



Cite this: *Soft Matter*, 2024, 20, 7246

Deep-learning optical flow for measuring velocity fields from experimental data†‡

Phu N. Tran, ^a Sattvic Ray, ^c Linnea Lemma, ^{ac} Yunrui Li, ^b Reef Sweeney, ^c Aparna Baskaran, ^a Zvonimir Dogic, ^{acd} Pengyu Hong ^{*b} and Michael F. Hagan ^{*a}

Deep learning-based optical flow (DLOF) extracts features in adjacent video frames with deep convolutional neural networks. It uses those features to estimate the inter-frame motions of objects. We evaluate the ability of optical flow to quantify the spontaneous flows of microtubule (MT)-based active nematics under different labeling conditions, and compare its performance to particle image velocimetry (PIV). We obtain flow velocity ground truths either by performing semi-automated particle tracking on samples with sparsely labeled filaments, or from passive tracer beads. DLOF produces more accurate velocity fields than PIV for densely labeled samples. PIV cannot reliably distinguish contrast variations at high densities, particularly along the nematic director. DLOF overcomes this limitation. For sparsely labeled samples, DLOF and PIV produce comparable results, but DLOF gives higher-resolution fields. Our work establishes DLOF as a versatile tool for measuring fluid flows in a broad class of active, soft, and biophysical systems.

Received 24th April 2024,
Accepted 20th August 2024

DOI: 10.1039/d4sm00483c

rsc.li/soft-matter-journal

1 Introduction

Accurate measurement of flow fields is a cornerstone for modeling diverse phenomena ranging from fluid dynamics¹ and active matter² to biology.³ A conventional approach to estimating flow fields is particle image velocimetry (PIV), where flow velocities are computed by correlating features of two consequent images.^{4–6} However, PIV has limitations. One arises from the dependence of the interrogation window size on seeding particle speed. Consequently, PIV cannot estimate turbulent flows smaller than the interrogation window, leading to potential errors in the velocity field.⁷ Furthermore, significant Brownian motion can introduce uncertainty into PIV measurements.⁸ Another limitation is that tracer particles must be within an optimal range of density and size.⁷ This requirement can be impractical in biological systems using fluorescent proteins as markers, preventing the use of smaller window sizes

as a workaround for issues related to Brownian motion or smaller turbulent flows.⁹ To overcome these limitations we explore a deep learning-based optical flow (DLOF) algorithm for the estimation of the flow fields.

In computer vision, optical flow describes the apparent motions of objects in a sequence of images.¹⁰ Various rule-based techniques for optical flow estimation have been developed, including differential methods,^{11–14} variational methods,^{15–18} and feature-based methods.^{19–22} Implementations of rule-based optical flow algorithms can be advantageous over PIV for applications in biological images.^{23–28} Rapid advancements in machine learning resulted in deep learning optical flow (DLOF) algorithms, where the automatic feature extraction offered by deep convolutional neural networks has significantly improved the algorithm accuracy.^{29–55}

Although recent efforts used DLOF to estimate velocity fields in applications that would otherwise rely on PIV,^{56–61} these works trained and evaluated DLOF with synthetic data from fluid dynamics simulations or computer-generated and augmented PIV datasets that mimic noisy data in real-world experiments. Obtaining ground-truth velocities required for training machine learning models has been challenging with real-world data. We overcome this limitation by investigating the performance of DLOF on experimental data from extensively studied active nematic liquid crystals.^{62–72} We image microtubule (MT)-based active nematics under conditions that are beyond the limitations of PIV and present a significant challenge to its performance. We then develop a computational

^a Department of Physics, Brandeis University, Waltham, MA 02453, USA.

E-mail: hagan@brandeis.edu

^b Department of Computer Science, Brandeis University, Waltham, MA 02453, USA.

E-mail: hongpeng@brandeis.edu

^c Department of Physics, University of California at Santa Barbara, Santa Barbara, CA 93106, USA

^d Biomolecular and Engineering Science, University of California at Santa Barbara, Santa Barbara, CA 93106, USA

† Data and code available: <https://github.com/trannogocphu/opticalflow-activenematics>.

‡ Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d4sm00483c>



framework to apply DLOF to quantify the microtubule velocity fields. We test the framework with ground truth velocity fields obtained by particle tracking methods. We compare the velocity fields obtained by PIV and DLOF against this ground-truth data. Importantly, this data is characteristic of flow fields from diverse soft matter and biophysical systems, suggesting that our conclusions are broadly applicable.

Microtubule (MT)-based active nematics are powered by ATP-consuming kinesin molecular motors. In such materials the extensile MT bundles generate internal active stresses, which in turn give rise to motile topological defects and associated autonomous flows.⁶⁴ Active nematics are described by two continuous fields, the director field, which describes the average orientation of the anisotropic MT filaments, and the velocity field, which describes their motions. Accurate measurement of the director field requires samples in which all the filaments are labeled. However, such samples yield low variations in spatial intensity, which makes application of PIV techniques challenging.⁷³ In fully labeled active nematics PIV underestimates the velocity component along the nematic director,^{68,74–76} which can be attributed to the nematic anisotropy; the intensity of MT bundles is fairly uniform along the nematic directors, which presents challenges for implementation of PIV. Alternatively, obtaining accurate PIV fields requires samples with a low volume fraction of labeled MTs, which creates highly speckled patterns suitable for PIV application, but from which the director field cannot be extracted. Overcoming these competing challenges requires active nematics containing high-concentration MTs in one color and dilute tracer MTs in a different wavelength.⁷⁶ The former are suitable for director field measurement while the latter allow for accurate application of PIV techniques. However, these samples are cumbersome to prepare, and sequential imaging can introduce a time lag between the measurement of the two fields.

We show that DLOF produces an accurate measurement of the flow field irrespective of the fraction of labeled filaments. Thus, DLOF techniques can fully characterize the instantaneous state of an active nematic from one set of images. Furthermore, the DLOF results are higher resolution and less noisy than those from PIV. Importantly, while we use the MT-based active nematic system to test optical flow, the implications of our results are more general. The velocity fields from the MT-based active nematic system closely resemble the chaotic flow fields that arise in different soft matter and biophysical systems, such as unstable elastic polymer solutions,⁷⁷ bacterial suspensions,⁷⁸ interface dynamics of confined active droplets,⁷⁹ tissues dynamics driving biological morphogenesis,^{80,81} and flows generated by biological swimmers.⁸² Thus, our results suggest that DLOF models can be used for more accurate and robust measurements of the velocity fields across this wide range of active, soft, and biological systems.

2 Deep learning optical flow (DLOF)

DLOF uses convolutional neural networks for the automatic extraction of relevant features from two adjacent frames in a

video and uses the extracted features to estimate the movements of objects between the two video frames.^{29,32,38,83,84} DLOF models are typically trained using supervised learning algorithms, in which training data are synthetic videos that include the true motions of all the objects in the videos across the video frames.^{85–91} Synthetic data are required by this approach because obtaining the true displacements of objects in real-world videos is highly challenging. Thus, the ability of the models to properly adapt to unseen data from a different domain becomes crucial for the trained models to be useful in real-world scenarios. A recent study suggested that a model called RAFT (Recurrent all-pairs field transforms for optical flow), which was originally trained using synthetic data, could generalize well to unseen fluid dynamics videos.^{58,84} However, this study evaluated the model's performance on simulation-generated videos and did not evaluate the performance on challenging videos obtained in experiments, such as the active nematics described above.

2.1 Architecture of the RAFT model

RAFT estimates the optical flow from a pair of images (I_1, I_2) in three main stages: (1) extract features of the input images using a convolutional neural network, (2) use those extracted features to construct a correlation volume that computes the visual similarity of the images, and (3) compute the final flow through an iterative process.

2.1.1 Feature map extraction. The model uses an encoder g_θ , which is a convolutional neural network, to extract features from the two input images. In particular, g_θ extracts features at 1/8 resolution; *i.e.*, $g_\theta: \mathbb{R}^{H_0 \times W_0 \times C} \mapsto \mathbb{R}^{H_0/8 \times W_0/8 \times D}$, where H_0 and W_0 are the height and width of the images, C the number of color channels ($C = 3$ for RGB and $C = 1$ for grayscale images), and D the number of desired feature maps to be extracted. The encoder gradually reduces the resolution of the output feature maps; *i.e.*, it successively outputs feature maps at 1/2, 1/4, and finally 1/8 resolution. For each of these steps, the resolution reduction is performed by convolutional residual neural network blocks (Fig. 2A). In general, feature maps produced at lower resolutions extract spatial correlations at higher levels with a wider receptive field, and it has been shown empirically that learning features at the aforementioned resolutions offers a balance between the model's performance and complexity.⁸⁴

2.1.2 Construction of correlation feature map. Visual similarity between the two input frames is required to find the correspondences of moving objects between them. RAFT computes the visual similarity by constructing a correlation between all pairs of extracted features of first the image $g_\theta(I_1) \in \mathbb{R}^{H \times W \times D}$, and then that of the second image $g_\theta(I_2) \in \mathbb{R}^{H \times W \times D}$ (right part of Fig. 2A). The elements of a correlation volume $C(g_\theta(I_1), g_\theta(I_2)) \in \mathbb{R}^{H \times W \times H \times W}$ are given by $C_{ijkl} = \sum_h g_\theta(I_1)_{ijh} \cdot g_\theta(I_2)_{klh}$. Correlations are further computed as a 4-layer pyramid $\{C^1, C^2, C^3, C^4\}$, where C^k has dimensions $H \times W \times H/2^k \times W/2^k$ (Fig. 2B). Here, the reduction of the last two dimensions of the correlation volume C by a factor of 2^k is achieved by pooling the last two dimensions of C with kernel



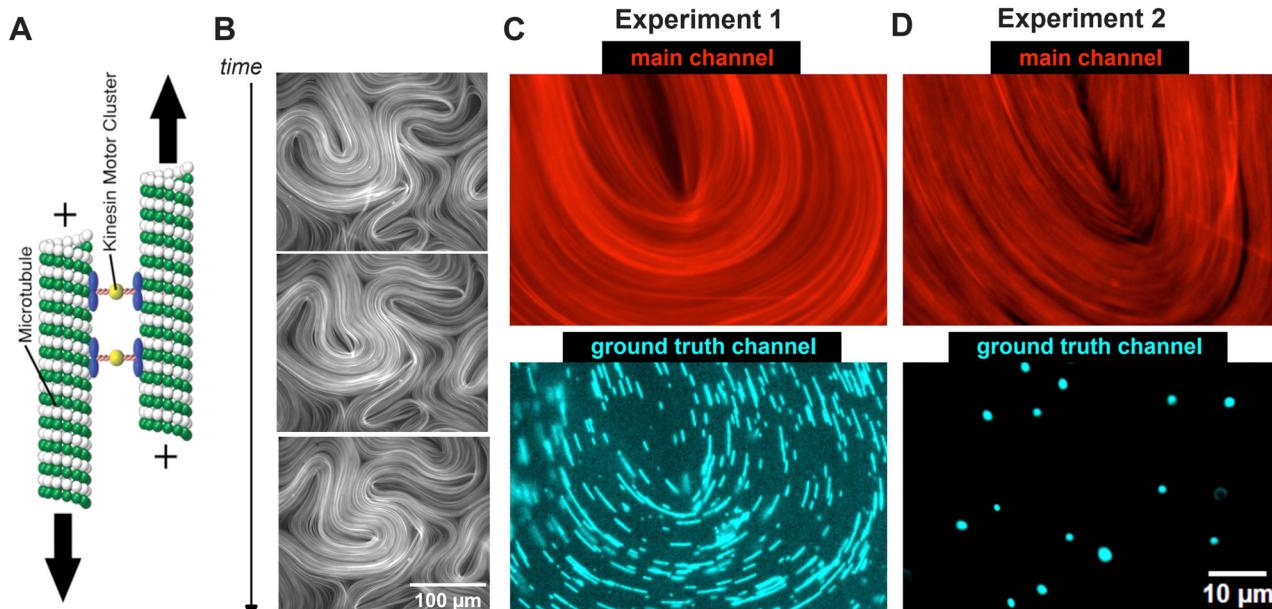


Fig. 1 Microtubule (MT)-based active nematics. (A) Microscopic components of the active nematic liquid crystal. Kinesin motor clusters consume energy to actively slide neighboring MTs against each other. (B) The active nematic exhibits the spontaneous flow that deforms the nematic texture over time. All MTs are fluorescently labeled at 647 nm. Increased local intensity indicates a higher local filament concentration. The time step is 7.5 s. (C) In Experiment 1, the fully labeled MTs (top panel) are mixed with a sparse population of MTs that fluoresce at 488 nm (bottom panel), which are used to generate ground-truth velocity points. (D) In Experiment 2, the fully labeled MTs (top panel) are mixed with passivated microbeads, which are used to generate the ground-truth velocities (bottom panel).

size k and equivalent stride. Having correlations at multiple levels through $\{C^1, C^2, C^3, C^4\}$ allows the model to handle both small and large displacements. The first two dimensions (that belong to I_1) are maintained to preserve high-resolution information, enabling the model to detect motions of small fast-moving objects.

The link between an object in I_1 and its estimated correspondence in I_2 is determined through correlation lookup using the correlation pyramid, as described in Fig. 2C. The correspondence $x' \in I_2$ of a pixel $x = (u, v) \in I_1$ is estimated by $x' = (u + f^1(u), v + f^2(v))$, where (f^1, f^2) is the current estimate of DLOF between I_1 and I_2 . A local grid around x' is then defined as $\mathcal{N}(x')_r = \{x' + \Delta x | \Delta x \in \mathbb{Z}, \|\Delta x\|_1 \leq r\}$, a set of integer offsets that are within a radius of r of x' (using ℓ_1 distance). The local neighborhood $\mathcal{N}(x')_r$ is used to index from all levels of the correlation pyramid using bilinear sampling, such that the grid $\mathcal{N}(x'/2^k)_r$ is used to index the correlations C^k . At a constant searching radius r across all levels, a local neighborhood on a lower level implies a larger context; for example, at $k = 4$, a neighborhood of $r = 4$ effectively includes a range of 256 pixels at the video's resolution. The interpolated correlation scores at all levels are concatenated to form a single feature map, which serves as an input for iterative flow refinement described below.

2.1.3 Iterative flow refinement. The flow between the two input images is determined through an iterative process, such that the final flow \mathbf{f}_N is obtained from the sequence $\mathbf{f}_{k+1} = \mathbf{f}_k + \Delta \mathbf{f}$ where $0 \leq k \leq N - 1$, $\mathbf{f}_0 = \mathbf{0}$, N is the number of iterations, and $\Delta \mathbf{f}$ is being produced by the model at each of the iterations.

The flow updating is performed by a convolutional gated recurrent unit (ConvGRU) cell,⁹² in which convolutions have replaced fully connected layers:

$$z_t = \sigma(\text{Conv}_{3 \times 3}([h_{t-1}, x_t], W_z)) \quad (1)$$

$$r_t = \sigma(\text{Conv}_{3 \times 3}([h_{t-1}, x_t], W_r)) \quad (2)$$

$$\tilde{h}_t = \tanh(\text{Conv}_{3 \times 3}([r_t \odot h_{t-1}, x_t], W_h)) \quad (3)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (4)$$

where x_t , z_t , r_t , \tilde{h}_t , h_t are the input, update gate, reset gate, internal memory state, and hidden state at time t , respectively; $\sigma(\cdot)$ is the sigmoid function, $\tanh(\cdot)$ the hyperbolic tangent, and $\text{Conv}_{3 \times 3}(\cdot, W)$ the convolution operator with kernel size 3×3 and bias W . Here, the hidden state h_t is further processed by two convolutions to produce the flow update $\Delta \mathbf{f}$ at time t .

In the above set of equations, at a current time t , the input x_t is the concatenation of the current flow estimate, correlation, and context features. The update gate z_t , which is calculated using the last hidden state h_{t-1} and the current input x_t , controls how much past knowledge should be considered in the computation of the current hidden state h_t . The reset signal r_t is a function of the current input x_t and the last hidden state h_{t-1} , and determines how much of the past knowledge to forget. The internal memory \tilde{h}_t of the GRU cell is calculated using the current input x_t and the last hidden state h_{t-1} weighted by the reset gate r_t . Finally, the hidden state is updated by the weighted sum of the last hidden state h_{t-1}



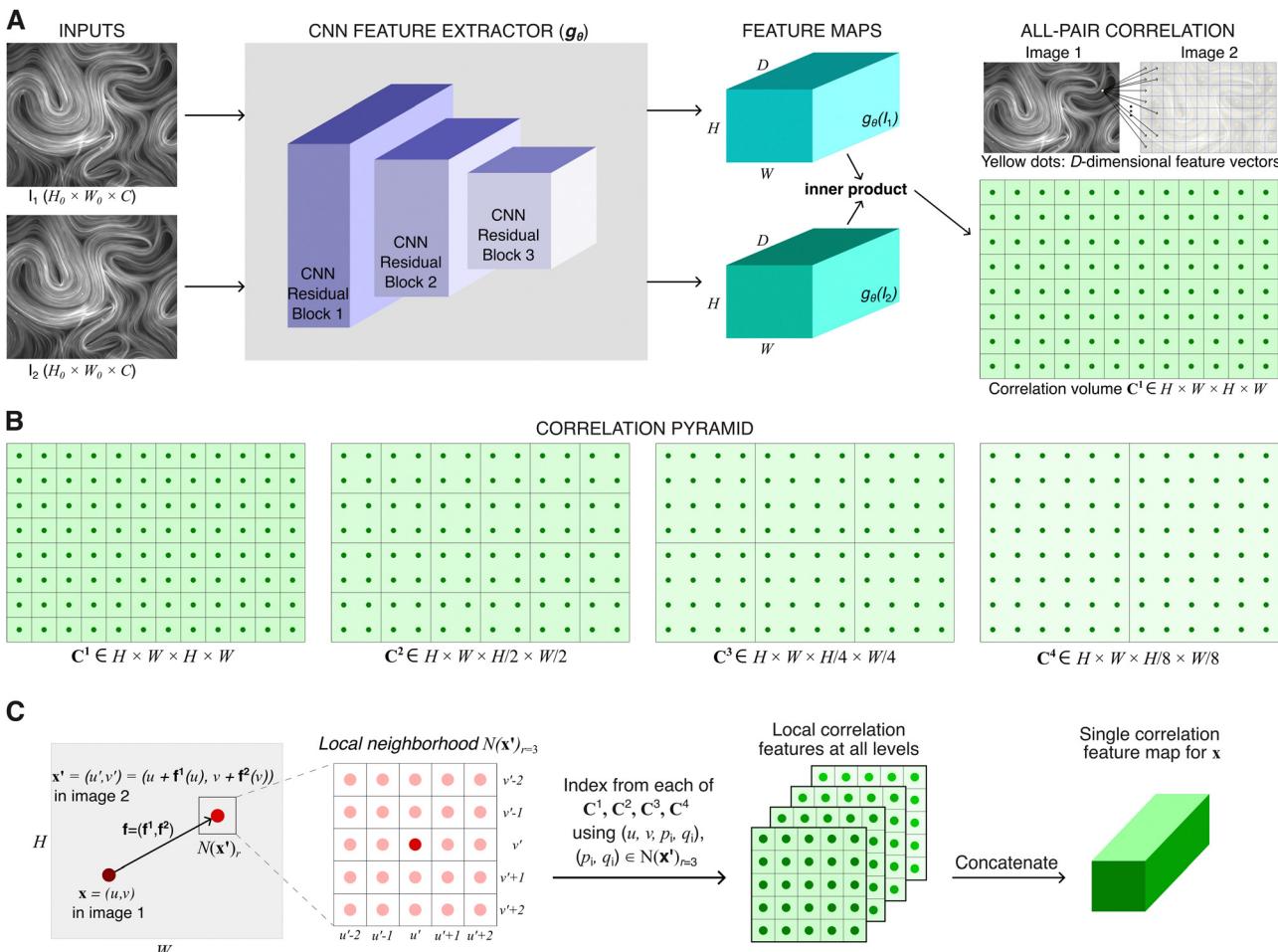


Fig. 2 Main components of the DLOF model. (A) Feature extraction and construction of feature-level correlations: a convolutional neural network (CNN) is used to extract D feature maps of resolution $H \times W$ for each of the input images. Taking the inner product of the features maps of two images produces all-pair feature-level correlation volumes \mathbf{C}^1 of dimension $H \times W \times H \times W$. (B) Correlation pyramid: multi-scale feature correlations are constructed by pooling the last two dimensions of \mathbf{C}^1 , such that those dimensions are reduced by 1/2, 1/4, and 1/8, resulting in \mathbf{C}^2 , \mathbf{C}^3 , and \mathbf{C}^4 , respectively. The first two dimensions preserve high-resolution information while multi-scale correlations enable the model to capture the motions of small fast-moving objects. (C) Correlation lookup for a pixel \mathbf{x} in I_1 : an estimate of the location of the correspondence \mathbf{x}' in I_2 is initialized by displacing \mathbf{x} using the current flow estimate \mathbf{f} . The model then looks for the most correlated features in a neighborhood $N(\mathbf{x}')_r$, centered at \mathbf{x}' ($r = 3$ in the figure), where all locations within $N(\mathbf{x}')_r$ are used to index from the correlation pyramid $\{\mathbf{C}^1, \mathbf{C}^2, \mathbf{C}^3, \mathbf{C}^4\}$ to produce correlation features at all levels, which are further concatenated to form a single correlation feature map for the pixel \mathbf{x} in I_1 .

and the current cell memory \tilde{h}_t , with the update gate z_t controlling the weights distribution.

2.2 Training DLOF

Most DLOF models are trained by supervised learning using synthetic data, where flow ground truths can be obtained straightforwardly during data generation. The supervised loss \mathcal{L}_s used to optimize RAFT's parameters compares the sequence of predictions $\{\mathbf{f}_1, \dots, \mathbf{f}_N\}$ with the flow ground truth \mathbf{f}_{gt} , with exponentially increasing weights:

$$\mathcal{L}_s = \sum_{i=1}^N \gamma^{N-1} \|\mathbf{f}_{gt} - \mathbf{f}_i\|_1 \quad (5)$$

where $\gamma < 1$. RAFT is trained using supervised learning, and it has been shown to generalize well to data in other domains.^{58,84} When it is required, the model's parameters can be further

fine-tuned using the real-world data in the target domain; however, unsupervised learning is generally required because ground truths of those data are often unavailable.

2.2.1 Unsupervised training. An approach to unsupervised training is to generate realistic pseudo-flow ground truth data using the current model, and then use that pseudo ground truth data for further optimizing the model's parameters. In this approach, the current model is first used to warp the image I_1 to produce an estimate of the image I_2 , i.e., $\tilde{I}_2 = \Omega(I_1, \mathbf{f})$ where Ω is the warping function that displaces the pixels in I_1 according to the current estimate \mathbf{f} of the flow. \tilde{I}_2 can be then used as a pseudo ground truth to compute a simple unsupervised loss

$$\mathcal{L}_u = w_{photo} \cdot \mathcal{L}_{photo} + w_{smooth} \cdot \mathcal{L}_{smooth} \quad (6)$$

where \mathcal{L}_{photo} denotes the photometric loss between I_2 and \tilde{I}_2 , \mathcal{L}_{smooth} flow smoothness regularization, and w_{photo} , w_{smooth}

are the weights. The photometric loss quantifies the structural and visual differences between I_2 and \tilde{I}_2 , being aware of occluded regions in which pixels in I_1 do not have their correspondences in I_2 . A common metric used for photometric loss is the occlusion-aware structural similarity index (SSIM).^{25,93} A major challenge in unsupervised training of DLOF models is to obtain an accurate estimate of occlusions,⁸³ which cannot be directly measured when dealing with real-world data. The unsupervised loss above also has a second term to encourage the smoothness of the resultant velocity fields. For example, the k -th order smoothness is defined as⁸³

$$\mathcal{L}_{\text{smooth}(k)} = \frac{1}{k} \sum \exp\left(-\frac{\nabla I}{\sigma}\right) \cdot \|\nabla^{(k)} V\| \quad (7)$$

where ∇I detects the edges in the current image, $\nabla^{(k)} V$ is the k -th order gradient of the corresponding velocity field, σ controls the strength of the regularization, and n is total number of samples.

We obtained the results in the benchmarks of this work using a RAFT model that was trained with the FlyingThings synthetic datasets,⁹¹ which yielded the highest performance in our investigation. During velocity computation, we empirically set the number of iterations for flow refinement to 24.

3 Active nematics samples

We tested the performance of the DLOF framework using a MT-based active nematic liquid crystal.⁶⁴ An active nematic is a quasi-2D liquid crystal comprised of locally aligned filamentous MTs. When powered by kinesin molecular motors, extensible MTs spontaneously generate a chaotic flow field that varies over space and time, and in turn reorients the nematic texture (Fig. 1A and B). Typically, the velocity field is computed by performing PIV on images of active nematics comprised of fluorescently labeled MTs. However, this method can be inaccurate when all the MTs are labeled, as these samples have poor contrast variations in fluorescence intensity, especially in the direction of the MT alignment.⁷³

We performed two distinct experiments, each containing a different type of tracer that we used to estimate the ground truth. In both experiments, a large fraction of MTs were labeled with a fluorescent dye that emits 647 nm wavelength photons. In Experiment 1, samples contained a very low concentration of 488 nm labeled MTs. They were dilute enough so that individual filaments could be distinguished (Fig. 1C, bottom panel). However, accurately linking the detected MTs into time trajectories was only possible for a small fraction of the dilute population. In Experiment 2, instead of relying on dilute labeling, we mixed passivated 488 nm fluorescent microbeads into the active nematic (Fig. 1D, bottom panel). Although not directly incorporated into the quasi-2D active nematic, these beads were located right above the nematic layer and followed the same flow field. Compared to the sparsely-labeled MTs, the beads could be reliably tracked across several frames with

an automated algorithm [ESI‡], thus providing a larger set of velocity values that served as the ground truth.

4 Results and discussion

4.1 Experiment I: ground truth provided by sparsely labeled MTs

We first studied active nematics containing both densely and sparsely labeled MTs with different fluorophores. The sample was imaged sequentially in the dense and sparse channel. Using these samples we first performed PIV and DLOF on densely labeled samples. This data was compared to particle tracking of sparsely-labeled active nematics, which served as ground truth (Fig. 3). The velocities estimated by PIV for densely labeled systems are inaccurate. DLOF overcomes this limitation providing more accurate estimates of both the velocity magnitude and direction. We hypothesize that the breakdown of the PIV for densely labeled systems arises because the algorithm cannot reliably distinguish contrast variations at high densities. As we show below, the breakdown is strongest in directions parallel to the director field. PIV significantly underestimates the velocity tangent to the MT bundles because the contrast is more uniform in that direction, as was previously reported.^{58,73-76}

To quantify the above-described observations, we used PIV and DLOF to estimate the velocity fields from the dense and dilute channels. We compared these to the ground truth based on single-particle tracking. PIV and DLOF estimate the flow field everywhere while single particle tracking yields velocities only at the location of tracked points. The velocity magnitude error is calculated by $|||\mathbf{v}|| - ||\mathbf{v}^*||| / ||\mathbf{v}^*||$ where \mathbf{v}^* is the true displacement vector obtained from particle tracking at a particular position and \mathbf{v} is the velocity obtained at the same position from either the PIV or DLOF. The orientation error θ is calculated using the cosine similarity, where $\cos(\theta) = \mathbf{v} \cdot \mathbf{v}^* / (||\mathbf{v}|| \cdot ||\mathbf{v}^*||)$. By repeating the procedure for all tracked particles we obtained the distribution of measurement errors (Fig. 4). PIV and DLOF have comparable errors for sparse labels (Fig. 4B). However, with dense labels, PIV results were more unreliable. In contrast, the DLOF estimates are nearly as good as those with sparse labels (Fig. 4A). Similarly, the mean orientation errors of PIV and DLOF are also comparable when using sparse labels, 14 and 17 degrees, respectively (Fig. 4D). The discrepancy between orientation errors produced by PIV and DLOF becomes significant when using dense labels, where the mean orientation error of PIV increases to 44 degrees while that of DLOF is 29 degrees (Fig. 4C).

Previous studies^{68,74-76} had shown that uniform contrast along densely labeled MT bundles poses a major challenge to PIV, resulting in significantly underestimated velocity component tangent to the MT bundles. We therefore evaluated the contribution of this effect to our observed breakdown of PIV as follows. We extracted the director, *i.e.*, the local orientation, of the MT bundles using the dense labels and computed average errors of velocities obtained by PIV and DLOF as functions of



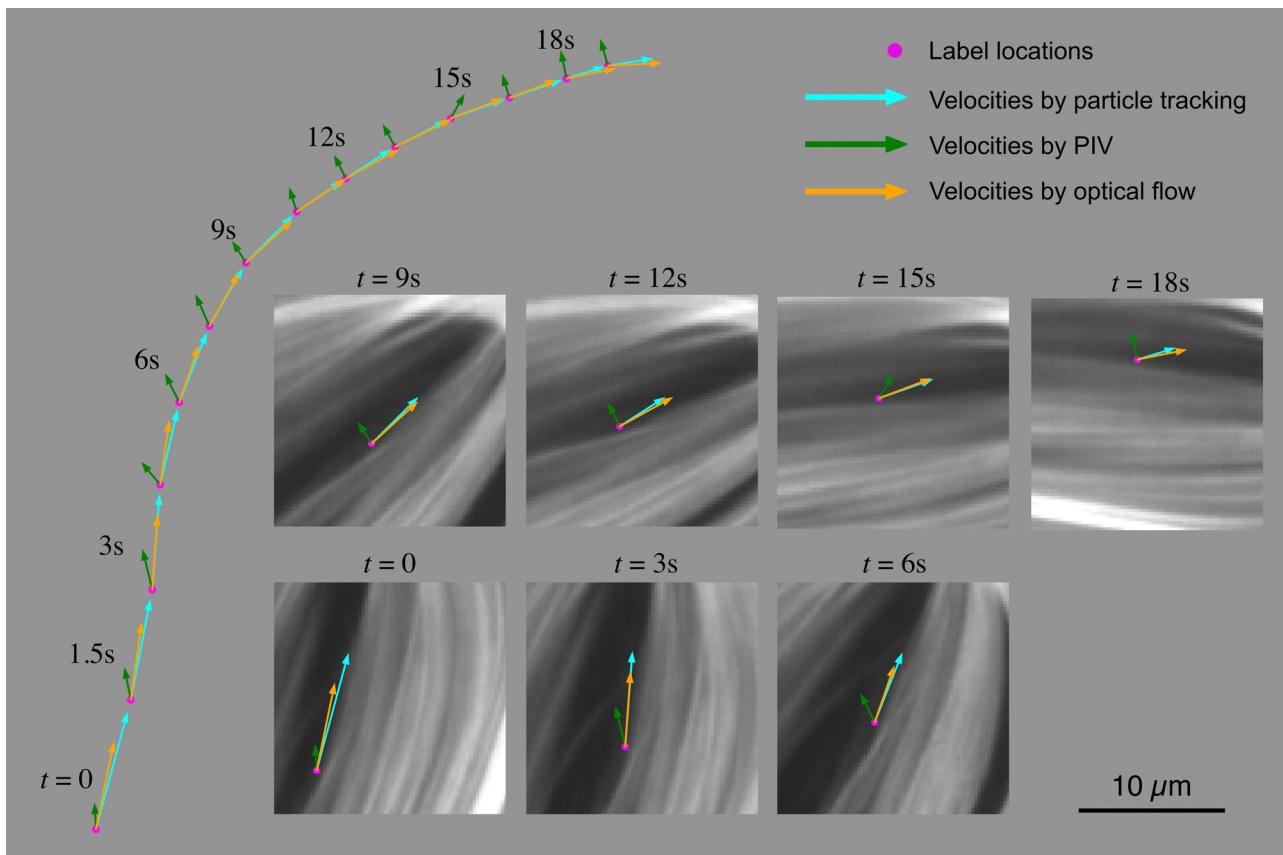


Fig. 3 DLOF outperforms PIV for densely labeled samples. (left) The trajectory of an individual MT, which is imaged every 1.5 seconds. MT true velocities (cyan arrows) are obtained by particle tracking. The velocity vectors estimated by PIV and DLOF are indicated respectively with green and orange arrows. The insets depict the densely labeled MTs in local neighborhoods of the tracked labels at the indicated times. The high densities of the labels in the images pose a significant challenge to PIV, resulting in inaccurate velocity estimates. In contrast, DLOF produces highly accurate velocities. Particle tracking was extracted from a simultaneously imaged sparsely labeled channel.

the angle between ground truth velocity and director (Fig. 5). We find that when the MTs are moving in directions with significant components along the director, PIV produces high relative speed errors (Fig. 5A) and orientation errors (Fig. 5B). DLOF strongly improves the estimation of velocities in these directions. In particular, the improvement of DLOF over PIV uniformly increases as the velocity direction approaches the director field. When the velocities are parallel to the directors (*i.e.*, angles between velocity and director are less than 1 degree), the average relative speed error is reduced by 37% with DLOF (compared to PIV), and average orientation error reduced by 31%. This analysis shows that DLOF can resolve this well-known limitation of PIV, and thus establishes DLOF as an alternative method capable of obtaining accurate velocity fields with dense labels.

4.1.1 Comparing PIV and DLOF spatial flow fields. Thus far, our analysis has focused on the accuracy of the PIV and DLOF methods in estimating the velocities of individual traced labels. Next, we evaluate the quality of the two-dimensional flow fields produced by each method. In this case, we do not have ground truth to compare against, since the tracked dilute MTs do not yield the spatial flow fields. Previous analysis

showed that PIV and DLOF are comparable for sparsely labeled systems. Therefore, we use the flow fields determined by PIV with sparse labels as the baseline. For a meaningful comparison, we note that PIV produced the velocity fields on lower-resolution spatial grids when compared to DLOF. Therefore, we interpolate the DLOF results onto the lower-resolution grid of the PIV results. PIV and DLOF produce consistent flow fields for sparsely labeled samples (Fig. 6). However, the DLOF results are significantly smoother. While the DLOF results are somewhat noisier for the densely labeled system, the correct flow structure is maintained. In comparison, PIV on densely labeled systems produces an inaccurate flow structure. Importantly, the DLOF model correctly estimates velocities across different regions and different scales of the flow speed. For example, MT bundles move faster in the vicinity of $+1/2$ topological defects and slower near $-1/2$ defects.

We compared the flow speeds obtained from PIV and DLOF averaged over the entire field (Fig. 7). Consistent with the previous analysis above, the PIV and DLOF estimates are nearly identical for sparsely labeled samples. The DLOF estimates for dense labels fall within the 95% confidence interval. In contrast, PIV significantly underestimates the velocities for dense labels.

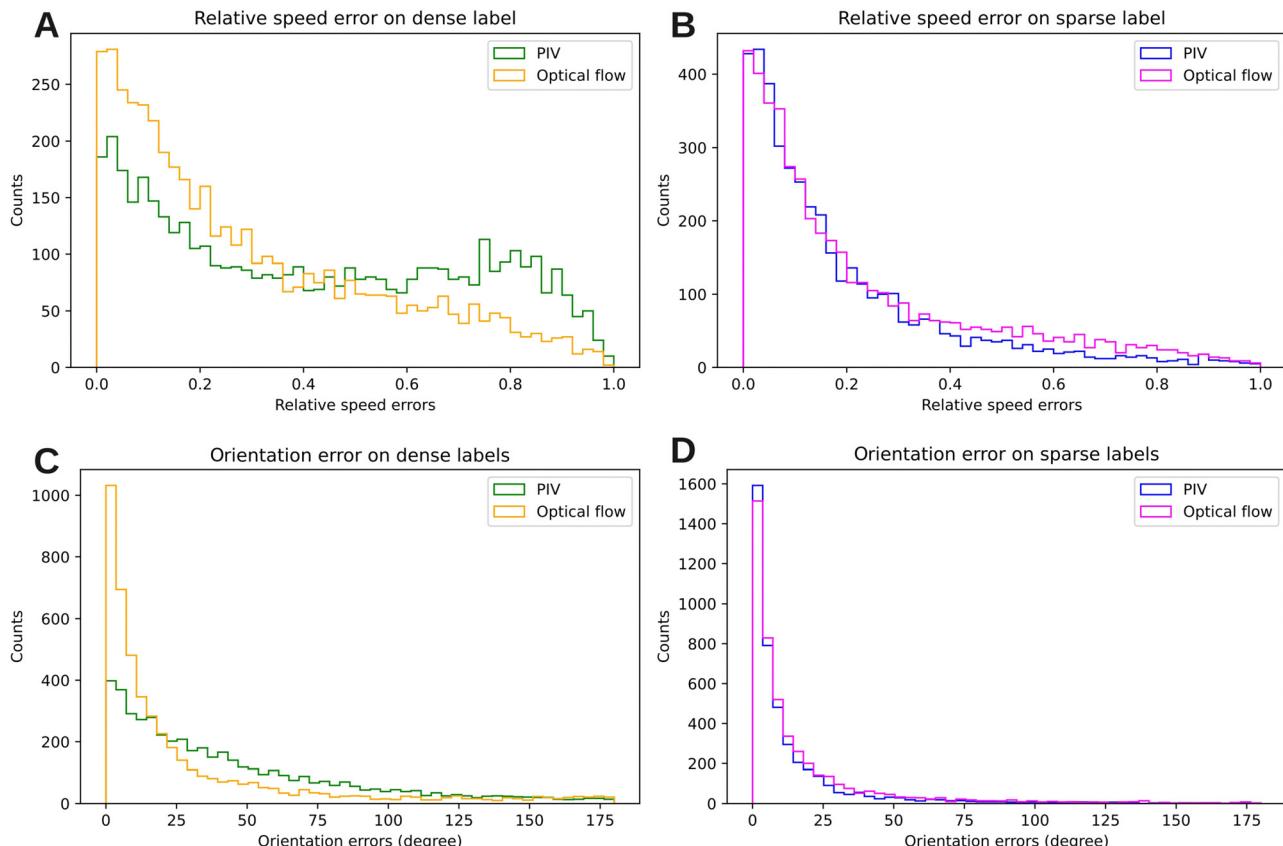


Fig. 4 Comparing PIV and DLOF to single-filament tracking. Distribution of errors when comparing PIV and DLOF velocity fields from sparsely and densely labeled samples to single-filament tracking. The distributions of errors in the magnitude and orientation of the velocity (defined in the text) for PIV and DLOF. Errors are computed by comparing different estimates with particle tracking results. The mean relative speed errors for PIV are 42% and 19% for densely and sparsely labeled systems; errors for DLOF are 29% and 23%. The mean orientation errors for PIV are 44 degrees and 14 degrees for densely and sparsely labeled systems; errors for DLOF are 29 degrees and 17 degrees. The distributions are obtained from 4738 traced labels across 44 frames in Experiment 1.

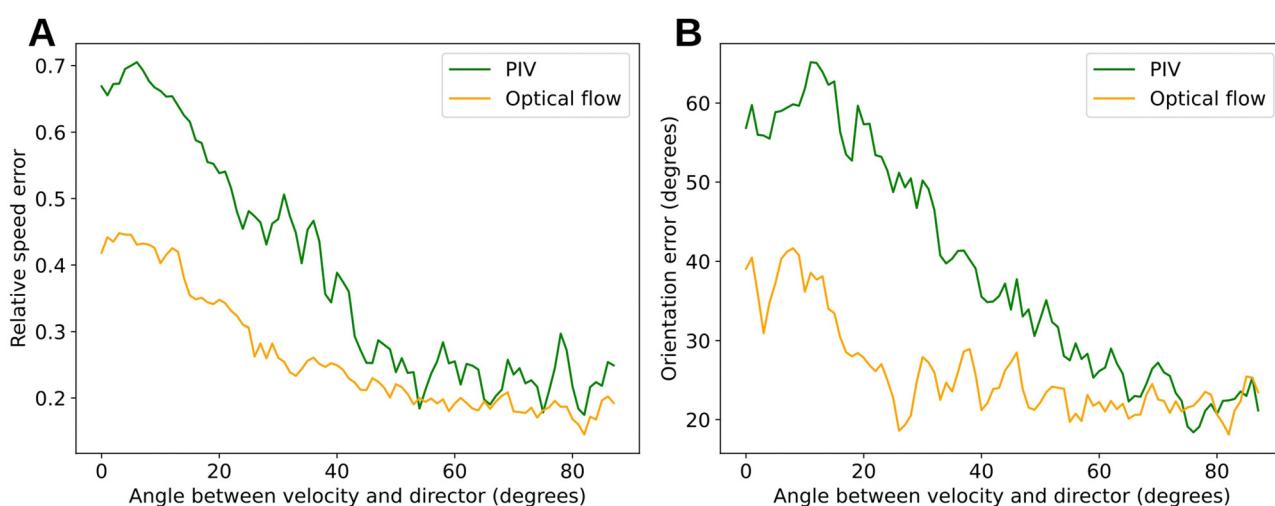


Fig. 5 The improvement of DLOF over PIV increases as the velocity becomes parallel to the director field (for dense labels). Average relative speed error (A) and average orientation error (B) of PIV and DLOF as a function of the angle between ground truth velocity and director. PIV particularly breaks when the velocities are tangent to the MT bundles due to the uniform contrast of the dense labels along MT bundles. DLOF can handle the uniform contrast along MT bundles and thus produces much more accurate velocities.



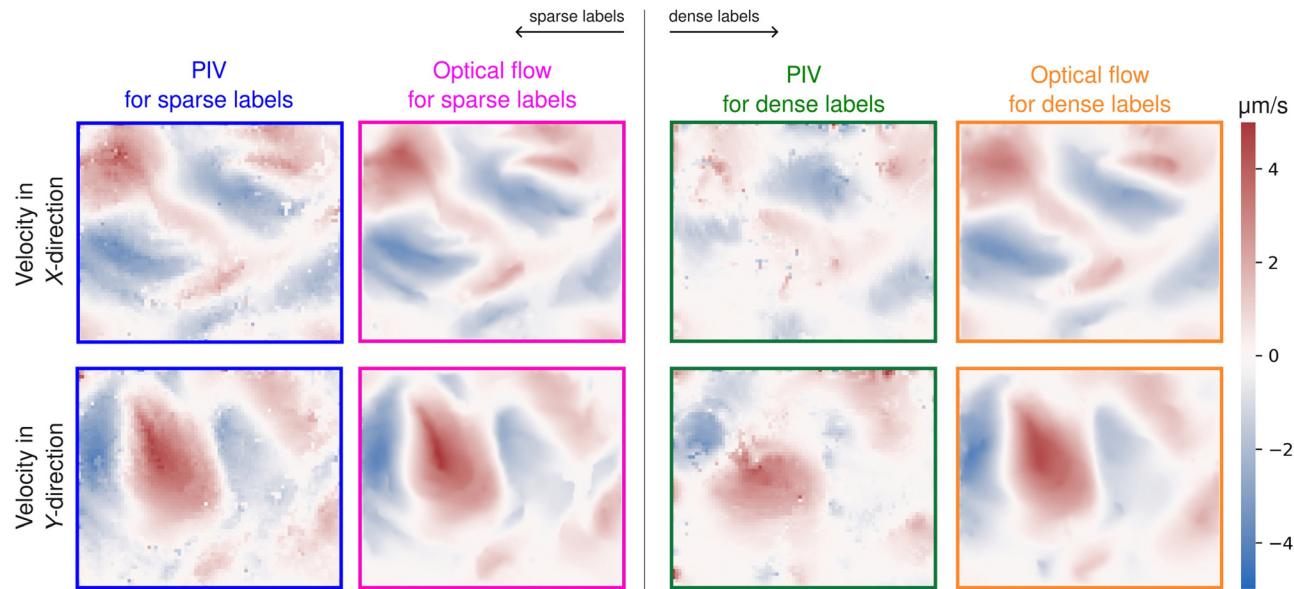


Fig. 6 Comparison of the velocity fields in the x -direction (top row) and y -direction (bottom row) produced by PIV and DLOF for sparse labels (blue and magenta highlighted), and by PIV and DLOF for dense labels (green and orange highlighted). The velocity fields are calculated for the first frame obtained from Experiment 1. DLOF always produces smoother fields, due to its capability to estimate displacements on a pixel-level. Remarkably, when dealing with dense labels, velocity fields estimated by DLOF are significantly more accurate than those produced by PIV (by comparing green and orange boxes for each velocity component).

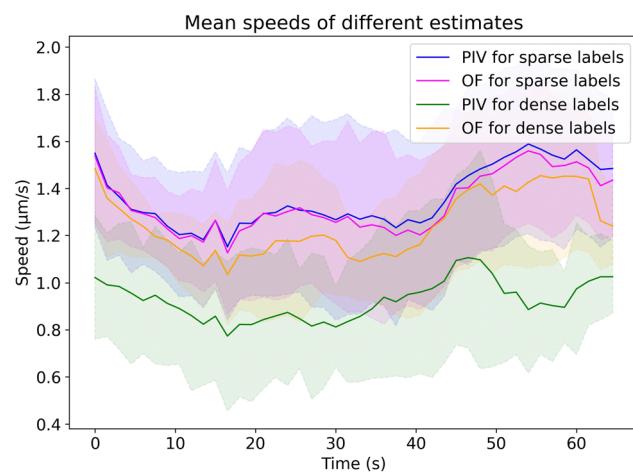


Fig. 7 Comparison of mean flow speeds as a function of time. The flow speeds ($\mu\text{m s}^{-1}$) averaged over the entire spatial domain are shown as a function of time over the 44 frames of the benchmark video using the dense labels. The frame interval is 1.5 seconds, and results are shown for sparse and dense labels for PIV and optical flow. The shaded areas show 95% confidence levels of the mean speeds.

For a final comparison, we define the normalized zero-lag cross-correlation between an estimated velocity and the ground truth as

$$\mathcal{C} = \frac{\sum_i \mathbf{v}_i \cdot \mathbf{v}_i^*}{\sum_i \mathbf{v}_i^* \cdot \mathbf{v}_i^*} = \frac{\sum_i \mathbf{v}_i \cdot \mathbf{v}_i^*}{\sum_i \|\mathbf{v}_i^*\|^2}, \quad (8)$$

where \mathbf{v}_i and \mathbf{v}_i^* are the estimated and the ground truth velocities of the traced label i , and \sum_i sums over all the traced

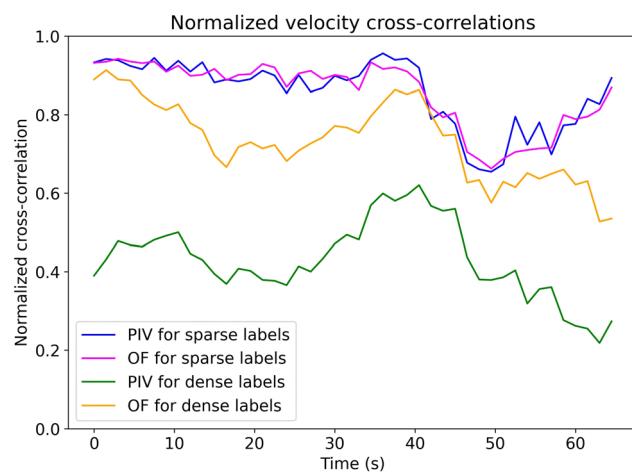


Fig. 8 Normalized zero-lag cross-correlation between velocity estimates and ground truth. The normalized spatial correlation (eqn (8)) is shown for optical flow on sparsely and densely labeled systems, as well as PIV on densely labeled systems, as a function of time.

labels in the current frame. A perfect velocity estimation would yield $\mathcal{C} = 1$, while $\mathcal{C} > 1$ indicates that, on average, flow speeds are overestimated and $\mathcal{C} < 1$ underestimated. PIV and DLOF perform similarly for sparse labels (Fig. 8). The performance discrepancy between PIV and DLOF becomes significant for dense labels, where velocities produced by DLOF are still highly correlated with the ground truths. In contrast, velocities estimated by PIV result in significantly lower correlations.

4.2 Experiment 2: ground truth provided by passive beads

We also compared DLOF and PIV against tracked passive beads, which served as the ground truth. This measurement assumes



that passive beads within active nematic samples provide good estimates of MT velocities, as previously shown by Tayar *et al.* (Fig. 12).⁷³ We include it here for two reasons. First, it provides an alternative means to compare the relative accuracy of DLOF and PIV, without relying on the PIV measurement itself. Second, the fact that (as shown next) we observe nearly identical results from both methods is a significant observation. Active nematics are hierarchical materials, and the velocity field can depend on the length scale on which it is characterized. At the microscopic scale, clusters of molecular motors induce relative extensile sliding of adjacent MTs. In principle, tracking individual filaments could at least partially include this microscopic dynamics, whereas micron-sized beads will not. Hydrodynamic theories coarse-grain over such jittery microscopic motions, suggesting that micron-sized beads could be a better tracer for measuring such coarse-grained velocity fields. The fact that we observe an indistinguishable velocity field between the two approaches means that both methods provide a good description of the coarse-grained velocity field.

In each frame, we compared the instantaneous velocity of each bead to the velocities at the same position generated by PIV and DLOF. Since we computed PIV on a sparse grid,

we interpolated its values as necessary to correspond to bead positions. As in Experiment 1, the comparison shows that DLOF is more accurate than PIV (Fig. 9). In particular, the difference in speeds between the beads and the DLOF velocities was significantly smaller than that between the beads and PIV (Fig. 9A). Similarly, the angular orientations of DLOF velocities were also closer to the bead velocities (Fig. 9B). At each time point, the spatially averaged mean speed of the DLOF field was closer to that of the beads, while the mean speed of PIV was systematically lower (Fig. 9C). This result is consistent with the notion that PIV systematically underestimates the motion of MTs when their motion is locally parallel, rather than perpendicular, to intensity gradients in the image on a length scale larger than the size of PIV's interrogation region.^{6,7} Lastly, the zero-lag cross-correlation eqn (8) between the DLOF and bead velocities was consistently higher than the correlation between PIV and bead velocities (Fig. 9D).

Our benchmarks demonstrate the accuracy of DLOF for extracting velocities from active nematics, surpassing the limitations of traditional PIV methods. Although we have trained and demonstrated the model on 2D active nematics samples captured with a 60 \times magnification objective, we note that it appears to generalize well to other magnifications and situations, such as 2D slices from a 3D isotropic active MT system⁶⁴ captured at lower magnification (10 \times), provided that: there is sufficient contrast between labeled MTs and the background, the illumination of MTs does not change significantly between the two input frames, and the movements between two input frames are smaller than the algorithm's search window and the scale of the moving textures in the images.

5 Conclusions

We compared DLOF and PIV for estimating the velocity fields of active nematics, by generating ground truth velocity fields that enabled quantitative comparison of the two techniques. DLOF produces spatially smoother velocity fields. It also generates more accurate flows than PIV for high densities of fluorescent filaments. The high performance of DLOF arises because it determines displacements between frames by finding maximum cross-correlations in the rich feature space extracted by deep neural networks across multiple scales of spatial resolution. Furthermore, unlike PIV, DLOF eliminates the need to manually tune and readjust the model's parameters when working with data that have high contrast variances across the entire data. This is essential for analyzing large amounts of data, or for real-time control applications where it is impractical to manually tune parameters of algorithms such as PIV.

Importantly, these results have implications that extend beyond the field of active nematics. Active nematics velocity fields resemble those of a wide variety of soft matter and biophysical systems,^{77–79} suggesting that DLOF might be more accurate than PIV in these systems as well. In particular, it is likely that DLOF will significantly outperform PIV in other systems with anisotropic constituents, based on our

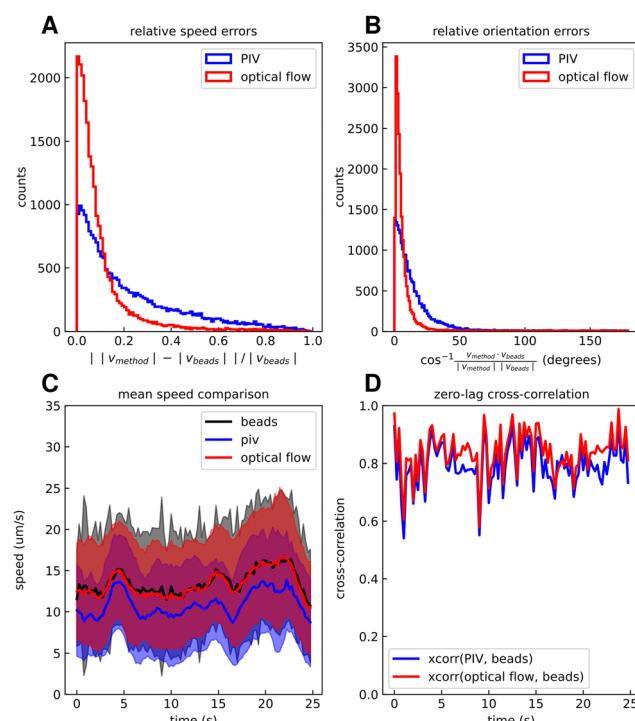


Fig. 9 Comparison of PIV and DLOF where passive tracer beads generate the ground-truth velocities. (A) Histograms of speed differences between PIV and bead velocities, and between optical flow and bead velocities. (B) Histograms of angular orientation differences between PIV and bead velocities, and between optical flow and bead velocities. (C) Mean speed of the beads, PIV, and optical flow over time. The speed is averaged over all available points for the given field (note that there are far more optical flow points than PIV points, and far more PIV points than beads, in each frame). Error bars indicate the standard deviation. (D) Zero-lag cross-correlation between PIV and bead velocities, and between PIV and optical flow velocities over time (eqn (8)).



observation that PIV is especially inaccurate in estimating velocities along the long direction of particles. However, as noted above, PIV may be more accurate than DLOF when there is highly nonuniform illumination on the sample or the resolution is insufficient to visually represent the moving textures in the data between successive frames.

There is growing interest in applying data-driven and machine-learning approaches to physics and materials discovery,^{94–102} but these approaches are limited by the availability of training data. The ability of DLOF to autonomously generate high-quality velocity fields is a crucial step for advancing these applications.

Data availability

The data that support the findings of this study are openly available at <https://github.com/trangocphu/opticalflow-activenematics>.

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

This work was supported by the Department of Energy (DOE) DE-SC0022291. Computing resources were provided by the NSF XSEDE allocation TG-MCB090163 and the Brandeis HPCC which is partially supported by the NSF through DMR-MRSEC 2011846 and OAC-1920147. The authors thank Michael M. Norton (Physics Department, Brandeis University) for insightful feedback on the manuscript.

Notes and references

- 1 T. Corpetti, E. Memin and P. Perez, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, **24**, 365–380.
- 2 M. C. Marchetti, J. F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, M. Rao, R. A. Simha and M. Curie, *Rev. Mod. Phys.*, 2013, **85**, 1143–1189.
- 3 S. Vogel, *Life in moving fluids: the physical biology of flow—revised and expanded*, Princeton University Press, 2nd edn, 2020.
- 4 W. Thielicke and R. Sonntag, *J. Open Res. Softw.*, 2021, **9**, 12.
- 5 L. Sarno, A. Caravetta, Y. C. Tai, R. Martino, M. N. Papa and C. Y. Kuo, *Adv. Powder Technol.*, 2018, **29**, 3107–3123.
- 6 M. Raffel, C. E. Willert, F. Scarano, C. J. Kähler, S. T. Wereley and J. Kompenhans, *Particle image velocimetry: a practical guide*, Springer, 2018.
- 7 S. Scharnowski and C. J. Kähler, *Opt. Laser Eng.*, 2020, **135**, 106185.
- 8 M. G. Olsen and C. J. Bourdon, *Meas. Sci. Technol.*, 2007, **18**, 1963.
- 9 C. J. Kähler, S. Scharnowski and C. Cierpka, *Exp. Fluids*, 2012, **52**, 1629–1639.
- 10 J. L. Barron, D. J. Fleet and S. S. Beauchemin, *Int. J. Comput. Vis.*, 1994, **12**, 43–77.
- 11 A. Verri, F. Girosi and V. Torre, *JOSA A*, 1990, **7**, 912–922.
- 12 A. Bainbridge-Smith and R. G. Lane, *Image Vis. Comput.*, 1997, **15**, 11–22.
- 13 P. Baraldi, A. Sarti, C. Lamberti, A. Prandini and F. Sgallari, *IEEE Trans. Biomed. Eng.*, 1996, **43**, 259–272.
- 14 A. Bruhn, J. Weickert and C. Schnörr, *Int. J. Comput. Vis.*, 2005, **61**, 211–231.
- 15 A. Bruhn, J. Weickert, C. Feddern, T. Kohlberger and C. Schnörr, *IEEE Trans. Image Process.*, 2005, **14**, 608–615.
- 16 I. Cohen, Proceedings of the Scandinavian conference on image analysis, 1993.
- 17 A. Bruhn, J. Weickert, C. Feddern, T. Kohlberger and C. Schnörr, *Computer Analysis of Images and Patterns*, Berlin, Heidelberg, 2003, pp. 222–229.
- 18 Z. Tu, W. Xie, D. Zhang, R. Poppe, R. C. Veltkamp, B. Li and J. Yuan, *Signal Process. Image Commun.*, 2019, **72**, 9–24.
- 19 C.-C. Cheng, K.-H. Ho, H.-T. Li and G.-L. Lin, Proceedings of the IEEE International Symposium on Intelligent Control, 2002, pp. 350–355.
- 20 C.-C. Cheng and H.-T. Li, *Int. J. Inf. Technol.*, 2006, **12**, 82–90.
- 21 S. S. Beauchemin and J. L. Barron, *ACM Comput. Surv.*, 1995, **27**, 433–466.
- 22 A. Becciu, H. van Assen, L. Florack, S. Kozerke, V. Roode and B. M. ter Haar Romeny, *Scale Space and Variational Methods in Computer Vision*, Berlin, Heidelberg, 2009, pp. 588–599.
- 23 J.-Y. Bouguet, *et al.*, *Intel corporation*, 2001, vol. 5, p. 4.
- 24 G. Farnebäck, *Image Analysis*, Berlin, Heidelberg, 2003, pp. 363–370.
- 25 T. Brox, A. Bruhn, N. Papenberg and J. Weickert, Computer Vision-ECCV 2004: 8th European Conference on Computer Vision, Prague, Czech Republic, May 11–14, 2004, Proceedings, Part IV 8, 2004, pp. 25–36.
- 26 C. Zach, T. Pock and H. Bischof, *Pattern Recognition*, Berlin, Heidelberg, 2007, pp. 214–223.
- 27 X. Yong, C.-K. Huang and C. T. Lim, *J. R. Soc., Interface*, 2021, **18**, 20210248.
- 28 D. K. Vig, A. E. Hamby and C. W. Wolgemuth, *Biophys. J.*, 2016, **110**, 1469–1475.
- 29 A. Ranjan and M. J. Black, Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4161–4170.
- 30 P. Fischer, A. Dosovitskiy, E. Ilg, P. Häusser, C. Hazrbas, V. Golkov, P. Van der Smagt, D. Cremers and T. Brox, IEEE International Conference on Computer Vision (ICCV), 2015, pp. 2758–2766, DOI: [10.1109/ICCV.2015.316](https://doi.org/10.1109/ICCV.2015.316).
- 31 Z. Ren, J. Yan, B. Ni, B. Liu, X. Yang and H. Zha, Proceedings of the AAAI conference on artificial intelligence, 2017.
- 32 C. Yu, X. Bi and Y. Fan, *Ocean Eng.*, 2023, **271**, 113693.
- 33 S. Bai, Z. Geng, Y. Savani and J. Z. Kolter, Deep Equilibrium Optical Flow Estimation, Proceedings of the IEEE/CVF



Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 620–630, DOI: [10.1109/CVPR52688.2022.00070](https://doi.org/10.1109/CVPR52688.2022.00070).

34 A. Bar-Haim and L. Wolf, Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 7995–8004, DOI: [10.1109/CVPR42600.2020.00802](https://doi.org/10.1109/CVPR42600.2020.00802).

35 Y. Han, K. Luo, A. Luo, J. Liu, H. Fan, G. Luo and S. Liu, European conference on computer vision, 2022, DOI: [10.1007/978-3-031-19800-7_17](https://doi.org/10.1007/978-3-031-19800-7_17).

36 Z. Huang, X. Shi, C. Zhang, Q. Wang, K. C. Cheung, H. Qin, J. Dai and H. Li, European conference on computer vision, 2022, pp. 668–685, DOI: [10.1007/978-3-031-19790-1_40](https://doi.org/10.1007/978-3-031-19790-1_40).

37 T.-W. Hui, X. Tang and C. C. Loy, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, pp. 2555–2569, DOI: [10.1109/TPAMI.2020.2976928](https://doi.org/10.1109/TPAMI.2020.2976928).

38 E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy and T. Brox, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2462–2470.

39 J. Jeong, H. Cai, R. Garrepalli and F. Porikli, Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023, pp. 13691–13700, DOI: [10.1109/CVPR52729.2023.01316](https://doi.org/10.1109/CVPR52729.2023.01316).

40 S. Jiang, D. Campbell, Y. Lu, H. Li and R. Hartley, Proceedings of IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 9752–9761, DOI: [10.1109/ICCV48922.2021.00963](https://doi.org/10.1109/ICCV48922.2021.00963).

41 S. Jiang, Y. Lu, H. Li and R. Hartley, Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 16587–16595, DOI: [10.1109/CVPR46437.2021.01632](https://doi.org/10.1109/CVPR46437.2021.01632).

42 L. Liu, J. Zhang, R. He, Y. Liu, Y. Wang, Y. Tai, D. Luo, C. Wang, J. Li and F. Huang, 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 6488–6497.

43 S. Liu, K. Luo, N. Ye, C. Wang, J. Wang and B. Zeng, *IEEE Trans. Image Process.*, 2021, **30**, 6420–6433.

44 K. Luo, C. Wang, S. Liu, H. Fan, J. Wang and J. Sun, 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 2021, pp. 1045–1054.

45 A. Luo, F. Yang, X. Li and S. Liu, 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 2022, pp. 8896–8905.

46 C. Min, T. Kim and J. Lim, 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 2144–2153.

47 A. Nebisoy and S. Malekzadeh, *arXiv*, 2021, preprint, arXiv: 2103.05101, DOI: [10.48550/arXiv.2103.05101](https://doi.org/10.48550/arXiv.2103.05101).

48 Z. Pan, D. Geng and A. Owens, Proceedings of the 37th International Conference on Neural Information Processing Systems, 2023, pp. 253–273, DOI: [10.5555/3666122.3666135](https://doi.org/10.5555/3666122.3666135).

49 A. Stone, D. Maurer, A. Ayvaci, A. Angelova and R. Jonschkowski, Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 3886–3895, DOI: [10.1109/CVPR46437.2021.00388](https://doi.org/10.1109/CVPR46437.2021.00388).

50 D. Sun, X. Yang, M.-Y. Liu and J. Kautz, *PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume*, 2018.

51 D. Sun, C. Herrmann, F. Reda, M. Rubinstein, D. Fleet and W. T. Freeman, European Conference on Computer Vision, 2022, pp. 165–182, DOI: [10.1007/978-3-031-20047-2_10](https://doi.org/10.1007/978-3-031-20047-2_10).

52 A. Ullah, K. Muhammad, J. Del Ser, S. W. Baik and V. H. C. de Albuquerque, *IEEE Trans. Ind. Electron.*, 2019, **66**, 9692–9702.

53 H. Xu, J. Yang, J. Cai, J. Zhang and X. Tong, Proceedings of IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 10478–10487, DOI: [10.1109/ICCV48922.2021.01033](https://doi.org/10.1109/ICCV48922.2021.01033).

54 H. Xu, J. Zhang, J. Cai, H. Rezatofighi and D. Tao, Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 8111–8120, DOI: [10.1109/CVPR52688.2022.00795](https://doi.org/10.1109/CVPR52688.2022.00795).

55 S. Zhao, Y. Sheng, Y. Dong, E. I.-C. Chang and Y. Xu, 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 6277–6286.

56 S. Cai, S. Zhou, C. Xu and Q. Gao, *Exp. Fluids*, 2019, **60**, 73.

57 S. Discetti and Y. Liu, *Meas. Sci. Technol.*, 2022, **34**, 021001.

58 C. Lagemann, K. Lagemann, S. Mukherjee and W. Schröder, *Nat. Mach. Intell.*, 2021, **3**, 641–651.

59 C. Yu, X. Bi, Y. Fan, Y. Han and Y. Kuai, *IEEE Trans. Instrum. Meas.*, 2021, **70**, 1–15.

60 C. Yu, X. Bi and Y. Fan, *Ocean Eng.*, 2023, **271**, 113693.

61 W. Zhang, X. Dong, Z. Sun and S. Xu, *Phys. Fluids*, 2023, **35**, 077108.

62 R. Aditi Simha and S. Ramaswamy, *Phys. Rev. Lett.*, 2002, **89**, 058101.

63 V. Narayan, S. Ramaswamy and N. Menon, *Science*, 2007, **317**, 105–108.

64 T. Sanchez, D. T. N. Chen, S. J. DeCamp, M. Heymann and Z. Dogic, *Nature*, 2012, **491**, 431–434.

65 P. Guillamat, J. Ignés-Mullol, S. Shankar, M. C. Marchetti and F. Sagués, *Phys. Rev. E*, 2016, **94**, 1–5.

66 N. Kumar, R. Zhang, J. J. de Pablo and M. L. Gardel, *Sci. Adv.*, 2018, **4**, eaat7779.

67 C. Blanch-Mercader, V. Yashunsky, S. Garcia, G. Duclos, L. Giomi and P. Silberzan, *Phys. Rev. Lett.*, 2018, **120**, 208101.

68 A. J. Tan, E. Roberts, S. A. Smith, U. A. Olvera, J. Arteaga, S. Fortini, K. A. Mitchell and L. S. Hirst, *Nat. Phys.*, 2019, **15**, 1033–1039.

69 L. Giomi, L. Mahadevan, B. Chakraborty and M. F. Hagan, *Phys. Rev. Lett.*, 2011, **106**, 2–5.

70 L. Giomi, L. Mahadevan, B. Chakraborty and M. F. Hagan, *Nonlinearity*, 2012, **25**, 2245–2269.

71 S. P. Thampi, R. Golestanian and J. M. Yeomans, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2014, **90**, 1–5.

72 T. N. Shendruk, A. Doostmohammadi, K. Thijssen and J. M. Yeomans, *Soft Matter*, 2017, **13**, 3853–3862.

73 A. M. Tayar, L. M. Lemma and Z. Dogic, *Microtubules: Methods and Protocols*, Springer, 2022, pp. 151–183.

74 A. Opathalage, M. M. Norton, M. P. N. Juniper, B. Langeslay, S. A. Aghvami, S. Fraden and Z. Dogic, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 4788–4797.

75 F. L. Memarian, D. Hammar, M. M. H. Sabbir, M. Elias, K. A. Mitchell and L. Hirst, *Phys. Rev. Lett.*, 2024, 228301.



76 M. Serra, L. Lemma, L. Giomi, Z. Dogic and L. Mahadevan, *Nat. Phys.*, 2023, 1–7.

77 C. A. Browne and S. S. Datta, *Sci. Adv.*, 2021, 7, eabj2619.

78 X. Wen, Y. Sang, Y. Zhang, F. Ge, G. Jing and Y. He, *ACS Nano*, 2023, **17**, 10104–10112.

79 P. Ramesh, B. V. Hokmabad, D. O. Pushkin, A. J. T. M. Mathijssen and C. C. Maass, *J. Fluid Mech.*, 2023, **966**, A29.

80 N. P. Mitchell, D. J. Cislo, S. Shankar, Y. Lin, B. I. Shraiman and S. J. Streichan, *eLife*, 2022, **11**, e77355.

81 S. J. Streichan, M. F. Lefebvre, N. Noll, E. F. Wieschaus and B. I. Shraiman, *eLife*, 2018, **7**, e27454.

82 D. Mondal, R. Adhikari and P. Sharma, *Sci. Adv.*, 2020, **6**, eabb0503.

83 R. Jonschkowski, A. Stone, J. T. Barron, A. Gordon, K. Konolige and A. Angelova, Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16, 2020, pp. 557–572.

84 Z. Teed and J. Deng, Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16, 2020, pp. 402–419.

85 D. J. Butler, J. Wulff, G. B. Stanley and M. J. Black, European Conf. on Computer Vision (ECCV), 2012, pp. 611–625.

86 M. Menze and A. Geiger, Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

87 A. Geiger, P. Lenz and R. Urtasun, Conference on Computer Vision and Pattern Recognition (CVPR), 2012.

88 A. Geiger, P. Lenz, C. Stiller and R. Urtasun, *Int. J. Robot. Res.*, 2013, 1231–1237.

89 J. Fritsch, T. Kuehnl and A. Geiger, International Conference on Intelligent Transportation Systems (ITSC), 2013.

90 A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. van der Smagt, D. Cremers and T. Brox, Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015.

91 N. Mayer, E. Ilg, P. Häusser, P. Fischer, D. Cremers, A. Dosovitskiy and T. Brox, IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

92 R. Dey and F. M. Salem, 2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS), 2017, pp. 1597–1600.

93 A. Gordon, H. Li, R. Jonschkowski and A. Angelova, Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 8977–8986.

94 B. W. Brunton, L. A. Johnson, J. G. Ojemann and J. N. Kutz, *J. Neurosci. Meth.*, 2016, **258**, 1–15.

95 E. de Bézenac, A. Pajot and P. Gallinari, *J. Stat. Mech.: Theory Exp.*, 2019, **2019**, 124009.

96 F. Cichos, K. Gustavsson, B. Mehlig and G. Volpe, *Nat. Mach. Intell.*, 2020, **2**, 94–103.

97 J. Colen, M. Han, R. Zhang, S. A. Redford, L. M. Lemma, L. Morgan, P. V. Ruijgrok, R. Adkins, Z. Bryant, Z. Dogic, M. L. Gardel, J. J. de Pablo and V. Vitelli, *Proc. Natl. Acad. Sci. U. S. A.*, 2021, **118**, 10.

98 C. Joshi, S. Ray, L. M. Lemma, M. Varghese, G. Sharp, Z. Dogic, A. Baskaran and M. F. Hagan, *Phys. Rev. Lett.*, 2022, **129**, 258001.

99 Z. Zhou, C. Joshi, R. Liu, M. M. Norton, L. Lemma, Z. Dogic, M. F. Hagan, S. Fraden and P. Hong, *Soft Matter*, 2021, **17**, 738–747.

100 S. L. Brunton and J. N. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*, Cambridge University Press, 2022.

101 M. Golden, R. O. Grigoriev, J. Nambisan and A. Fernandez-Nieves, *Sci. Adv.*, 2023, **9**, eabq6120.

102 Y. Li, Z. Zarei, P. N. Tran, Y. Wang, A. Baskaran, S. Fraden, M. F. Hagan and P. Hong, *Soft Matter*, 2024, **20**, 1869–1883.

