Investigating Optimal Training and Uncertainty Quantification for CNN-based Optical Flow

D. Kurihara[1], G. Blois[1], H. Sakaue[1], D.E. Schiavazzi[1,2]

[1] University of Notre Dame, Department of Aerospace and Mechanical Engineering, Notre Dame, USA

[2] University of Notre Dame, Department of Applied and Computational Mathematics and Statistics, Notre Dame, USA

1 Introduction

Optical Flow (OF) techniques provide "dense estimation" flow maps (i.e. pixel-level resolution) of time-correlated images and thus are appealing to applications requiring high spatial resolutions. OF methods revolve around mathematical descriptions of the image as a collection of *features*, in which the pixel-level light intensity is the primary variable (Horn and Schunck, 1981). Feature tracking often involves the notion of scale invariance. Traditional OF approaches, merely based on mathematical formulations, have suffered from many challenges, especially when directly applied to images of fluid flows textured with tracer particles (hereafter PIV-like images). Due to the limited number of computationally manageable features and sub-optimal regularization methods, successful implementation of past approaches has been limited to highly textured images and small displacement dynamic ranges.

Recent deep learning-based methods have effectively removed several limitations, offering an opportunity to revitalize OF techniques. Being these based on structural features, the natural ability of artificial neural networks (ANN) with deep architectures to extract a large collection of such features at multiple resolutions through convolutions with learnable kernels can be brought to bear. In this context, a number of newly proposed Convolutional Neural Networks (CNNs), originally developed for computer vision applications, have been successfully applied to the estimation of *flows* from dynamic scenes with moving objects. A recent example is LiteFlowNet (Fig. 1(a)), which provides state-of-the-art performance on a number of datasets including rigid body motion of objects in space, vehicles moving in traffic and, most notably, computer animated sequences (Hui et al., 2018). However, there is still limited understanding of which network setup (e.g. architecture, hyperparameter selection, etc.) provides optimal flow accuracy for PIV-like images.

In this context, the goal of this study is to 1) provide a quantitative understanding of how LiteFlowNet performs for PIV-like images under different training paradigms and hyperparameter setups, and also to 2) extended the capabilities of LiteFlowNet to quantify flow uncertainty.

2 Methods

LiteFlowNet is a family of recently proposed deep neural networks for optical flow estimation (Hui et al., 2018, 2020). Its architecture consists of pyramidal feature extraction (*NetC*) followed by a cascade of flow inference modules (*NetE*), i.e., a matching module, a subpixel module, and a regularization module. The matching module identifies the corresponding features in the image pair, the subpixel module corrects the flow estimation and provides subpixel accuracy, whereas the regularization module is used to refine the flow estimate near the boundary of moving objects. The network progressively identifies flow features in a coarse-to-fine resolution pipeline, starting from level 6 (coarser resolution) to level 2 (finer resolution). The original LiteFlowNet was trained using a staged process (see Fig. 1(a)), and did not support quantification of flow uncertainty. The first portion of this work assesses the predictive performance of LiteFlowNet using available pre-trained weights. Tests targeted synthetic PIV-like image sets with varying particle density, size and displacement. We then explored the potential of LiteFlowNet when trained from PIV-specific examples, following two different training paradigms proposed in the literature. In addition, we augmented the LiteFlowNet architecture with dropout layers providing both a regularization mechanism and the possibility to estimate uncertainty from prediction ensembles.

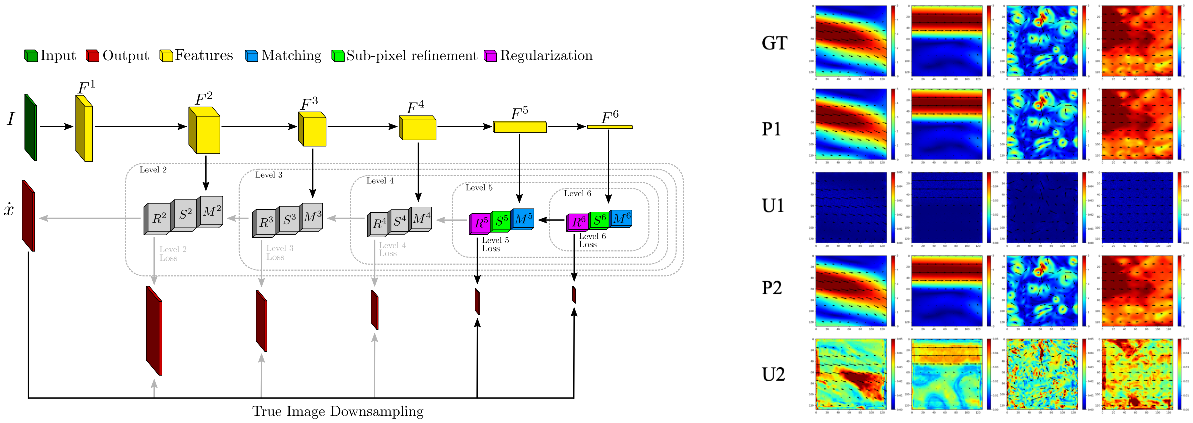

Figure 1: (Left) LiteFlowNet architecture and schematics of a staged training sequence. (Right) Prediction of instantaneous flow fields: the first row (GT) refers to the ground truth. Predictions are obtained from end-to-end training with dropouts (P1) and with optimal loss penalties from all levels (row P2). Standard deviations from ensemble predictions for P1 and P2 are shown in U1 and U2, respectively.

3 Results

First, three sets of weights were tested (referred to as *Default*, *KITTI* and *Sintel*), and their predictions compared against PIV image processing using three interrogation window sizes (64, 32, and 16), with 50% overlap. The comparisons were made by statistical analysis of large samples, using average absolute pixel differences under different up- and down-scaling interpolation strategies. Sintel weights offered the best performance due to the presence of image deformation and shear in the training dataset, unlike *Default* and *KITTI* which are mainly trained from rigidly moving examples. Sintel results were comparable to those obtained using PIV correlations. The first newly tested training approach consisted of multiple sessions where finer image resolution levels are progressively activated at each successive *stage* (*staged training*, see Fig. 1(a)). The second is referred to as *end-to-end* training, where the complete network is trained all at once, and the loss is calculated as a penalized aggregation of prediction errors from all levels. End-to-end training generally produces both minimal losses and best accuracy. However, it is less clear how to *weight* the losses from each level, particularly for a general system which may operate on a wide range of particle densities, for which the importance of the features at various levels may vary within the same training dataset. Finally, we studied layouts with a dropout layer positioned after each flow inference module and after all modules, achieving the best performance when placing a dropout layer after each matching module for all levels.

To compare all the approaches described, we trained our modified network on three PIV-specific datasets generated through numerical simulation (Cai et al., 2019). Results in Fig. 1(b) show consistent performance improvements of the proposed network leveraging dropouts (P1) which contained small uncertainty (U1) for all the datasets considered. We also determined optimal weighted losses from various levels, obtaining reduced accuracy (P2) and larger uncertainties (U2), confirming the key importance of high resolution losses and providing grounds for designing more efficient network layouts.

References

Cai S, Liang J, Gao Q, Xu C, and Wei R (2019) Particle image velocimetry based on a deep learning motion estimator. *IEEE Transactions on Instrumentation and Measurement* 69:3538–3554

Horn BK and Schunck BG (1981) Determining optical flow. in *Techniques and Applications of Image Understanding*. volume 281. pages 319–331. International Society for Optics and Photonics

Hui TW, Tang X, and Loy C (2018) LiteFlowNet: A Lightweight Convolutional Neural Network for Optical Flow Estimation. in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pages 8981–8989

Hui TW, Tang X, and Loy C (2020) A Lightweight Optical Flow CNN - Revisiting Data Fidelity and Regularization