Supplementary Material

On-line in-situ prediction of 3D flame evolution from its 2D projections *via* deep learning

Jianqing Huang, Hecong Liu, Weiwei Cai\*

*Key Lab of Education Ministry for Power Machinery and Engineering, School of Mechanical Engineering, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, China, 200240*

*\*cweiwei@sjtu.edu.cn*

**1. VOLUMETRIC TOMOGRAPHY SYSTEM**

The experimental setup of volumetric tomography (VT) system consists of a swirl-stabilized burner, a customized fiber bundle (Nanjing Chunhui Science and Technology Industrial Co. Ltd), one high-speed camera (Photron FASTCAM Mini AX100) and a workstation. Nine objective lenses were placed in front of the input ends of the fiber bundle to simultaneously collect different projections of the flame generated by the swirl-stabilized burner. The projections were arranged to be a $3×3$ array at the output end and then captured by the camera. The camera was equipped with a micro-lens and an extension ring to achieve a proper magnification. There were two input channels for the burner, the outer one for swirling air and the inner one for methane. The flows were controlled by Alicat mass flow controllers (21-1-00-1-100-KM6001, for air; 21-1-12-1-20-KM6001, for methane). Two types i.e., a simple laminar diffusion flame (Flame #1) and a complex turbulent swirl-stabilized flame (Flame #2) were tested in the experiments. More details about the experiments are listed in Table 1.

The nine projections recorded by the camera can be used to recover the 3D structures of the flames and the imaging model can be mathematically established as ([Liu *et al.* 2019](#_ENREF_4)):

 $A\rightharpoonaccent{x}=\rightharpoonaccent{p } $, (2)

where *A* is the weight matrix which relates the 3D distribution of the flame chemiluminescence $\rightharpoonaccent{x}$ with the projections $\rightharpoonaccent{p }$. Equation (2) is a linear equation system and is typically ill-posed, and can be solved with a well-established iterative algorithm such as ART ([Yu *et al.* 2018](#_ENREF_5)) which can be mathematically described as:

 $\rightharpoonaccent{x}^{(k+1,i)}=\rightharpoonaccent{x}^{(k,i)}-λ\_{ART}×\frac{A\_{i}×\rightharpoonaccent{x}^{\left(k,i\right)}-p\_{i}}{\left‖A\_{i}\right‖\_{2}^{2}}×A\_{i}^{T}$, (3)

where $\rightharpoonaccent{x}^{(k+1,i)}$ indicates the solution within the (*k+*1)-th iteration after updating according to the *i*-th equation; $λ\_{ART}$ is a relaxation factor which controls the convergence rate; symbol $\left‖∙\right‖\_{2}^{2}$ indicates the square of the 2-norm of a vector; and superscript *T* means the transpose of a matrix. In this work it typically took about 30 minutes to process one frame of tomographic data.

Table 1. Operation conditions of the experiment.

|  |  |
| --- | --- |
| Conditions for the tested flames | Condition for the camera |
|  | Flame #1 | Flame #2 |  |
| Equivalence ratio | 0.5 | 0.1 | Framing rate | 1 kHz |
| Air flow rate | 7.6 L/min | 59.4 L/min | Exposure time | 1 ms |
| Methane flow rate | 0.4 L/min | 0.6 L/min | Pixel resolution | $$1024×1024$$ |
| Swirl number | 0.87 | 0.87 | Focal length | 55 mm |

**2. DETAILS OF NEURAL NETWOKS TRAINING AND OPTIMIZATION**

For the readers’ convergence, the training process of our hybrid CNN-LSTM model will be described in two separate parts i.e., the CNN part and the LSTM part.

Convolutional neural network (CNN) features local receptive fields, shared weights and pooling, which makes it a highly effective approach to extracting features from images ([Huang *et al.* 2018](#_ENREF_1)). The designed CNN framework is shown in Table 2. There are nine input channels corresponding to nine projections of the flame from different perspectives. The hidden layer consists of three convolutional layers, three max-pooling layers and two dense layers. To avoid overfitting and obtain a robust training model, three dropout layers are embedded in the hidden layer. The activation function was selected to be leaky ReLU and the learning rate was set to automatic adjustment mode, which are both conducive to obtain better training results. When we train the CNN to extract useful features from projections, the ground truth was the corresponding 3D flame structure at the same time instant. The parameters of the network were optimized using Adam ([Kingma & Ba 2017](#_ENREF_2)). As per other key hyper-parameters e.g. the learning rate and the batch size were optimized to be 0.001 and 50, respectively. After about 30 epochs, the loss function L2 converged, which means that the training process of CNN has been finished. Subsequently, using the trained CNN model, each 3D flame structure in the dataset can be represented by its corresponding feature vector (i.e., the output of the flatten layer).

Table 2. The framework of the CNN network

|  |  |  |
| --- | --- | --- |
| Layer | Name | Size |
| 1 | Input | 120×60×9 |
| 2 | Conv2D + LeakyReLU | 116×56×64 |
| 3 | Dropout | 116×56×64 |
| 4 | Max-Pooling2D | 58×28×64 |
| 5 | Conv2D + LeakyReLU | 54×24×128 |
| 6 | Dropout | 54×24×128 |
| 7 | Max-Pooling2D | 27×12×128 |
| 8 | Conv2D + LeakyReLU | 24×9×32 |
| 9 | Dropout | 24×9×32 |
| 10 | Max-Pooling2D | 12×4×32 |
| 11 | Flatten | 1536 |
| 12 | Dense + LeakyReLU | 500 |
| 13 | Dense + LeakyReLU | 275000 |
| 14 | Output | 50×50×110 |

The long short-term memory (LSTM) neural network is a powerful deep learning method for sequence modeling due to its special memory blocks i.e., the LSTM cell in the hidden layer ([Lipton 2015](#_ENREF_3)). There are three gates including an input one, a forget one and an output one in each LSTM cell, which controls the information addition, removal and passage to the next block, respectively. As shown in Table 3, when we train the LSTM to predict the frame in the future, the window of the history input data was set to be 10 frames (i.e., *h*=10 ms). There were totally 30 epochs for the training process. The optimizer was Adam and the learning rate was determined to be 0.001 as well. It has to be noted that the LSTM network was trained to predict the feature vector of 3D flame structure, instead of directly predicting the full flame structure. When both the CNN and LSTM networks were trained, LSTM was connected to the CNN behind its flatten layer. Thus, the hybrid CNN-LSTM model was obtained and can be used to predict the 3D flame evolutions based on its history 2D projections.

Table 3. The framework of the LSTM network

|  |  |  |
| --- | --- | --- |
| Layer | Name | Size |
| 1 | Input | 10×1356 |
| 2 | LSTM | 1024 |
| 3 | Dense | 1536 |
| 4 | Output | 1536 |

**References**

Huang, J., Liu, H., Dai, J., Cai, W., 2018. Reconstruction for limited-data nonlinear tomographic absorption spectroscopy via deep learning. Journal of Quantitative Spectroscopy and Radiative Transfer 218, 187-193

Kingma, D.P., Ba, J., 2017. Adam: A Method for Stochastic Optimization. arXiv preprint arXiv 1412, 6980

Lipton, Z.C., 2015. A critical review of recurrent neural networks for sequence learning. arXiv preprint arXiv 1506, 00019

Liu, H., Sun, B., Cai, W., 2019. kHz-rate volumetric flame imaging using a single camera. Optics Communications 437, 33-43

Yu, T., Ruan, C., Liu, H., Cai, W., Lu, X., 2018. Time-resolved measurements of a swirl flame at 4 kHz via computed tomography of chemiluminescence. Applied Optics 57, 5962-5969