Commun. Comput. Phys.
doi: 10.4208/cicp.OA-2018-0035

Vol. **25**, No. 3, pp. 625-650
March 2019

# Data-Driven, Physics-Based Feature Extraction from Fluid Flow Fields using Convolutional Neural Networks

Carlos Michelén Ströfer[1,*], Jin-Long Wu[1], Heng Xiao[1] and Eric Paterson[1]

[1] *Kevin T. Crofton Department of Aerospace and Ocean Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA.*

Received 5 February 2018; Accepted (in revised version) 31 May 2018

**Abstract.** Feature identification is an important task in many fluid dynamics applications and diverse methods have been developed for this purpose. These methods are based on a physical understanding of the underlying behavior of the flow in the vicinity of the feature. Particularly, they require the definition of suitable criteria (i.e. point-based or neighborhood-based derived properties) and proper selection of thresholds. However, these methods rely on creative visualization of physical idiosyncrasies of specific features and flow regimes, making them non-universal and requiring significant effort to develop. Here we present a physics-based, data-driven method capable of identifying any flow feature it is trained to. We use convolutional neural networks, a machine learning approach developed for image recognition, and adapt it to the problem of identifying flow features. This provides a general method and removes the large burden placed on identifying new features. The method was tested using mean flow fields from numerical simulations, where the recirculation region and boundary layer were identified in several two-dimensional flows through a convergent-divergent channel, and the horseshoe vortex was identified in three-dimensional flow over a wing-body junction.

**AMS subject classifications**: 76G25, 76M99

**Key words**: Machine learning, feature identification, fluid dynamics, convolutional neural network.

## 1 Introduction

Feature detection is an important component of data post-processing in fluid dynamics experiments and simulations, and plays an important role in our physical understanding

*Corresponding author. *Email addresses:* `cmich@vt.edu` (C. Michelén Ströfer), `jinlong@vt.edu` (J.-L. Wu), `hengxiao@vt.edu` (H. Xiao), `egp@vt.edu` (E. Paterson)

 

of flow phenomena and fluid-structure interactions. In this study we use convolutional neural networks, a machine-learning approach developed for image recognition, to automatically detect features of interest in fluid flow fields. As with traditional methods, feature detection is done by identification of patterns within the relevant physics-based scalar fields of the flow. Unlike traditional methods, the specific patterns to use are not explicitly specified, but inferred from a set of human-labeled training examples.

### 1.1 Feature detection in fluid flow fields

A flow feature is a physically meaningful structure within the flow that is of interest for the application at hand. Examples include recirculation regions, shed vortices, boundary layers, and shock waves. Applications of flow feature extraction include fundamental physical understanding of flow dynamics (e.g. relation between coherent structures and turbulence dynamics [1, 2]), engineering design (e.g. reduction of shock wave drag [3]), on-line steering of large simulations (e.g. feature-based adaptive mesh [4, 5]), among others. Despite the important role flow features play in so many applications, the task of accurately identifying these features remains a laborious one.

Post et al. [6] describe the general *feature extraction pipeline* as consisting of four steps: selection, clustering, attribute calculation, and iconic mapping. *Selection* consists of identifying all points that are part of the feature of interest, and the output of this step is a binary map. In the *clustering* step these points are divided into discrete instances of the feature of interest. For instance, the input to the clustering step might be a binary map of points that belong to vortices, and the output are groupings of these points into coherent regions, e.g. distinct vortices. The last two steps, *attribute calculation* and *iconic mapping*, consist of calculating properties of each coherent region (e.g. volume, length, centroid) and fitting a parametrized icon (e.g. ellipsoid). This has the effect of reducing the dimensionality of the problem, allowing the description of the evolution of features with only a few parameters. In this paper we focus on the first two steps, *selection* and *clustering*, and only imply these steps when referring to *feature extraction* or *identification*.

Many methods exist for identifying features based on an understanding of the underlying physics of the phenomena. As an example, methods for identifying vortices include identification based on local field values (e.g. vorticity, helicity, $Q$-criterion), as well as methods based on global flow properties (e.g. curvature center, looping streamlines) [6]. The disadvantages with these methods are that they are specific to the feature in question, limited to certain types of flows (e.g. external flows, turbo-machinery) [6], and rely on creative visualization of the physical phenomena at hand. This has lead to many disjoint methods particularly suited for very specific problems, and places a large burden on developing methods for identifying new features.

As a specific example of the need for improved flow feature extraction we consider the problem of turbulence. Turbulent flows contain temporally-coherent structures (flow features) at various scales. These coherent structures are known to play an important role in mass, momentum, and energy transport [1, 2] and subsequently have an effect on