Semi-Supervised Machine Learning in Data-Driven Flow Measurement

Junwei Chen* , Stefano Discetti and Marco Raiola[1]
1: Dept. of Aerospace Engineering, Universidad Carlos III de Madrid, Spain
* Correspondent author: junwei.chen@uc3m.es

Keywords: PIV processing, Machine learning, Data driven, POD

ABSTRACT In the previous work of Chen el at. (2022a), data-driven techniques demonstrated to be able to enrich snapshot PIV (Particle Image Velocimetry) with the time resolution of fast probe data. In this paper, a semi-supervised machine learning method is introduced, which allows to train not only on probe data labeled with simutaneous PIV outputs but also on unlabeled data from probe-only measurements. Then the new machine learning method is compared with EPOD and MLP, shows it's prevailence both in the accuracy and reducing the computation cost.

1 INTRODUCTION

Time-resolved PIV (TR-PIV) provides both spatial and temporal information, which is a powerful tool in locating the evolution of coherent structures, identifying physical phenomena and allows computing pressure fields from integration of the Navier-Stokes equations. However, it is not always achievable due to the budget and the restriction of laser and camera repetition rate. A good surrogate of TR-PIV can be obtained by combining snapshot PIV (no need in temporal resolution) and fast probe measurement. This has been recently applied in Chen el at. (2022a) using the EPOD (Extended Proper Orthogonal Decomposition) and Chen el at. (2022b) using the classical multilayer perceptron (MLP), aiming to pressure field reconstruction. Benefited from the capability to map nonlinear relations, the MLP outperforms the EPOD. In detail, the flow field $\boldsymbol{U}$ can be temporal-spatial decoupled via POD,

$$\boldsymbol{U}(\boldsymbol{x},t)=\sum_i \sigma_i a_i(t)\boldsymbol{\phi}_i(\boldsymbol{x}), \tag{1}$$

where the $\boldsymbol{x}$ and t stands for position and time, σ_i is the weight of the i$^{\text{th}}$ mode, and a_i is the coefficient for normalized orthogonal spatial basis $\boldsymbol{\phi}_i(\boldsymbol{x})$. The MLP regresses the relationship f between the probe data $\boldsymbol{p}$ and the coefficients $\boldsymbol{a}$,

$$\widehat{\boldsymbol{a}}(t)=f(\boldsymbol{p}(t)), \tag{2}$$

where $\widehat{\boldsymbol{a}}$ is the coefficient predicted via the model f, and the task is to find the optimal model,

$$\underset{f}{\operatorname{argmin}}\,|\boldsymbol{a}-f(\boldsymbol{p})|. \tag{3}$$

Despite the advancements in machine learning, the prediction accuracy of MLP still exhibits noticeable deviations from the ground truth, particularly when the POD spectrum lacks compactness. One approach to further enhance the mapping capability of the MLP is using semi-supervised techniques (Chapelle et al., 2006).

It must be remarked that the operative cost of recording $\boldsymbol{a}$ and $\boldsymbol{p}$ is different. The coefficients $\boldsymbol{a}$ are obtained by PIV field, which requires storage and processing recourses far superior than point probes $\boldsymbol{p}$. This limits the amount of simultaneous $\boldsymbol{p}$ and $\boldsymbol{a}$ available for supervised machine learning (i.e. fulfilling Eq. 3). However, if the model f confirms the assumption of translation invariance, which means when the model is trained well at time t_j, $\left|\boldsymbol{a}(t_j)-f\left(\boldsymbol{p}(t_j)\right)\right|\to 0$, it should work well also at time t_{j+k}, $\left|\boldsymbol{a}(t_{j+k})-f\left(\boldsymbol{p}(t_{j+k})\right)\right|\to 0$, where $k=0,\pm1,\pm2,\ldots$ This can be fulfilled by introducing another model g on the time derivative of $\boldsymbol{a}$,

$$\widehat{\boldsymbol{a}}_t(t)=g(\boldsymbol{p}(t)), \tag{5}$$

and the optimization of f and g does not require to be supervised with the coefficient $\boldsymbol{a}_j$,

$$\underset{f,g}{\operatorname{argmin}}\left|\frac{f\left(\boldsymbol{p}(t_{j+1})\right)-f\left(\boldsymbol{p}(t_{j-1})\right)}{t_{j+1}-t_{j-1}}-g\left(\boldsymbol{p}(t_j)\right)\right|_j. \tag{6}$$

The training session includes 3 steps, and the current f and g are two MLP models,

(1) training f as a starter, which is actually the single MLP,
(2) training g using the model f in (1),
(3) training f and g together.

The training is supervised for f in step (1) and partially in step (3), while unsupervised for and g for partially f in step (3). This allows enriching the training by adquiring additional sequences of $\boldsymbol{p}$ without corresponding PIV measurements. In addition, the supervised machine learning for data-driven measurement is only trained in several individual snapshots without knowing the evolution of the flow, but the unsupervised part can be trained in series of frames and tracking the variation of $\boldsymbol{a}$ in adjacent frames, thus suppressing unphysical temporal fluctuations.

2 VALIDATION

The semi-supervised machine learning was tested in different data sets including synthetic and experimental ones. In this abstract the same experiment on the wake of a wing already tested in Chen el at. (2022a) is included. The snapshot PIV is down sampled from the time-resolved one, and probe data are extract in several points of the PIV field to simulate probe data. Finally the pressure integrated from the velocity estimation of EPOD, MLP and the semi-supervised machine learning are compared to that from time-resolved PIV. An example of reconstructed pressure field is reported below.

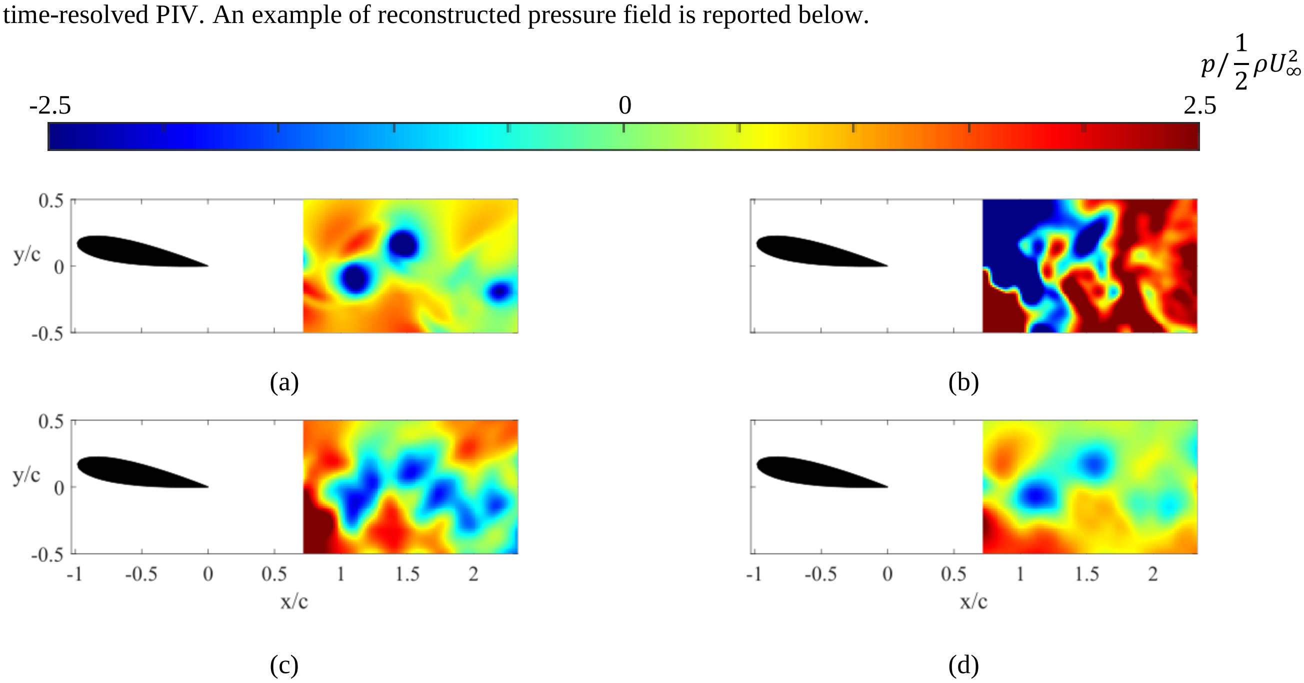

Figure 1. The pressure integrated from (a) time-resolved PIV field, (b) EPOD, (c), MLP and (d) semi-supervised machine learning.

Figure 1 shows the semi-supervised machine learning depicts high and low pressure areas more accurate to the MLP and EPOD. The MLP reduced the root-mean-square error of the domain shown in the figure and over 1000 frames by 50% and 73% comparing to the EPOD in velocity and pressure field reconstruction, and the semi-supervised machine learning gives a further reduction by 6% and 10%.

3 ACKNOWLEDGEMENT

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 949085).

4 REFERENCES

- Chapelle, O., Schölkopf, B., & Zien, A. (2006). A discussion of semi-supervised learning and transduction. In Semi-supervised learning (pp. 473-478). MIT Press.
- Chen, J., Raiola, M., & Discetti, S. (2022a). Pressure from data-driven estimation of velocity fields using snapshot PIV and fast probes. Experimental Thermal and Fluid Science, 136, 110647.
- Chen, J., Raiola, M., & Discetti, S. (2022b). Data-driven estimated velocity and pressure fields using snapshot PIV and fast probes. 20th International Symposium on Application of Laser and Imaging Techniques to Fluid Mechanics.