Machine Learning for Super-resolution and Coherent Structure Identification in Turbulent Boundary Layer flows.

Sahil Kommalapati

ksahil@uw.edu

Objective

Develop Machine Learning based frameworks to analyze turbulent boundary layer data acquired from Particle Image Velocimetry (PIV) for coherent structure identification and super-resolution inference. The first portion of this project aims to identify and characterize turbulent coherent structures using a Bayesian optimization technique. This approach allows the automated identification of structures that take a large number of parameters to be modelled posing a high-dimensional problem. The second part of the thesis will evaluate the effectiveness of a number of current super-resolution techniques to infer small scale details in a fully turbulent flow from a lower resolution dataset of PIV data of turbulent boundary layer flow over a flat plate.

Justification for Research

Traditionally Machine Learning based approaches have been extensively used in Flow control applications. However, recently there has been a heightened interest targeting most fundamental aspects of Fluids and specifically turbulence research. This recent boom in interest can be attributed to increased computational power and the development of efficient machine learning training algorithms. Of particular interest is the super-resolution and sub-grid scale modelling problem, that has many applications in fluid dynamics. Turbulent flow evolves on many scales. Direct numerical simulation (DNS) of all turbulent length scales is extremely expensive. An alternative is to use Large eddy simulations (LES) of only the larger length scales. On the other hand, experimental techniques to extract all scales is restricted by resolution limits of measuring equipment. Hence, it is extremely valuable to reconstruct smaller scales of turbulence from large-scale data.

We propose utilizing Machine Learning algorithms for this task. However, it is still unclear as to precisely estimate the amount of data required to obtain a converged solution for a multi-scale problem, such as turbulence, or how generalizable the solution will be to other conditions. Ideally, known conservation laws should be incorporated into such models, however this is challenging and has not yet been extensively explored.

Utilizing DNS data, researchers have implemented max-pooling or average pooling [1] to generate low-resolution data from which the high resolution velocity fields are reconstructed. However, the application of these methods on experimental data has partially been limited by a need for large, statistically independent datasets. My thesis is designed to address this problem, by acquiring a large dataset of turbulent velocity fields using particle Image velocimetry (PIV) and then applying machine learning for the super-resolution on this data set. With these results, comparisons can be made on the effectiveness of ML models that have been developed on DNS data, against experimental PIV data.

A second challenge for the analysis of turbulent flows has been the robust identification of complex turbulent coherent structures. While many approaches have been proposed and employed, there are still challenges automating the identification of hairpin vortex signatures from instantaneous velocity fields, something that can be done visually by eye. This thesis will aim to rectify this issue by using a Bayesian approach. This approach has been examined over the last few quarters as it holds promise for its ability to accurately match high-dimensional flow models to experimental data. This technique has been applied to the extraction of convective velocities, vortex circulation strength and radius in a turbulent boundary layer dataset. A paper detailing the Bayesian framework and its results is currently being prepared for publication.

Approach and Methodology

Approach for Vortex Identification.

A single scale PIV data of boundary layer flow over a flat plate is available at the Williams lab. After initial pre-processing like removing noise and interpolating missing vector data, this data is ready to be matched with coherent structure models. The Emcee open source Python module [2] is an excellent Bayesian resource to implement Markov chain Monte Carlo (MCMC) samplers that are Affine Invariant. The results obtained from the Bayesian optimization framework are compared against the solutions obtained from traditional minimization optimization for validation. Finally, a Python based open-source package is to be developed that can be released to simplify the process our integrating this research by other research groups. Work towards this package has already gained a significant pace.

Approach for super-resolution.

Multi-scale turbulent PIV information of a turbulent boundary layer flow will be collected by using 2 cameras in the 3x3 wind tunnel of UW-AA. These cameras will employ an overlapping large and smaller fields of view. One camera will focus on the large scale coherent structures in the flow. The other camera is zoomed in and focuses on the smaller scales observed on a finer grid. A large dataset is acquired to obtain a large number of samples of all scales of turbulence.

Previously, several algorithms and frameworks have been developed by employing Convolutional Neural Networks [1], Shallow [3] and Deep Neural Networks, Extreme Learning Machines (ELMs) and Generative Adversarial Networks [4] (GANs) amongst a plethora of others, to address the super-resolution problem using data generated through Direct Numerical Simulations (DNS).

Shallow neural networks have simple architectures and take lower computational effort to tune and train. They have shown high efficiencies on super-resolution of forced isotropic turbulent DNS data. This architecture is to be reconstructed and implemented on PIV data. Based on analogies between non-linear neural networks and principle component analysis, the dominant modes contributing towards the super-resolution can be visualized. Finally, results from this implementation can be validated against super-resolution approaches using other ML models. For example, GANs (with temporal discriminators) and the Hybrid Down sampled Skip-Connection Multi-Scale models.

Schedule of Work

The thesis work is proposed to span over the upcoming three quarters, concluding in Spring 2021.

Autumn 2020

Results of the Bayesian coherent structure identification project will be prepared for publication. This work will be presented at APS DFD 2020.

PIV data will be collected from the 3x3 wind tunnel, using two cameras capturing two different scales of the flow. This would ultimately be the training data for the machine learning models. In the meantime, there is an existing PIV dataset of a low Reynolds number turbulent boundary layer with 750 velocity fields that can be used for initial testing of SNN based super-resolution testing. For preliminary computations, max-pooling or average pooling can be employed to down-sample and generate the testing, training and validation data sets for the ML models.

Winter 2021

The Shallow neural network (SNN) framework employed single scale PIV data last quarter. This quarter, we can test the framework by utilizing multi-scale PIV data captured in Autumn. Since, this dataset is much larger, validation accuracies are expected to be improved.

Further, Super-resolution using the Hybrid Down sampled Skip-Connection Multi-Scale (DSC/MS) model and Generative Adversarial networks (GAN) are implemented by utilizing the newly captured PIV data. A rigorous comparative study is to conducted to evaluate the individual performances of all the three processes that have been employed so far.

Tuning the hyper-parameters in Machine Learning models is as crucial as finalizing the model itself. Hyperband [5] (from UW-CSE) based hyper-parameter tuning is promising to automate the optimization process. This could be employed to tune the above models. Based on time constraints, a Bayesian optimization based hyper-parameter tuning can also be implemented to get another estimate. The open-source ML module in Python, Keras, provides a simple package to implement both of these techniques.

Spring 2021

Conclude fine-tuning the ML models i.e. Shallow Neural Networks, DCS/MS model and GANs and training on the models using the (extremely) big PIV datasets. The advantages and short comings of each implementation would be documented and compared, which help in laying steps for future directions.

Finish completing the writing for the thesis that would have been concomitantly going on, through all of the previous quarters. Prepare for Graduation by the end of this quarter.

Estimated Cost

The cost associated with the thesis is the time required to be spent at the 3x3 wind tunnel and the computers associated with the PIV setup. These resources, upon proper scheduling are available to all students at UW. Hence, there are no explicit costs to be associated with the research.

Committee Member names and E-mail addresses

Prof. Owen Williams (Chair), ojhw@uw.edu
Prof. Steven Brunton, sbrunton@uw.edu
Prof. Brian Polagye, bpolagye@uw.edu

Committee Chair's Signature: ______________________
Date Signed by Committee Chair: 10/9/20

[1] K. Fukami, K. Fukagata and K. Taira, "Super-resolution analysis with machine learning for low-resolution flow data," *11th International Symposium on Turbulence and Shear Flow Phenomena, TSFP,* 2019.

[2] D. Foreman-Mackey, D. W. Hogg, D. Lang and J. Goodman, "emcee: the MCMC hammer," *Publications of the Astronomical Society of the Pacific,* vol. 125, no. 925, p. 306, 2013/2/25.

[3] N. B. Erichson, L. Mathelin, Z. Yao, S. L. Brunton, M. W. Mahoney and J. N. Kutz, "Shallow neural networks for fluid flow reconstruction with limited sensors," *Proceedings of the Royal Society A,* vol. 476, no. 2238, p. 20200097, 2020.

[4] Y. Xie, E. Franz, M. Chu and N. Thuerey, "tempoGAN: A temporally coherent, volumetric GAN for super-resolution fluid flow," *ACM Transactions on Graphics (TOG),* vol. 37, no. 4, p. 95, 2018.

[5] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh and A. Talwalkar, "Hyperband: A novel bandit-based approach to hyperparameter optimization," *The Journal of Machine Learning Research,* vol. 18, no. 1, pp. 6765-6816, 2017.