Key aspects of unsupervised optical flow models in PIV applications

C. Lagemann\textsuperscript{1}\textsuperscript{*}, W. Schröder\textsuperscript{1}

\textsuperscript{1} RWTH Aachen University, Chair of Fluid Mechanics and Institute of Aerodynamics, Aachen, Germany
\textsuperscript{*} c.lagemann@aia.rwth-aachen.de

Abstract

In recent years, several algorithms have been proposed that apply deep learning techniques to the data analysis workflow of Particle-Image Velocimetry (PIV) experiments. These techniques have shown potential to match or even surpass classical algorithms in terms of efficiency, accuracy, and spatial resolution. However, the diversity in dynamic flows and particle image conditions can pose a challenge to fully supervised learning-based PIV tools. If conditions vary significantly from the synthetic training data, the performance can degrade substantially. Unsupervised learning offers a solution to this issue by eliminating the need for a general synthetic dataset. This approach can theoretically improve the inference capability of deep learning models and might tailor them better to real-world experimental data. However, designing an adequate unsupervised loss objective is not trivial due to the variety of different penalty terms and requires application-specific consideration. Therefore, we investigate systematically key components of unsupervised losses and regularizers for PIV applications in this study. We derive empirically a proxy loss well-suited for PIV and demonstrate the effectiveness of our model called URAFT-PIV under varying image and lighting conditions using both synthetic and experimental data. Our results provide new insights into the capabilities of unsupervised deep learning for PIV processing.

1 Introduction

In contrast to cross-correlation based PIV schemes which typically compute displacement estimates for the entire interrogation window \cite{raffel2018}, optical flow based methods operate at a much higher resolution and compute a displacement estimate for every pixel individually. Especially in the field of PIV, the performance of optical flow techniques has recently seen dramatic improvements originating from the adoption of deep learning approaches. Recent work has demonstrated that these deep learning based optical flow techniques have the potential to match or outperform state-of-the-art classical algorithms in terms of efficiency, accuracy, and spatial resolution. Valuable examples can be found in \cite{cai2019, lagemann2019, zhang2020, lagemann2021b, morimoto2021} etc. However, while work on deep learning for PIV is promising, there remain open questions concerning the application of the almost exclusively supervised models to general PIV data. These challenges originate from the fact that supervised optical flow techniques are exclusively trained on synthetic data since ground truth information is required but difficult to obtain for real image pairs. As a result, the feature distribution of synthetic and realistic PIV images might differ substantially causing an inherent distribution mismatch that supervised approaches might struggle to overcome. In contrast, unsupervised training procedures do not rely on known ground truth data for supervision since all applied penalty terms are based on geometric constraints. Therefore, the required training data exhibiting real-world measurement conditions for training an unsupervised model is abundant since unsupervised training only requires unlabeled image data. This fundamental difference is at the basis of our motivation to investigate the important research question if unsupervised approaches could leverage this diverse data source to overcome drawbacks of supervised learning. Presumably, such a model is free of any rendering pipeline induced biases and enables higher-quality results due to generalization to arbitrary measurement conditions.

Recent unsupervised approaches for general optical flow problems \cite{zhang2020, zhong2019, lagemann2021a, Jonaschkowski2020, jiang2021, liu2019b, Meister2017, Zhu2017} mostly share the basic assumption that an object's appearance
does not change within two consecutive frames. Hence, it allows us to estimate the optical flow by warping one image to match the other. However, further loss entities accompanying this constant feature assumption to regularize the network vary greatly, but largely determine the effectiveness of the unsupervised learner in the end. To shed some light on this methodological zoo of penalty terms, we investigate systematically key components of common photometric losses and accommodating regularizers to derive a proxy loss well-suited for the specific test case of PIV applications. Our findings highlight that a loss objective comprising the generalized Charbonnier loss to penalize photometric differences between source and warped target image, an adaptive loss objective to address forward/backward flow inconsistencies, and an adaptive second-order smoothness objective is very robust and convincingly accurate in real-world applications. That is, we demonstrate that an unsupervised optical flow model trained under this proxy loss can reliably estimate optical flow for different synthetic and real-world applications. More importantly, this is achieved at a spatial resolution which goes far beyond traditional cross-correlation based counterparts.

The remainder of the paper is organized as follows. We first introduce basic conceptual ideas and assumptions and the applied architecture. We then present an extensive body of empirical work, spanning many different, challenging PIV experiments. Finally, we close with some concluding comments.

2 Deep learning architecture for estimating optical flow in particle image pairs

RAFT-PIV is a deep learning-based method for estimating the optical flow of PIV images (Lagemann et al. (2021b, 2022)). It is based on the successful RAFT architecture by (Teed and Deng (2020), operates entirely on the input resolution, and updates its flow predictions iteratively. RAFT-PIV has been shown to achieve state-of-the-art accuracy on a public PIV database and outperforms other supervised and unsupervised learning-based approaches. Moreover, it generalizes well to various real-world measurements.

As shown in Fig. 1, the working principle of RAFT-PIV involves two main steps. First, a feature extraction is performed using a shared CNN-based feature encoder which maps the input image pair into a dense feature embedding. This high-level feature map extracts the most important image patterns such as edges and textures. In addition to the feature encoder, RAFT-PIV employs a context encoder that is only applied to the first particle image.

In the second stage of RAFT-PIV’s optical flow estimation process, a 4D correlation volume is calculated between the feature embeddings of both input images. This is done using a dot product to measure similarity. The correlation provides insight into the probability that a particular pixel in the first frame resembles pixels in the second frame. A 4-layer correlation pyramid is then created by pooling the last two dimensions of the correlation volume and stacking the initial and subsequent volumes. This approach allows RAFT-PIV to accurately track large movements of small objects across multiple scales.

Finally, the optical flow is updated using a convolutional gated recurrent unit (ConvGRU), which takes into account some motion features, context information, and its current hidden state. The motion features are derived from a combination of context features and local extracts of the correlation volume, which are updated iteratively. Hence, the final flow prediction is a combination of a sequence of residual flow updates during which the ConvGRU balances knowledge of previous optical flow predictions with its current hidden state. Further architectural details and a more intuitive description of specific network parts can be found in Lagemann et al. (2021b, 2022).

In contrast to the original RAFT-PIV network which was trained in a supervised fashion, the current work explores new unsupervised loss formulations to train the optical flow network directly on real-world PIV images. Unsupervised learning paradigms possess a significant advantage as the loss objective is predicated on geometric penalty terms, obviating the need for ground truth labels. However, the formulation of an appropriate unsupervised loss objective is a non-trivial task and not as straightforward as in supervised learning. To date, typical unsupervised loss objectives for general optical flow estimation comprise photometric, forward/backward flow consistency, and discretized smoothness penalty terms. Thus, unsupervised loss functions are themselves compositions of simpler loss entities.

In this work, three different photometric losses are investigated to penalize the photometric difference between the source and the warped target image, namely the generalized Charbonnier loss, the structural similarity index measure (SSIM), and the ternary census loss. The Charbonnier loss builds upon a half-quadratic loss regularization which ensures edge-preservation in ill-posed image processing problems. By introducing a hyperparameter $\varepsilon$, the Charbonnier loss is of linear order for absolute values smaller than 1.
and of quadratic order elsewhere. To further improve the robustness of this loss function, a shape parameter $\alpha$ can be used in the exponent. The generalized version of the Charbonnier loss reads $L_{\text{Charb}} = (a^2 + e^2)^{\alpha/2}$ with $a$ denoting an arbitrary tensor.

The structural similarity index measure (SSIM) is used as a metric to measure the similarity between two images and extracts three key features from an image: luminance, contrast, and structure. The luminance $\mu$ is the mean of all pixel values whereas the contrast $\sigma$ is expressed by the standard deviation of the pixels. Structural information is computed by subtracting the luminance and dividing by the contrast. In combination, the SSIM function reads $L_{\text{SSIM}} = \frac{(2\mu_1\mu_2 + \epsilon_1)(2\sigma_1\sigma_2 + \epsilon_2)}{[\mu_1^2 + \mu_2^2 + \epsilon_1][\sigma_1^2 + \sigma_2^2 + \epsilon_2]}$, where $\mu_1$ and $\mu_2$ denote the luminance and $\sigma_1$, $\sigma_2$, and $\sigma_1,2$ the contrast of the source and the target image. The quantities $e_1$ and $e_2$ are used to avoid division by zero.

The ternary census transform is a non-parametric matching cost that relies on the ordering of pixel intensities and provides an illumination-robust constancy assumption for solving correspondence problems in computer vision. It is based on the census transform, which computes a binary string (census signature) for every pixel by comparing its grey value with the grey values in its neighborhood. The census transform can be expressed as $x_i(i, i^*) = \begin{cases} 0 & \text{if } i > i^* \\ 1 & \text{if } i \leq i^* \end{cases}$, where $i$ denotes a specific pixel and $i^*$ the values of its eight neighboring pixels. The similarity between two images can be measured by computing the Hamming distance between both census signatures and averaging it to obtain the final census loss.

During training, photometric penalty terms are applied bi-directionally meaning that optical flow fields are predicted for forward and backward flow individually by switching the input image order.

Moreover, a consistency check between forward and backward flow is performed to further penalize occluded regions. Two formulations are considered: the generalized Charbonnier loss and an adaptive
robust loss function. The latter is a combination of various losses and can be written as

\[
L = \frac{\|\alpha - 2\|}{\alpha} \left( \left( \frac{(a/s)^2}{\|\alpha - 2\| + 1} \right)^{(a/2)} - 1 \right),
\]

where \(\alpha\) is a parameter to control the robustness of the loss and \(s > 0\) represents a scale parameter. This loss function has increased expressiveness and the effect of outliers can be optimized based on empirical data.

Finally, an optical flow regularization is applied to the forward and backward flow to obtain smoothness in the predicted displacement fields. Here, we use an edge-aware smoothness penalty which formally can be expressed as the \(k\)-th order accurate smoothness loss

\[
L_{\text{smooth}} = \sum \exp \left(-\left| \frac{\partial I}{\partial x} \right| \right) \left| \frac{\partial^k V}{\partial x^k} \right| + \exp \left(-\left| \frac{\partial I}{\partial y} \right| \right) \left| \frac{\partial^k V}{\partial y^k} \right|
\]

where \(I\) denotes the gray color image intensity and \(V\) is the optical flow prediction. In summary, the applied loss objective for unsupervised learning can be written as

\[
L = \omega_{\text{photo}} \cdot L_{\text{photo}} + \omega_{\text{con}} \cdot L_{\text{con}} + \omega_{\text{smooth}} \cdot L_{\text{smooth}},
\]

where \(L_{\text{photo}}, L_{\text{con}}, \) and \(L_{\text{smooth}}\) denote the penalty function of the photometric, forward-backward consistent and smoothness oriented objective. Their individual weight \(\omega\) is obtained by extensive hyperparameter studies. We trained the RAFT-PIV model for 200 epochs on the training dataset introduced in Lagemann et al. (2021b) and used the same hyperparameters as detailed in Lagemann et al. (2022).

3 Results

To investigate the performance of our unsupervised deep learning based optical flow tool, three benchmark are investigated. First, we evaluate a general benchmark scenario characterized by idealized particle image conditions. Next, a test case presented in the third international PIV challenge is investigated to elaborate on the spatial resolution ability of the proposed unsupervised network. Finally, images of a real PIV measurement are analyzed. To enhance readability and concision of this paper, only results of the most successful loss combinations are presented.

3.1 Idealized benchmark

This dataset Cai et al. (2019) is characterized by a high particle image density, a maximum particle image displacement of \(\pm 10\)px and a particle image peak intensity ranging from 200 to 255 counts within an 8-bit grayscale. These images represent mostly high-quality experimental conditions which are hard to meet in real-world PIV measurements. Hence, this dataset is valuable and useful in providing a general benchmark for algorithm evaluation, but does not allow for profound conclusions with respect to arbitrary experimental measurements. The endpoint errors of the inference results are presented in Table 1 along with a comparison to state-of-the-art techniques described in literature. The loss combinations examined in this study demonstrate a considerable improvement over an unsupervised LiteFlowNet-based approach, resulting in a substantial reduction in the endpoint error. Interestingly, unsupervised RAFT-PIV networks also exhibit strong performance when compared to their supervised equivalents. Although the error values are somewhat higher due to the less informative nature of the unsupervised proxy loss, the overall performance remains noteworthy. Our investigation further identified three primary sources of increased error: occluded image domains at image boundaries, regions lacking texture, and local smoothing of extrema. Occluded image domains at image boundaries present a challenge for accurate motion prediction due to the failure of the brightness constancy assumption. An appropriate smoothness objective can partially compensate for this issue, but it becomes increasingly difficult to address as displacement values increase. Texture-less regions, characterized by the absence of particle images, also pose a challenge for motion prediction since no flow information is provided. The smooth flow field assumption can provide some correction, but larger particle image-free domains may result in spurious motion prediction. Utilizing high seeding densities can enhance the performance of unsupervised PIV networks in these scenarios. However, in regions
Table 1: Averaged Endpoint Error (AEE) for all test cases of the idealized PIV database introduced in Cai et al. (2019) for various unsupervised RAFT-PIV models. Test cases include flows around a cylinder and over a back-step, simulations of a turbulent channel flow and isotropic turbulence, and velocities fields based on the Surface Quasi-Geostrophic (SQG) model. The error unit is set to pixel per 100 pixels for easier comparison. Values in curly brackets correspond to supervised networks. Except for the “SSIM-1st” model, all other RAFT-PIV networks are optimized by an adaptive forward/backward consistency penalty, an adaptive self-supervision loss and a 2nd-order accurate smoothness regularization, but apply different photometric objectives. The "SSIM-1st" model employs a proxy loss based on the SSIM loss, an adaptive forward/backward flow penalty, an adaptive self-supervision loss, and an adaptive first-order accurate smoothness term.

with local flow extrema and vanishing gradients, smoothing can cause complementary errors. In particular, between regions with larger displacement gradients, peak values of high and low momentum streaks are underestimated and overestimated respectively, leading to increased error values. As a result, flow fields dominated by gradients, such as channel flows or isotropic turbulence, have higher endpoint errors as shown in Table 1. In addition, as shown in Table 1 using a first-order accurate smoothing term, denoted as “SSIM-1st”, results in a greater endpoint error than using its second-order counterpart, “SSIM-2nd”. This can be visually confirmed by looking at Figure 2. The reason for this difference is that first-order accurate smoothing objectives only apply filters in the horizontal and vertical directions. On the other hand, second-order accurate smoothing penalties include diagonal gradients, making them more effective in handling turbulence dominated flows such as isotropic turbulence and SQG test cases. Overall, a loss objective comprising the

Figure 2: Inference results - test case isotropic turbulence: Optical flow prediction using different loss combinations and the absolute error distribution between ground truth flow and network predictions. The first row illustrates the displacement and the second row the error distribution of the horizontal direction.
generalized Charbonnier loss to penalize photometric differences between source and warped target image, an adaptive loss objective to address forward/backward flow inconsistencies, and an adaptive second-order smoothness objective is the most robust and convincingly accurate loss combination in this test case. For reasons of brevity, only results of this specific proxy loss will be investigated and compared against gold-standard baselines in the following.

3.2 Spatial resolution ability

In order to rigorously evaluate the spatial resolution capabilities of URAFT-PIV, we subject it to a series of open-source PIV test cases specifically designed to elucidate the limitations of standard PIV algorithms when confronted with small spatial scales and strong displacement gradients. We utilize test case A4 from the third international PIV challenge Stanislas et al. (2008) and compare the results with those obtained from a state-of-the-art PIV algorithm. The selected test case entails a modulated one-dimensional sinusoidal shear displacement characterized by a wavelength that oscillates between 10 and 400 px and a sine amplitude of 2 px. The results depicted in Fig. 3 incorporate third-order accurate B-spline interpolated results for pixelwise prediction derived from cross-correlation based in-house code called PascalPIV Marquardt et al. (2019); Feldhusen-Hoffmann et al. (2021). As demonstrated in Fig. 3 and supported by Stanislas et al. (2008), a standard cross-correlation approach has difficulties resolving the lowest wavelengths. In contrast, an unsupervised RAFT-PIV model can accurately estimate local displacements even at very fine spatial scales and low wavelengths, which is a significant advantage of this neural optical flow method. However, one can also note that the displacement magnitude decreases with decreasing spatial wavelength due to the smoothness regularization applied during unsupervised training. This causes peak displacements in areas with small sine wavelengths to be smoothed and underestimated.

3.3 PIV measurement of a turbulent jet flow

The final test case explores the effectiveness of an unsupervised RAFT-PIV network when applied to a high-speed PIV measurement of a turbulent jet flow. The experimental setup details can be found in Stanislas et al. (2008). The PIV images present a challenge due to the relatively low density of particle images and their comparably small diameters. Results from PascalPIV and both supervised and unsupervised RAFT-PIV methods are displayed in Fig. 4, where only the displacement magnitude is shown as the radial displacement is minimal and does not provide valuable information. From a broader perspective, Fig. 4 effectively demonstrates the resolution advantage of both supervised and unsupervised RAFT-PIV approaches, as several flow details detected by (U)RAFT-PIV are lost when PascalPIV is used. Additionally, Fig. 4 visually reveals the benefit of fine-tuning a network on the specific test case. An unsupervised optical flow estimator trained solely on a synthetic dataset struggles with structure-less regions at the left and right image boundaries, likely because such particle-free image conditions are not captured in the training dataset. However, fine-tuning on the test images significantly improves the prediction accuracy and enables the network to predict correct displacements close to the image boundaries.
Figure 4: Experimental data - turbulent jet flow: The displacement magnitude of the benchmark PascalPIV, RAFT-PIV, and a raw and fine-tuned URAFT-PIV model (see text) is shown. Fine-tuning URAFT-PIV can substantially enhance the performance of the unsupervised learning method.

4 Conclusion

The majority of machine learning based optical flow methods for Particle Image Velocimetry (PIV) measurements are trained on synthetic data in a supervised manner resulting in a discrepancy between training and testing. Unsupervised training, which only requires unlabeled data, can help overcome this issue by eliminating biases induced by the rendering pipeline. We explored the performance of unsupervised optical flow estimation for PIV applications by examining key components of photometric losses and regularizers. We considered various loss entities such as the generalized Charbonnier loss, the ternary loss based on a census transform, and the Structural Similarity Index (SSIM) loss to address photometric differences. Forward/backward flow consistency was penalized using either the generalized Charbonnier loss or the adaptive robust objective. Smoothness penalties included first- and second-order accurate formulations as well as penalties based on adaptive robust loss functions. We conducted large-scale hyperparameter studies to investigate each combination of loss objectives and identified appropriate objective weights. The trained networks were tested on PIV data of varying complexity, including a general benchmark scenario with idealized particle image conditions, a test case to evaluate the spatial resolution ability of the proposed unsupervised networks, and results from real PIV measurements. Overall, it is shown that unsupervised training is indeed a valuable alternative for supervised learning, especially for test cases where no synthetic data is available during training. URAFT-PIV demonstrates a convincing performance relative to both, synthetic benchmarks and real-world measurement data, and enhances the generalization ability of future RAFT-PIV based learners.
Acknowledgements

This research was funded by the Deutsche Forschungsgemeinschaft within the research project “Learning Deep Optical Flow Estimation for Particle-Image Velocimetry” (DFG SCHR 309/79). The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. for funding this project by providing computing time on the GCS Supercomputers.

References


Lagemann C, Klaas M, and Schröder W (2021a) Unsupervised recurrent all-pairs field transforms for particle image velocimetry. in *14th International Symposium on Particle Image Velocimetry*. volume 1


Teed Z and Deng J (2020) RAFT: Recurrent All-Pairs Field Transforms for Optical Flow. in *European Conference on Computer Vision*. pages 402–419. Springer
