cyclic behavior by finding the approximate change in the flame images. In Figure 4, we show the lumped changes in the video data as the combustion moves from stable to unstable. The Euclidean distance between the images are calculated from an initial reference image at time point k by computing the 2-norm of the image residual with respect to the reference. In Figure 4a, we show the approximate empirical density of the Euclidean norm for the sequential image data during the two regime. As we can see, there is a change in the empirical density from approximately a unimodal Gaussian to multi-modal. In Figure 4b, we show the time-series of the residual norm which shows a near-periodic behavior (limit cycle) during the unstable regime. The autocorrelation of the image residuals during unstable regime is also shown in Figure 5 which shows a periodic behavior with a frequency ~ 26 showing the limit-cycle behavior during

unstable combustion. For the results presented here in Figures 4 and 5, the initial reference image is $k = 1$, but it can be seen from the time-series and the autocorrelation plots that the cyclic pattern is independent of k upto a phase shift. The reference image can also be a mean image calculated using a number of initial images without changing the results. From Figure 4, it is clear that these patterns can be used to classify or detect the two regimes; however, as we have lumped a lot of information from images on to a single dimension – the 2-norm of the residual from reference image, we would need a longer sequence of data to detect and classify.

At this point, we would like to point out that the analysis presented in (Sarkar, Lore, Sarkar, Ramanan, et al., 2015; Sarkar, Lore, & Sarkar, 2015; Hauser et al., 2016) using symbolic analysis of deep learning features and HOG features could potentially be done using the Euclidean norm between the images. Using this intuition, our first baseline consists of learning a Gaussian Process on the norm of residual of images from a reference image (we choose $k = 1$ for reference image). The residual norm is first normalized by removing the bias and dividing by the variance (so that the GP is not sensitive to flame luminosity but rather the temporal nature of data). An isotropic squared exponential function is used as the covariance function for the GP which is given by the following equation.
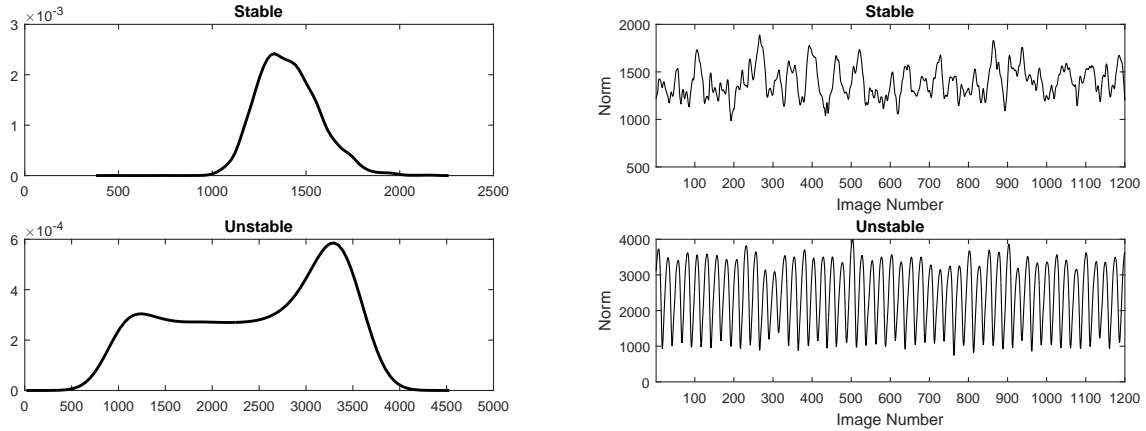
(a) Change in the approximate empirical density of the Euclidean distance between the images

(b) Change in the Euclidean distance between images with time shows the limit cycle during the unstable regime

Figure 4. Change in image data as the combustion process moves from stable to unstable. From the approximate empirical density and the limit cycle shown in sequence of the Euclidean norm between the images, change from minor fluctuations to a quasi-periodic behavior is visible
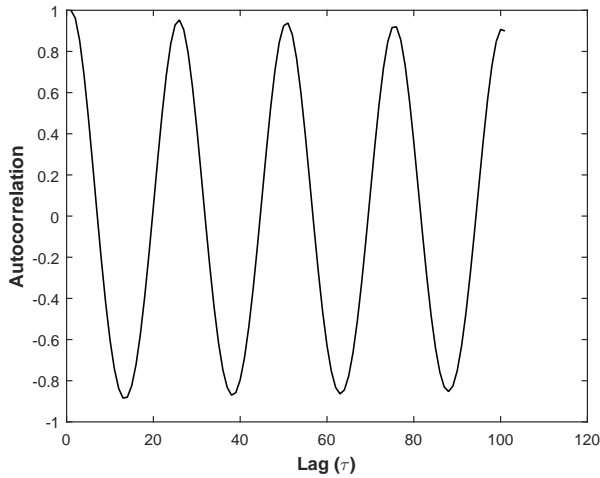

Figure 5. Autocorrelation of the image residuals (4b) during the unstable regime of combustion indicating the limit cycle and its frequency (~ 26)
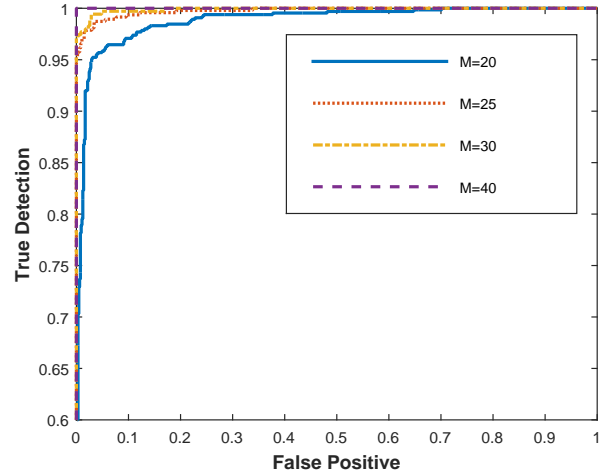
Figure 6. Performance of Gaussian Process on the residual of the sequential images. With sufficient memory, we are able to achieve perfect performance ($M = 40$); with reduction in memory, we see a reduction in performance.

$$k(x, x') = \sigma^2 \exp\left(\frac{-1}{2\ell^2}(x - x')^T(x - x')\right) \quad (2)$$

The corresponding free parameters to be inferred for the GP are just two $- \sigma$ and ℓ associated with the covariance function and they are inferred using the expectation propagation (Rasmussen & Williams, 2006). The likelihood function is modeled as the error function and thus, there are no free parameters associated with the likelihood function. We train two GP models for stable and unstable cases. These two trained GP models are used to calculate the likelihoods $P(x|y)$ for the unstable ($y = 1$) and stable ($y = 0$) classes

and use a Naive Bayes classifier to learn the optimal threshold using the training data set. The results of the binary classification (on the test set) are shown as receiver operating curve (ROC) in figure 6. We see that the GP can model the limit cycle perfectly with enough memory; with reduction in memory, the performance degrades. If looked closely, the performance degrades when the memory is lower than the limit cyclic behavior of the image residuals (the frequency of the limit cycle could be seen from the the image residuals and their autocorrelation in Figures 4 and 5). This shows that a Gaussian process efficiently models the temporal structure in

6

the image residuals. This performance is better than other published results in literature in (Sarkar, Lore, Sarkar, Ramanan, et al., 2015; Sarkar, Lore, & Sarkar, 2015; Hauser et al., 2016); the work in these papers ignores the fact that the combustion process locks on to a limit cycle and that it is this intrinsic dynamics driving the patterns in the time-series data. For example, in (Sarkar, Lore, & Sarkar, 2015) the authors use windows of length 1500 to estimate a measure for anomaly detection. Moreover, in (Hauser et al., 2016), the algorithm can't achieve perfect performance.

Even though we need much shorter memory than recently reported results, the disadvantage of this approach is that we need a relatively longer sequence of images to accumulate enough information to come to a conclusion. This happens as we are only looking at lumped changes in the images and any spatial difference between the images at stable and unstable regimes is ignored. This motivates the use of a deep CNN, as with a deep enough network and large data-set we might be able to extract relevant spatial features by treating the data as iid and learning a deep CNN to get separability between the classes based on just single frames. Therefore, for extracting spatial features, even though the governing physical dynamics has a finite memory, we assume a memory-less system.

The labeled data is divided into train, validation and test data such that out of the total images, half or $36,000$ are used as training set, $14,400$ are used as validation set and the remaining $21,600$ are used as test data. An equal proportion of stable and unstable classes is kept in all the three sets. These images are then used to train the deep CNN. The architecture of CNN is the similar to that shown in Figure 1 with two convolution layers and two pooling layers. The number of kernels (or feature maps) are varied to study the effect of size of parameter set on the performance of the network. Also the number of units are varied to do a study of the effect of the width of the network for the particular problem. The CNN is terminated using early stopping to allow some regularization in the network. The first convolutional layer has 10 kernels ($N_1 = 10$) and the second convolutional layer has 30 ($N_2 = 30$) kernels. The pooling layers perform a 2×2 max-pooling. The second convolutional layer is followed by a fully-connected layer with 80 units and another fully connected layer with 10 units. This is followed by an output layer which uses logistic regression for classification. A learning rate of 0.002 is used to train the CNN.

We are able to achieve fairly good classification performance with the deep CNN without considering the temporal dynamics in the image sequences. We get a test error rate of 4.68% and it took about 11.45 min to train this network (on an Nvidia geforce Titan X Graphics processing unit). This suggests that a sufficient spatial differences exist between single frames of combustion flames at the stable and unstable condition and the proposed deep network was able to extract those

discriminating features from the images. We use the trained models to detect the onset of instability in transient data sets from experiment set 2 that was collected as the system gradually moved from stable to unstable condition (Table 3). This is done by training a Naive Bayes classifier where the likelihood of stable and unstable class are computed using the deep CNN and are used to calculate the likelihood ratio. As we can see in Figure 7, the trained model detects the change as the combustion system gradually moves from stable to unstable behavior. It is noted that in both cases, the experiments were done where the system was stable initially and gradually becomes completely unstable; however the exact ground truth for this transition is not known (as the process is very fast). However, the flame images before and after the transition detected by the CNN model suggests that the onset of unstable behavior. To see this more clearly, we show some sequential images for Figure 7a before and after the transition in the likelihood ratio for the CNN model; the changes in the flame images can be seen in Figure 8. The images under the column *Before* in Figure 8 correspond to the point denoted by a red square in figure 7a and the images under the column *After* correspond to the point shown by an orange diamond in Figure 7a. Similar changes are also observed for Figure 7b and thus they are not shown here. The images under the *After* column can be seen to break near the flame entrance (on left) indicting the onset of periodic blow-off (unstable behavior). Thus, we conclude based on the results of these two experiments that the CNN model is useful for early detection of the phase transition from stable to unstable. Another point to be noted is that the experiments were done at different conditions than the experiments done to train the CNN model; this indicates that CNN model is quite general and that are common features in spatial structure of flames, even at different operating conditions.

We propose a spatio-temporal filter where a Gaussian process model is used to model the temporal dynamics of the features extracted using the deep CNN. We demonstrated earlier that the process has a definite temporal behavior with a finite memory. Therefore, modeling the temporal behavior is very desirable and the hope is to improve both detection performance and lead time to detection by explicitly modeling the temporal dynamics. To do this, we train a multi-dimensional Gaussian process using the output of the 2nd fully connected layer of the deep CNN with 10 units. A 10-dimensional squared exponential covariance function with automatic relevance detection determination (ARD) is chosen as the covariance function of the GP. The covariance function is parameterized as follows.

$$k(x, x') = \sigma^2 \exp\left(-(x - x')^T \Lambda^{-2} (x - x')/2 \right) \quad (3)$$

where the matrix Λ is a diagonal matrix of dimension equal to the dimension of the input space and σ^2 is the signal variance.
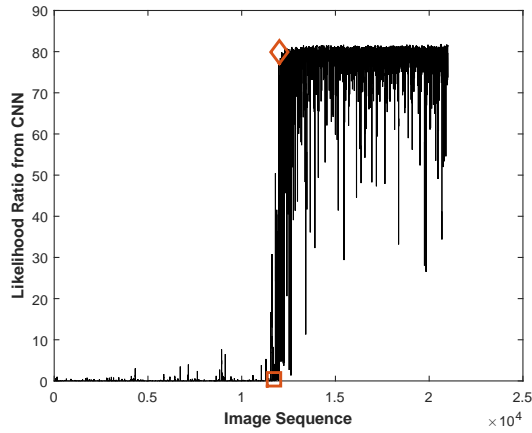
7

(a) Time series of likelihood ratio for transient Data Case-1 in Table 3
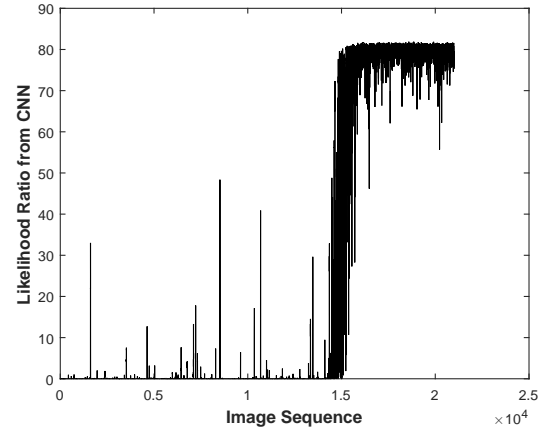
(b) Time series of likelihood ratio for transient Data Case-2 in Table 3

Figure 7. Performance of the trained deep learning algorithm on transient data
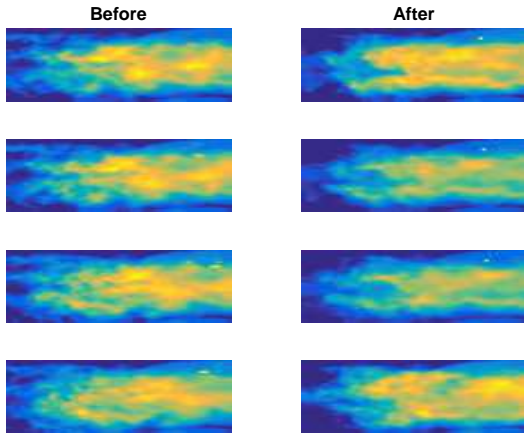

Figure 8. In this figure, we show typical snapshots of flames just before and after the switching detected by the CNN likelihood ratios for Figure 7a (similar changes are observed for figure 7b). The flame enters on the left side and moves towards the right in every image. If seen closely, one can see the flame breaks near the entrance which is due to the periodic behavior that occurs during unstable regime.

The likelihood function has the shape of a error-function (or cumulative Gaussian), which doesn't have any hyperparameters. Thus the total number of free parameters that need to be inferred from data are the hyperparemeters of the covariance function (i.e., the matrix Λ of size equal to the dimansion of the input space and the signal variance) which are inferred using the Laplace approximation or the expectation propagation (Rasmussen & Williams, 2006). During numerical experiments it was found that the Laplace approximation was much faster as compared to the expectation approximation and thus, the results are presented using the Laplace approximation.
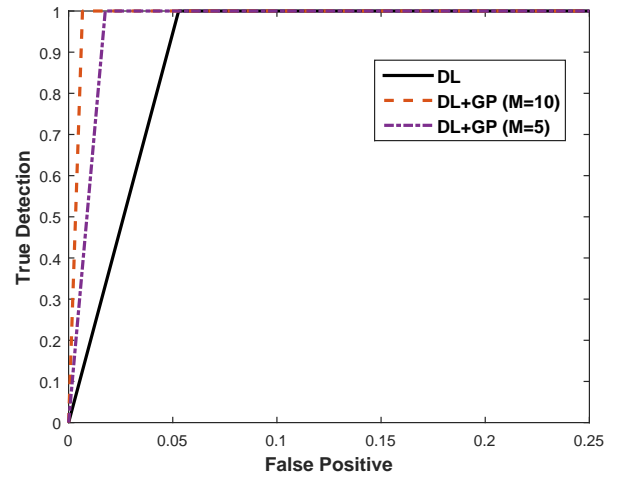

Figure 9. Performance of the trained deep learning algorithm and the Gaussian process on the deep learning features.

mation.

Similar to previous models, a Naive Bayes classifier is trained for the CNN-GP model using the training data set the ROC curve is calculated for the test set. For comparitive evaluation, al results are shown in figure 9. As can be seen, for M= 10, the GP trained on outputs of CNN gets a very small False Positive Rate of 0.7% when the true positive rate is 99.7%. Thus, we achieve near perfect classification with a temporal history of 10 frames as compared to 40 frames using GP alone on the image residuals or 300 frames in (Sarkar, Lore, & Sarkar, 2015). With the data being collected at $3\,\text{kHz}$, this leads to a perfect detection time of $3.3\,\text{ms}$. This is a big improvement over the earlier result in (Sarkar, Lore, & Sarkar, 2015), where the a detection was made every $100\,\text{ms}$. Using only

Table 4. Size of convolution filters

Image Size	Filter 1	Filter 2	Final Image Size
51×128	$15, 17$	$12, 13$	3×22

Table 5. Results of classification performance for baseline and **proposed** models. AUC= Area under curve (higher is better); FPR@99.7 is False positive rate at 99.7% detection rate (lower is better)

Model	Memory (M)	AUC	FPR@99.7
GP	20	0.9830	48.15%
GP	25	0.9918	19.15%
GP	30	0.9966	4.85%
GP	40	1.0000	0.00%
CNN	0	0.9750	5.28%
CNN-GP	5	0.9912	1.18%
CNN-GP	10	0.9987	0.67%

5 frames or in $1.7\,\mathrm{ms}$, the proposed approach is able detect the unstable regime with an accuracy of 99.7% and a false positive rate of 1.18%

Furthermore, the trained models are used to test the performance on the transient data using a memory of 5 frames. During test on the transient data, a window of length 5 is used for decision and every time a new frame is obtained, it replaces the last frame in the window (thus, there is 80% overlap). The likelihood ratios for the trained models are calculated using this sliding window. The results are shown in Figure 10 for the two transient conditions. The results are similar to those obtained by CNN model; however, it is noticed that it is more stable for case-1.

The current results of time for detection are comparable to the time scale of the combustion process (which is on the order of ms). The results of all the numerical experiments and different models trained are listed in Table 5. It is thus concluded that the proposed technique achieves very reliable and fast detection of combustion instabilities.

5. CONCLUSIONS AND FUTURE WORK

Combustion instability still remains a puzzle for researchers and the current state-of-the-art techniques heavily rely on physics-based models. The current analysis presented a data-driven spatio-temporal analysis of combustion flames using deep neural networks and Gaussian processes using high-speed images of flames during lean pre-mixed combustion. The present analysis presented several results on modeling of combustion process during stable and unstable phenomena.

In this paper, we presented a framework for learning hidden, stationary dynamics in video data using deep convolutional networks and Gaussian processes. The main idea was to reduce the size of the parameter set of a spatio-temporal convolutional network by using a Gaussian process to capture the temporal sequence of the features extracted by the deep CNN module. This study presented a rigorous machine learning-based approach to model combustion instability and suggests that statistical learning techniques could help understand and model the complex physical phenomenon to achieve accurate, real-time decisions. The proposed framework was used to model the behavior of lean-premixed flames in a swirl stabilized combustor as the combustion process moves from stable to unstable through a sharp transient phase. The CNN alone was able to achieve fairly good classification performance; however, based on the current results it is concluded that adding Gaussian process allows the filter to be more generalizable as compared to the CNN alone. Based on the numerical experiments done in this paper, it is also concluded that making a filter with appropriate depth (in the neural network) and memory (in the GP) allows to have generalizable data-driven models for the process which possibly can make correct predictions even with changes in lots of associated variables in the process (e.g. mixing-length, fuel-air ratio, combustor geometry).

Using the proposed method with other sensor modalities like pressure to make a data-driven, multi-sensor, hybrid model for combustion instability is a possible topic for future research. Some problems like making predictions on changes in system state and relating to the criterion like Rayleigh's is also suggested as a topic of future research. Also, it seems that for systems with structured dynamics (e.g., engineering systems), using a Bayesian model-based filter will help add reasoning for better understanding of the process; however, a more through analysis is required to understand the benefits of such a hierarchical reasoning. While the results in this paper are encouraging, further investigations using data from multiple operating conditions are required to make a model which also makes predictions on statistical margins of stability for the process.

REFERENCES

Banaszuk, A., Mehta, P. G., & Hagen, G. (2007). The role of control in design: From fixing problems to the design of dynamics. *Control Engineering Practice*, *15*(10),
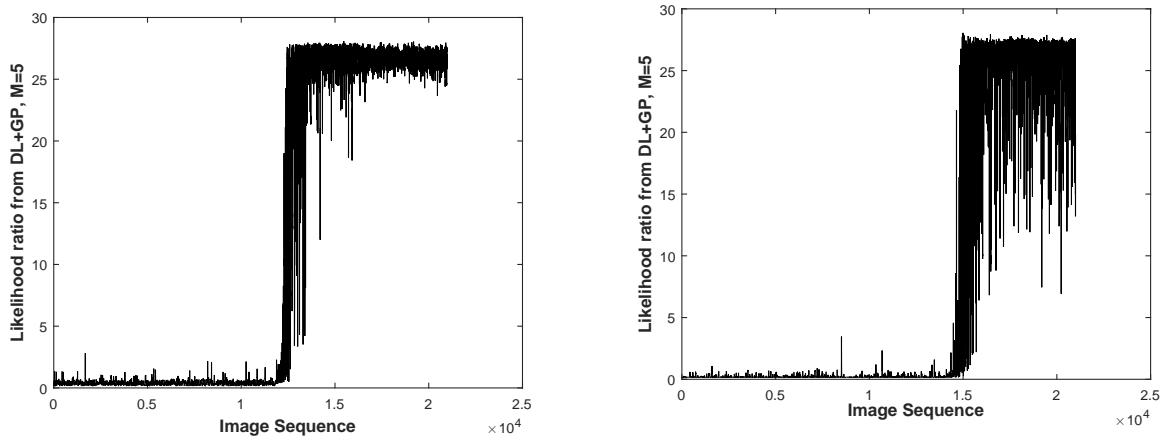
(a) Time Series of Likelihood ratio for Transient Data Case-1 in Table 3   (b) Time Series of Likelihood ratio for Transient Data Case-2 in Table 3

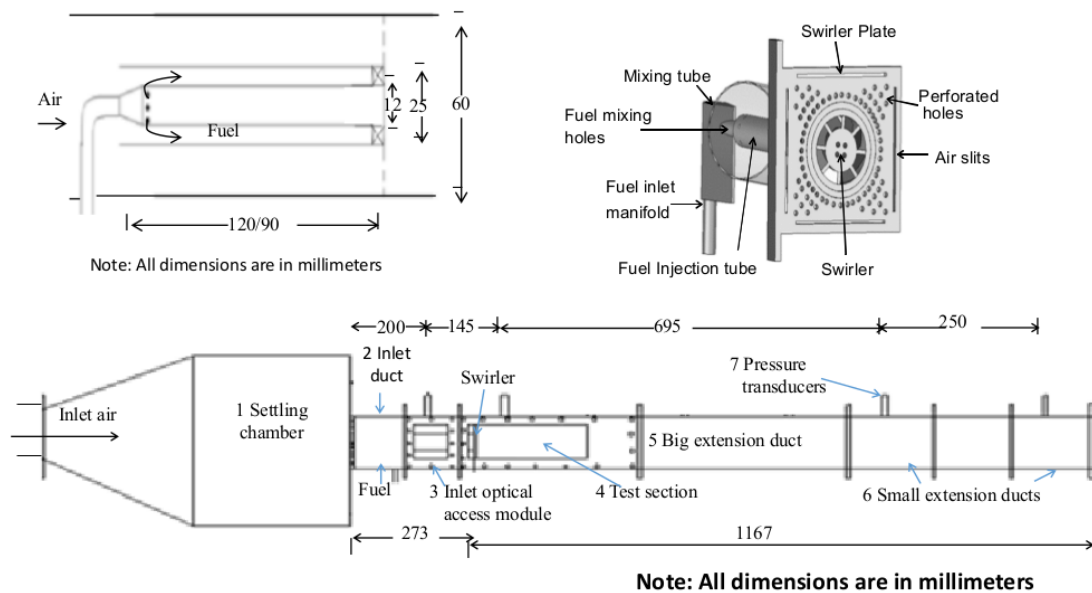Figure 10. Performance of the composite deep learning and Gaussian process algorithm on transient data

Figure 11. (a)Schematic of the experimental setup. 1 - settling chamber, 2 - inlet duct, 3 - IOAM, 4 - test section, 5 - big extension duct, 6 - small extension ducts, 7 - pressure transducers. The figure also shows the swirler plate geometry and the mixing tube of the swirler is also shown in detail.

1292–1305.

Banaszuk, A., Mehta, P. G., Jacobson, C. A., & Khibnik, A. I. (2006). Limits of achievable performance of controlled combustion processes. *Control Systems Technology, IEEE Transactions on*, *14*(5), 881–895.

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *35*(8), 1798–1828.

Berkooz, G., Holmes, P., & Lumley, J. L. (1993). The proper orthogonal decomposition in the analysis of turbulent flows. *Annual review of fluid mechanics*, *25*(1), 539–575.

Candel, S., Durox, D., Schuller, T., Bourgouin, J.-F., & Moeck, J. P. (2014). Dynamics of swirling flames. *Annual review of fluid mechanics*, *46*, 147–173.

Darema, F. (2005). Dynamic data driven applications systems: New capabilities for application simulations and measurements. In *International conference on computational science* (pp. 610–615).

Hauser, M., Li, Y., Li, J., & Ray, A. (2016). Real-time combustion state identification via image processing: A dynamic data-driven approach. In *2016 American Control Conference (ACC)* (p. 3316-3321).

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*(5786), 504–507.

Huang, Y., & Yang, V. (2009). Dynamics and stability of lean-premixed swirl-stabilized combustion. *Progress in Energy and Combustion Science*, *35*(4), 293–364.

Jha, D., Virani, N., & Ray, A. (2016). Markov moldeing of time-series data via spectral analysis. In 1st *International Conference on InfoSymbiotics/DDDAS, Hartford, CT*.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Le Cun Y., B. B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1990). Handwritten digit recognition with a back-propagation network. In *Advances in neural information processing systems*.

Lee, H., Grosse, R., Ranganath, R., & Ng, A. Y. (2009). Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In *Proceedings of the 26th annual international conference on machine learning* (pp. 609–616).

Moeck, J. P., Bourgouin, J.-F., Durox, D., Schuller, T., & Candel, S. (2012). Nonlinear interaction between a precessing vortex core and acoustic oscillations in a turbulent swirling flame. *Combustion and Flame*, *159*(8), 2650–2668.

O'Connor, J., Acharya, V., & Lieuwen, T. (2015). Transverse combustion instabilities: Acoustic, fluid mechanic, and flame processes. *Progress in Energy and Combustion Science*, *49*, 1–39.

Rasmussen, C. E., & Williams, C. (2006). *Gaussian processes for machine learning*. The MIT Press, Cambridge, MA, USA.

Sarkar, S., Jha, D., Lore, K., Sarkar, S., & Ray, A. (2016). Multimodal spatiotemporal information fusion using neural-symbolic modeling for early detection of combustion instabilities. In *2016 American Control Conference (ACC)* (p. 4918-4923).

Sarkar, S., Lore, K. G., & Sarkar, S. (2015). Early detection of combustion instability by neural-symbolic analysis on hi-speed video. In *Workshop on cognitive computation: Integrating neural and symbolic approaches (coco@ nips 2015), Montreal, Canada.*

Sarkar, S., Lore, K. G., Sarkar, S., Ramanan, V., Chakravarthy, S. R., Phoha, S., & Ray, A. (2015). Early detection of combustion instability from hi-speed flame images via deep learning and symbolic time series analysis.

Schmid, P. J. (2010). Dynamic mode decomposition of numerical and experimental data. *Journal of fluid mechanics*, *656*, 5–28.

Sé, Ducruix, b., Schuller, T., Durox, D., Sé, & Candel, b. (2003). Combustion dynamics and instabilities: Elementary coupling and driving mechanisms. *Journal of Propulsion and Power*, *19*(5), 722–734.

Virani, N., Jha, D., & Ray, A. (2016). Sequential hypothesis tests using markov models of time-series data. In 1st *ACM SIGKDD workshop on machine learning for prognostics and health management, San Fransisco, CA*.

APPENDIX

Experimental apparatus

The experimental apparatus is a swirl-stabilized combustor with a swirler of diameter $30\,\text{mm}$ with $60°$ vane angles (i.e., geometric swirl number of 1.28). Air is fed into the combustor through a settling chamber of diameter $280\,\text{mm}$ with a sudden contraction leading to a square cross section of side $60\,\text{mm}$, providing an acoustically open condition with area ratio of 17. A mesh and honeycomb structure at the immediate downstream of the contraction assures uniform flow to the swirler. The combustor, shown in figure 11 consists of a $200\,\text{mm}$ long inlet section, an inlet optical access module (IOAM) of length $100\,\text{mm}$, a primary combustion chamber of length $370\,\text{mm}$, and secondary duct of the same length. The overall length of the constant area ducts was chosen to be $1340\,\text{mm}$. The fuel injection tube is coaxial to a mixing tube which has the same diameter as that of the swirler. The bypass air that does not enter the mixing tube passes through slots on the swirl plate. The slots on the fuel injection tube are drilled at designated distance upstream of the swirler, which dictates the extent of premixing between fuel and air. The larger this distance, more homogeneous the air-fuel mixture is. Two upstream distances of $90\,\text{mm}$ and $120\,\text{mm}$ were chosen for fuel injections during the experiments, where the former of the two denotes partial premixing and the later provides full premixing. The hi-speed images were collected through IOAM at $3\,\text{kHz}$ using Photron High speed star with a spatial resolution of 1024×1024 pixels. Synchronized pressure data was acquired using piezoelectric transducers (PCB make) with resolution $225\,\text{mV}\,\text{kPa}^{-1}$ at a location downstream of the IOAM. The data acquisition was triggered simultaneously using NI card and taken for a duration of $3\,\text{s}$ yielding in a sequence of $9,000$ images for every operating condition. More details of the combustor could be found in (Sarkar, Lore, & Sarkar, 2015).