

Spike Imaging Velocimetry: Dense Motion Estimation of Fluids Using Spike Streams

Yunzhong Zhang^{1,2*}, You Zhou^{3*}, Changqing Su⁴, Zhen Cheng⁵, Zhaofei Yu⁴, Bo Xiong^{6†}, Tiejun Huang⁴, Xun Cao²

¹School of Physics, Nanjing University

²School of Electronic Science and Engineering, Nanjing University

³Medical School, Nanjing University

⁴State Key Laboratory for Multimedia Information Processing, Peking University

⁵Department of Automation, Tsinghua University

⁶Institute of Medical Technology, Peking University Health Science Center, Peking University

ltq@smail.nju.edu.cn, {zhouyou, caoxun}@nju.edu.cn, cqsu25@stu.pku.edu.cn, zcheng@mail.tsinghua.edu.cn, {yuzf12, xiongbo, tjhuang}@pku.edu.cn,

Abstract

Particle Image Velocimetry (PIV) is a widely adopted non-invasive imaging technique that tracks the motion of tracer particles across image sequences to capture the velocity distribution of fluid flows. It is commonly employed to analyze complex flow structures and validate numerical simulations. This study explores the untapped potential of spike cameras—ultra-high-speed, high-dynamic-range vision sensors—in high-speed fluid velocimetry. We propose a deep learning framework, Spike Imaging Velocimetry (SIV), tailored for high-resolution fluid motion estimation. To enhance the network’s performance, we design three novel modules specifically adapted to the characteristics of fluid dynamics and spike streams: the Detail-Preserving Hierarchical Transform (DPHT), the Graph Encoder (GE), and the Multi-scale Velocity Refinement (MSVR). Furthermore, we introduce a spike-based PIV dataset, Particle Scenes with Spike and Displacement (PSSD), which contains labeled samples from three representative fluid-dynamics scenarios: steady turbulence, high-speed flow, and high-dynamic-range conditions. Our proposed method outperforms existing baselines across all these scenarios, demonstrating its effectiveness.

Introduction

The need for accurate and non-invasive flow measurement techniques has led to the widespread use of Particle Image Velocimetry (PIV), a key diagnostic tool for fluid dynamics. PIV has been applied to critical areas such as studying aircraft aerodynamics (Kompenhans et al. 2000), optimizing vehicle designs for reduced air resistance (Nakagawa et al. 2016), improving energy efficiency in combustion processes (Chen, Li, and Zhu 2022), understanding blood flow dynamics in diseases (Raschi et al. 2012), and predicting water pollution spread (MoayeriKashani et al. 2016). In PIV experiments, tracer particles scatter light within a fluid,

*These authors contributed equally.

†Corresponding author.

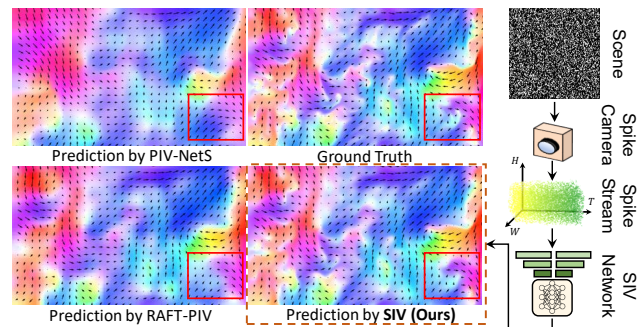


Figure 1: A schematic of fluid velocity field estimation based on spike cameras, and a comparison of the complex velocity field estimated using our proposed SIV network with PIV-NetS (Cai et al. 2019b) and RAFT-PIV (Lagemann et al. 2021).

and their displacements are used to infer the velocity field of the fluid. Recent advancements in optical flow networks (Dosovitskiy et al. 2015; Sun et al. 2018; Teed and Deng 2020) have inspired the development of learning-based PIV algorithms (Cai et al. 2019b,a; Lagemann et al. 2021). While optical flow networks are highly generalizable, fluid scenes present unique challenges that limit further performance improvements. Turbulence, a chaotic flow phenomenon at high Reynolds numbers, introduces specific issues in velocity field estimation: (1) High-speed fluid movement within the field of view. (2) Sparse regions without tracer particles, leading to discrete signals while still requiring the estimation of a continuous flow field. (3) Unstructured patterns and shapes with small-scale vortices and quasi-ordered structures. These challenges coexist in turbulent flows, necessitating advanced experimental hardware and algorithmic solutions. To address these, we propose Spike Imaging Velocimetry (SIV). In terms of hardware, we explore the use of spike cameras (Dong,

Huang, and Tian 2021; Huang 2019) for PIV. Spike cameras offer several advantages for fluid velocity measurement:

- **High Temporal Resolution (HTR):** With a temporal resolution of 20,000 Hz, spike cameras significantly reduce the time interval Δt between consecutive frames, providing rich temporal information that enables precise predictions. This addresses Challenge (1) effectively.

- **High Dynamic Range (HDR):** Spike cameras asynchronously accumulate photons at each pixel, providing a larger dynamic range, which is beneficial in scenarios with uneven lighting or high particle concentration. This helps mitigate limitations seen in conventional cameras, addressing Challenge (2).

- **Reduced Experimental Complexity:** Spike cameras can continuously capture high-speed images with low bandwidth and do not require synchronization controllers, simplifying experimental setups.

We propose a specialized network, **Spike Imaging Velocimetry (SIV)**, for estimating fluid velocity fields. To enhance network performance, we develop three novel modules tailored to the characteristics of fluid dynamics and spike cameras. These include the **Detail-Preserving Hierarchical Transform (DPHT)**, which enhances feature extraction from spike streams while mitigating downsampling artifacts; the **Graph Encoder (GE)**, which maps features into a graph structure for adaptive context aggregation; and the **Multi-scale Velocity Refinement (MSVR)**, which addresses over-smoothing in predicted flow fields. We also introduce a new dataset, **Particle Scenes with Spike and Displacement (PSSD)**, designed for learning fluid motion from spike data. It covers three representative motion estimation scenarios: steady turbulence, high-speed flow, and high dynamic range conditions. Extensive experiments demonstrate the superiority of SIV over both image-based and spike-based baselines. SIV achieves state-of-the-art performance across all evaluation settings. The effectiveness of each proposed module is further validated through ablation studies. The contributions of this work are summarized as follows:

1. We explore a feasible solution for applying spike cameras to fluid velocity measurement, opening new directions for PIV.
2. We design a dedicated network, SIV, tailored to fluid dynamics and turbulent flow properties, and propose three novel modules: the DPHT, GE, and MSVR.
3. We propose a new dataset, PSSD, which captures continuous fluid motion across four flow types and three challenging experimental scenarios.
4. We conduct extensive quantitative experiments and qualitative visual comparisons, demonstrating that SIV achieves state-of-the-art performance across multiple benchmarks. We further perform ablation studies to systematically validate the effectiveness of each proposed module.

Related Work

Spike-based Optical Flow Estimation

The spike camera (Dong, Huang, and Tian 2021; Xiong et al. 2024; Su et al. 2024) is a type of neuromorphic camera

(NeurCam). Inspired by biological visual neural structures and mechanisms, NeurCams emulate the encoding principles of the primate retina. Spike cameras mimic the encoding principles of the central fovea in the retina, providing continuous visual information with ultra-high temporal resolutions, high dynamic range (HDR), and significantly reduced energy consumption and latency, making them well-suited for high-speed motion scenarios. Their pixels independently accumulate photons and emit a spike when the accumulated intensity exceeds a certain threshold. This process maps the incoming photon stream into the spike stream, which preserves spatiotemporal relationships and can be reconstructed into images. For a detailed explanation of the working mechanism, refer to (Dong, Huang, and Tian 2021). The application of spike cameras in optical flow estimation provides significant insights for this study. Although optical flow estimation is generally used to estimate the motion of macroscopic objects, its principles and objectives are aligned with those of PIV. The techniques employed in PIV for image registration and displacement estimation are essentially a variant of optical flow estimation. Spike-based optical flow estimation is still a relatively new field. (Hu et al. 2022) introduced a specialized neural network architecture for spike-based optical flow estimation. This structure can adaptively eliminate motion blur in spike streams based on prior motion information. A framework based on RAFT is proposed by (Zhao et al. 2022) to handle continuous spike streams. (Zhao et al. 2024) demonstrates that using a multi-layer hierarchical structure to characterize spike streams can effectively suppress the influence of randomness in spikes and enhance the feature extraction capability of the network. (Chen, Yu, and Huang 2023) proposed a self-supervised approach to learn optical flow from spike cameras, which is then extended by (Xia et al. 2024).

Image-based PIV Algorithms

Many traditional image-based PIV algorithms (Lourenco and Krothapalli 1995; Nobach and Honkanen 2005) usually involve applying a fast Fourier transform (FFT) between two subregions of image pairs, known as interrogation windows. Displacement is then determined by finding the maximum correlation between the two interrogation windows, which is similar to the correlation method in optical flow estimation to some extent. However, traditional methods depend on finely tuned parameters and complex optimization techniques. With advancements in optical flow networks, various architectures (Dosovitskiy et al. 2015; Sun et al. 2018; Teed and Deng 2020) have been developed to predict dense velocity fields of macroscopic objects with high accuracy. Early PIV networks like PIV-DCNN (Lee, Yang, and Yin 2017) laid the groundwork for these developments. PIV-NetS (Cai et al. 2019b) is capable of generating full-resolution displacement outputs. (Cai et al. 2019a) adapt LiteFlowNet (Hui, Tang, and Loy 2018) to PIV. RAFT-PIV (Lagemann et al. 2021), derived from RAFT (Teed and Deng 2020), has shown exceptional performance on multiple datasets. (Yu et al. 2021) removes the computationally inefficient components of RAFT, such as the context encoder, to improve inference speed. Additionally, unsupervised learning frame-

works for PIV (Zhang and Piggott 2020; Zhang et al. 2023a) have been introduced, reducing reliance on the extremely difficult-to-obtain ground-truth displacements. Other recent PIV networks include LIMA (Manickathan, Mucignat, and Lunati 2023), CC-FCN (Gao et al. 2021), En-FlowNetC (Ji, An, and Hui 2024), ARAFT-FlowNet (Han and Wang 2023), OFVNetS (Ji, Hui, and An 2024), and PIV-PWCNet (Zhang et al. 2023b). However, the performance gains of most PIV networks are largely attributed to advances in optical flow estimation architectures, rather than targeted enhancements tailored to the specific characteristics of fluid dynamics, and most implementations remain unavailable as open-source.

Event-based PIV Algorithms

The event camera (Lichtsteiner, Posch, and Delbruck 2008; Moeys et al. 2017; Posch, Matolin, and Wohlgenannt 2010; Jiang et al. 2024) is another type of NeurCam, which is inspired by the peripheral retina (Rebecq et al. 2019). Unlike spike cameras, event cameras independently monitor the change in light intensity (increase or decrease) at each pixel and generate signals based on the local pixel information of dynamic changes. When the intensity change of a pixel exceeds a preset threshold, the camera records this change event (intensity increase as a positive event, intensity decrease as a negative event) and outputs its timestamp, position, and change value as data. The success of event cameras in optical flow estimation (Lee et al. 2020; Shiba, Aoki, and Gallego 2022) has paved the way for their application to PIV. (Willert and Klinner 2022) introduced the first event-camera PIV algorithm based on correlation and motion compensation. (Willert 2023) combined an event camera with pulsed illumination, substantially reducing estimation latency. These studies are insightful and practical; however, their algorithmic novelty is limited, and robustness is partly constrained by the event camera’s lack of absolute intensity.

Preliminary

The Working Principle of the Spike Camera

Each photosensitive element in a spike camera generates spike streams sequentially over time. A "1" in the stream indicates that a spike has occurred (i.e., the threshold has been reached), while a "0" indicates an ongoing accumulation state. The Integral-and-Fire (IF) mechanism of spike cameras can be mathematically represented as follows:

$$A(\mathbf{x}, t) = \left(\int_0^t \alpha \cdot I(\mathbf{x}, \tau) d\tau \right) \bmod \theta, \quad (1)$$

Where $A(\mathbf{x}, t)$ represents the voltage of the accumulator at pixel $\mathbf{x} = (x, y)$ at time τ , and θ is the threshold of the comparator. The spike camera’s output is a spatiotemporal binary stream S of size $H \times W \times N$, where H and W are the height and width of the sensor, respectively, and N represents the temporal dimension. Compared to traditional cameras, spike cameras preserve richer visual information regarding both temporal resolution and intensity representation.

Data Generation

Training neural networks using supervised strategies requires data with ground truth to optimize model parameters. However, it is difficult to obtain accurate velocity fields from experimental PIV data. We have to generate a synthetic dataset for training, following (Cai et al. 2019b). Our proposed dataset, Particle Scenes with Spike and Displacement (PSSD), includes three sub-datasets: Steady Turbulence (Problem 1), High-speed Flow (Problem 2), and HDR Scenes (Problem 3). Each sub-dataset comprises 10,000 sample sequences of spike streams and the corresponding ground-truth turbulent velocity fields when $\Delta t = 21$ (approximately 1.05 ms) and $\Delta t = 11$ (approximately 0.55 ms). Additionally, we provide corresponding pairs of images for comparative analysis for image-based methods. We use the reference flow field to simulate the motion of the scene and construct a virtual spike camera to trigger the spike stream. Both the images and the spike stream in the dataset include added noise. The reference flow fields are obtained from the Johns Hopkins Turbulence Database (JHTDB) (Li et al. 2008), a high-precision numerical simulation database widely used in turbulence research. JHTDB is based on direct numerical simulation (DNS) of the Navier–Stokes equations without any turbulence modeling, resolving all dynamically relevant scales in both space and time on fine grids. As a result, the simulated flow fields closely approximate real turbulent flows and can be regarded as ground truth in our study.

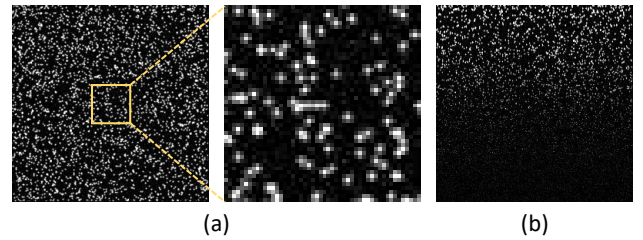


Figure 2: (a) Normal illumination scene (Problem 1, Problem 2); (b) High dynamic range scene (Problem 3).

Approaches

Overall Architecture

Problem definition. Let $\mathbf{S} \in \mathbb{B}^{H \times W \times T}$ represent the spike stream recorded by the spike camera. \mathbb{B} is the binary domain. We aim to estimate the displacement \mathbf{D} of particles in the fluid between time moments t and $t + \Delta t$. At the pixel level, for any given pixel $\mathbf{p} = (x, y)$ at t , we need to find a unique correspondence formulated by:

$$\mathcal{I}(\mathbf{p}, t) \leftarrow \mathcal{I}(\mathbf{p} + \mathbf{D}(\mathbf{p}), t + \Delta t). \quad (2)$$

Here, \mathcal{I} is the mapping of the spike stream \mathbf{S} in the high-dimensional feature space, representing the scene information.

Detail-Preserving Hierarchical Transform

To preserve particle signal information during feature extraction, we propose the Detail-Preserving Hierarchical

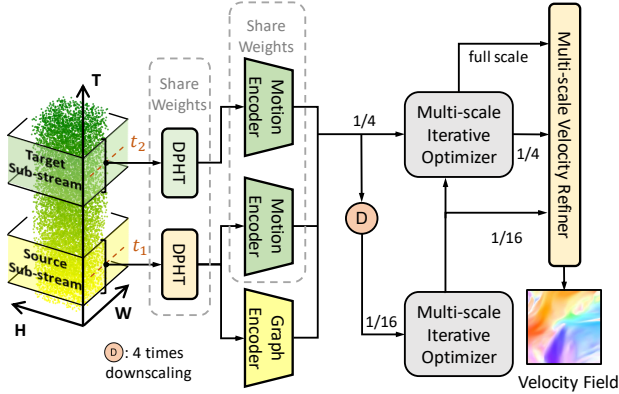


Figure 3: Spike Imaging Velocimetry (SIV) network.

Transform (DPHT), a multi-scale pyramidal architecture that integrates a 3D Detail-Preserving Downsampling module (see Figure 4(a)).

3D Detail-Preserving Downsampling. Given an input feature vector $\mathbf{x}^{(l)}$ at level l , the processing is performed as follows:

$$\mathbf{x}_{\text{out}}^{(l)} = \mathcal{D}^{(l)}(\mathbf{x}^{(l)}) = \mathcal{F}_{2C}(\mathcal{F}_{3C}(\mathbf{x}^{(l)}) + \mathcal{F}_{3MP}(\mathbf{x}^{(l)})), \quad (3)$$

where \mathcal{F}_{3C} is a $5 \times 3 \times 3$ 3D convolution, \mathcal{F}_{3MP} combines 3D max pooling and a $5 \times 1 \times 1$ 3D convolution, and \mathcal{F}_{2C} refers to a stack of 2D convolutions. The output of each level is passed to the next, i.e., $\mathbf{x}^{(l+1)} = \mathbf{x}_{\text{out}}^{(l)}$.

Global Temporal Aggregation. The output feature vectors from each level are refined through additional convolutions and then concatenated to form the final DPHT representation:

$$\mathbf{R} = \mathcal{A}_m[\text{Concat}(\mathbf{x}_{\text{out}}^{(l=1)}, \mathbf{x}_{\text{out}}^{(l=2)}, \mathbf{x}_{\text{out}}^{(l=3)})], \quad (4)$$

where \mathcal{A}_m denotes a convolution operation. The source and target sub-streams' signals are represented as \mathbf{R}_s and \mathbf{R}_t , respectively.

Graph Encoder

In turbulent flows with high Reynolds numbers, large vortices break into smaller ones, revealing an underlying topological structure in fluid motion. To capture this structure, we propose the Graph Encoder (GE), a feature extraction network based on Graph Neural Networks (GNNs) and Convolutional Neural Networks (CNNs). By representing features as graph nodes and constructing a dynamic adjacency matrix, GE aggregates and updates information from neighboring nodes, allowing for a larger receptive field than traditional CNNs. Graph Attention Networks (GATs) (Veličković et al. 2017) leverage attention mechanisms to learn the relative contribution of each neighbor, reducing interference from particle-sparse regions. This combination of GNN and CNN enables adaptive global perception and local refinement, effectively capturing fluid-specific contextual features.

Graph Projection. The first step is to map \mathbf{R}_s from the DPHT to the node space, i.e., $\mathbf{V}_c = \mathcal{P}_{R \rightarrow V}(\mathbf{R}_s)$. The context node representations, $\mathbf{V}_c = \{v_c^1, \dots, v_c^n\} \in \mathbb{R}^{C \times K}$,

contain appearance features related to the scene's shape and regional information, where $C = 128$ is the channel number and K is the number of nodes (set to $K = 128$ in this paper). The projection is formulated as:

$$v_c^i = \mathcal{N}\left(\sum_{\forall j} z_{ij} \cdot \mathbf{R}_{s,j}\right), \quad (5)$$

where $\mathcal{N}(\cdot)$ is the L_2 normalization, and z_{ij} denotes the normalized projection weights computed by a 1×1 convolution and a softmax function.

Graph Convolution Layer. After obtaining the nodes \mathbf{v}_c , the dynamic adjacency matrix $\tilde{\mathbf{A}} = \mathbf{V}_c^T \mathbf{V}_c$ is computed based on the similarity between node vectors. The operations of two GAT modules, \mathcal{G}_1 and \mathcal{G}_2 , update the node features as follows:

$$\mathbf{V}_c^{\text{out}} = \mathcal{G}_2[\mathcal{G}_1(\mathbf{V}_c, \tilde{\mathbf{A}}) + \mathbf{V}_c, \tilde{\mathbf{A}}] + \mathcal{G}_1(\mathbf{V}_c, \tilde{\mathbf{A}}) + \mathbf{V}_c, \quad (6)$$

Graph Re-projection. The re-projection operation $\mathcal{P}_{V \rightarrow R}(\cdot)$ takes the transformed vertex features $\mathbf{V}_c^{\text{out}}$ and the projection weight matrix \mathbf{Z} , producing a 2D feature map $\tilde{\mathbf{R}}_c$:

$$\tilde{\mathbf{R}}_c = \mathcal{P}_{V \rightarrow R}(\mathbf{V}_c^{\text{out}}, \mathbf{Z}) = \mathbf{Z}(\mathbf{V}_c^{\text{out}})^T. \quad (7)$$

This operation preserves spatial details, as even if two pixels share the same node, they will have distinct features. Finally, these results are integrated into a residual extractor network \mathcal{R} for local refinement:

$$\mathbf{F}_c = \mathcal{R}(\mathbf{R}_s + \alpha \tilde{\mathbf{R}}_c), \quad (8)$$

where α is a learnable parameter initialized to 0, gradually refining the feature sum.

Multi-scale Iterative Optimizer

We adopt the Multi-scale Iterative Optimizer (MSIO) from DIP (Zheng et al. 2022) as the backbone for iterative residual flow estimation in SIV. Unlike RAFT's full correlation volume construction, MSIO leverages Inverse PatchMatch to significantly reduce the computational cost introduced by large-scale feature maps (1/4 resolution of the full size). The detailed architecture of MSIO can be found in (Zheng et al. 2022); here, we denote it implicitly as \mathcal{U} . The output multi-scale velocity field is given by

$$\{\mathbf{u}_i\}_{i=1}^N = \mathcal{U}(\mathbf{F}_c, \mathbf{F}_m), \quad (9)$$

where \mathbf{F}_m denotes motion features extracted by the Motion Encoder, and N refers to the number of GRU iterations (Teed and Deng 2020).

Multi-scale Velocity Refinement

DIP adopts the same compact and effective convex upsampling strategy as RAFT to recover full-resolution optical flow. However, such upsampling is insufficient for accurately reconstructing the velocity field of fluids. To address this, we propose a Multi-scale Velocity Refinement (MSVR) module to assist in reconstructing the complete velocity field, including fine-scale vortices. As illustrated in Figure 5, we first apply a sequence of 2D convolutions followed by activation functions to process the multi-scale velocity fields

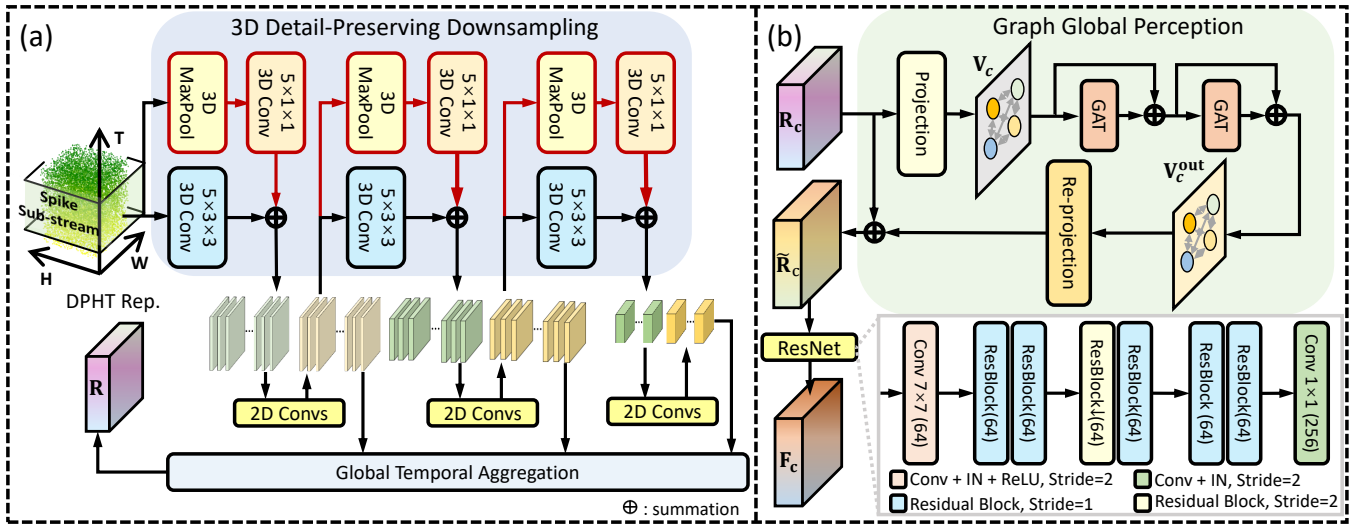


Figure 4: (a) Detail-Preserving Hierarchical Transform (DPHT) Module. (b) Graph Encoder (GE).

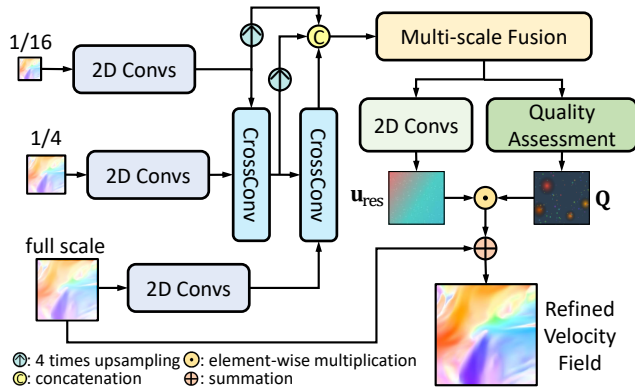


Figure 5: Multi-scale Velocity Refinement Module.

generated by MSIO. Information exchange across scales is achieved via cross convolution from coarse to fine levels. The extracted features are then fused and processed by distinct convolutional networks to generate a residual velocity field \mathbf{u}_{res} and a quality map \mathbf{Q} . The refined fluid velocity field is given by,

$$\mathbf{u}_{\text{ref}} = \mathbf{u}_N + \mathbf{u}_{\text{res}} \odot \mathbf{Q}, \quad (10)$$

where \odot denotes element-wise multiplication.

Loss Function

The loss function for the proposed network is composed of two parts: the flow loss and the gradient loss. Suppose the recurrent optimizer of the network has N iterations, and the estimated flow fields of each iteration are $\{\mathbf{u}_1, \dots, \mathbf{u}_N, \mathbf{u}_{N+1}\}$, $\mathbf{u}_{N+1} = \mathbf{u}_{\text{ref}}$. The flow loss is defined as:

$$L_{\text{flow}} = \sum_{i=1}^{N+1} \gamma^{N+1-i} \cdot \|\mathbf{u}_i - \mathbf{u}_{\text{gt}}\|_1, \quad (11)$$

\mathbf{u}_{gt} is the ground truth of velocity field. To preserve small vortical structures, we introduce a gradient loss:

$$L_{\text{grad}} = \sum_{i=1}^{N+1} \gamma^{N+1-i} \cdot (\|\nabla_x \mathbf{u}_i - \nabla_x \mathbf{u}_{\text{gt}}\|_1 + \|\nabla_y \mathbf{u}_i - \nabla_y \mathbf{u}_{\text{gt}}\|_1). \quad (12)$$

The total loss is:

$$L = L_{\text{flow}} + \beta \cdot L_{\text{grad}}. \quad (13)$$

γ is set to 0.8, and β is set to 0.3.

Experiments

We evaluate our method on the PSSD dataset and compare its performance with both traditional and learning-based approaches. Additionally, we conduct a series of ablation studies to assess the effectiveness of the proposed modules.

Experimental Details

SIV is implemented using the PyTorch framework on a single NVIDIA RTX 2080 Ti GPU. We use the Adam optimizer with a warm-up strategy during training. For each sub-dataset, the model is trained for 100 epochs with a batch size of 4 and an initial learning rate of 10^{-4} , applying an 85% decay every 10 epochs. The number of iterative refinements for the GRUs is set to 12. Following (Cai et al. 2019b) and (Lagemann et al. 2021), we use endpoint error (EPE) as the primary evaluation metric for comparing different algorithms. The EPE is the spatially mean value of Euclidean distance between the predicted flow \mathbf{u}_{pred} and its ground truth \mathbf{u}_{gt} .

Comparative Experiments

We compare our method with several open-source baselines, including classical PIV algorithms (Cai et al. 2018), image-based PIV networks (PIV-NetS (Cai et al. 2019b), RAFT-PIV (Lagemann et al. 2021)), spike-adapted optical flow

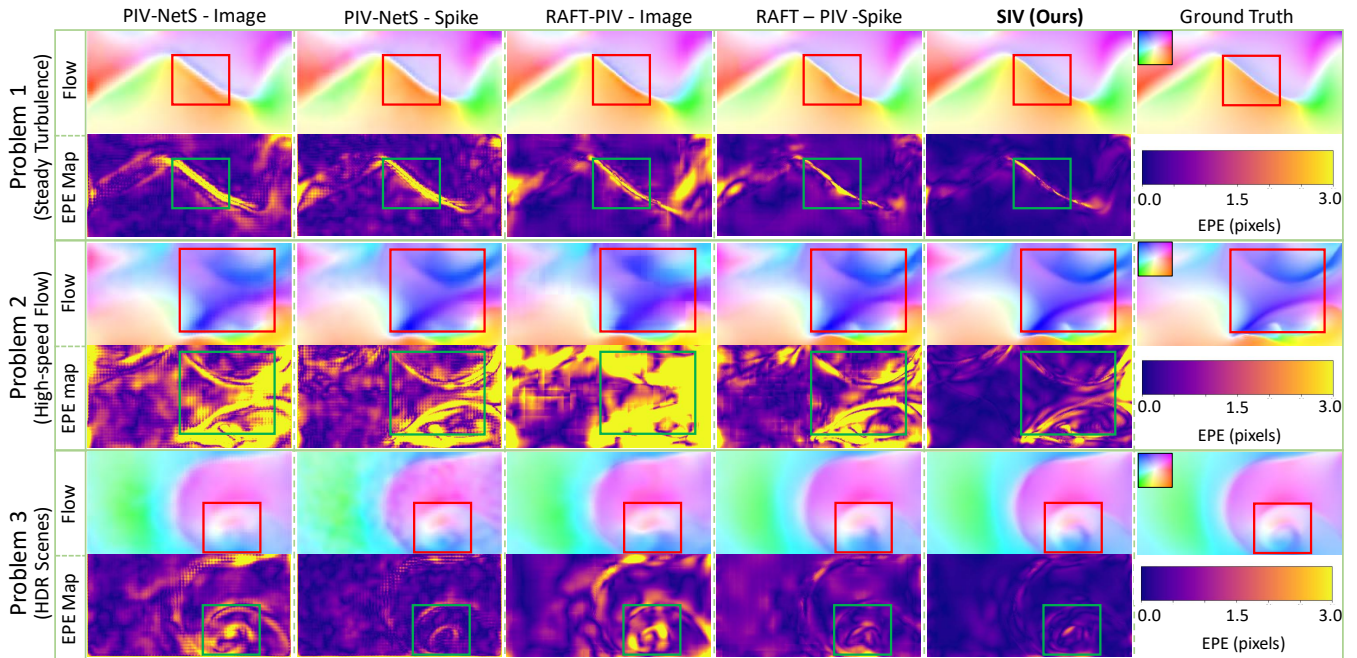


Figure 6: Visualization examples of different algorithms in Steady Turbulence (Problem 1), High-speed Flow (Problem 2), and HDR Scenes (Problem 3) are shown. For each problem, the red box in the top optical flow map and the green box in the bottom EPE error map indicate the same region of the flow field.

Methods	Channel		Isotropic		Mhd		Mixing		Average	
	$\Delta t=21$	$\Delta t=11$	$\Delta t=21$	$\Delta t=11$	$\Delta t=21$	$\Delta t=11$	$\Delta t=21$	$\Delta t=11$	$\Delta t=21$	$\Delta t=11$
(A) coarse-to-fine HS	6.790	0.509	7.415	1.998	7.683	2.558	9.388	1.954	7.819	1.755
(B) RAFT-PIV-Image	0.377	0.215	2.181	1.252	1.336	0.842	1.875	1.064	1.442	0.843
(C) RAFT-PIV-Spike	0.347	0.208	1.223	0.755	0.857	0.557	0.958	0.581	0.846	0.525
(D) PIV-NetS-Image	0.545	0.301	2.560	1.497	1.574	0.985	2.247	1.320	1.731	1.026
(E) PIV-NetS-Spike	0.553	0.289	1.507	0.874	1.252	0.734	1.382	0.782	1.173	0.670
(F) SCFlow	0.354	0.222	2.232	1.305	1.374	0.901	2.132	1.169	1.522	0.899
(G) HiST-SFlow	0.369	0.224	1.271	0.796	0.964	0.611	1.029	0.637	0.908	0.567
(H) Flowformer-Spike	0.286	0.147	1.069	0.642	0.727	0.437	0.806	0.472	0.722	0.425
(I) SIV (ours)	0.239	0.159	0.896	0.587	0.613	0.422	0.680	0.441	0.607	0.402

Table 1: EPE (\downarrow) results on PSSD dataset - Problem 1. The best results are marked in bold.

networks (e.g., Flowformer with spike input (Huang et al. 2022)), and spike-based optical flow networks (SCFlow (Hu et al. 2022), HiST-SFlow (Zhao et al. 2024)). All networks trained on spikes (incl. (C)(E)(F)(G)(H)(I) in Table 1, 2, and 3) use **voxel-grid inputs** to ensure a fair evaluation. All image baselines (incl. (A)(B)(D) in Table 1, 2, and 3) use sharp frames containing only artificially added noise (as shown in Figure 2, **not spike reconstructions**) simulating high-speed cameras to highlight the differences between image and spike inputs. The image is temporally aligned with the intermediate moments of the spike stream.

As shown in Tables 1, 2, and 3, $\Delta t = 21$ and $\Delta t = 11$ represent the time intervals between two spike streams used for velocity field prediction. Our SIV method achieves the lowest average EPE across all sub-datasets. Compared to baseline approaches, SIV demonstrates superior prediction accuracy in both low-velocity (Problem 1) and high-velocity

(Problem 2) scenarios. Additionally, in high dynamic range conditions (Problem 3), SIV effectively leverages the advantages of spike-based input modalities, addressing the fundamental limitations of conventional frame-based systems.

Figure 6 visualizes the predictions on three problems. The top row visualizes the velocity fields using the standard optical flow color coding (Baker et al. 2011), and the bottom row shows the EPE. Since shear layers are crucial for turbulence generation and development, it is noteworthy that SIV attains the lowest prediction error in both shear layers and small-scale vortices, while avoiding noticeable blurring, distortion, or discontinuities.

Ablation Study

Ablation of modules. All values in Table 4 and 5 are the arithmetic mean across scenes ($\Delta t = 21$), with the best results highlighted in bold. The results emphasize the

Methods	Channel		Isotropic		Mhd		Mixing		Average	
	$\Delta t=21$	$\Delta t=11$	$\Delta t=21$	$\Delta t=11$	$\Delta t=21$	$\Delta t=11$	$\Delta t=21$	$\Delta t=11$	$\Delta t=21$	$\Delta t=11$
(A) coarse-to-fine HS	20.691	8.244	17.164	8.051	17.469	8.170	17.131	7.936	18.114	8.100
(B) RAFT-PIV-Image	0.665	0.368	4.730	2.635	2.320	1.366	4.305	2.368	3.005	1.684
(C) RAFT-PIV-Spike	0.688	0.412	2.410	1.483	1.468	0.959	2.071	1.241	1.659	1.024
(D) PIV-NetS-Image	1.163	0.683	4.866	2.830	2.673	1.643	4.347	2.500	3.262	1.914
(E) PIV-NetS-Spike	0.887	0.540	3.172	1.818	2.392	1.410	3.067	1.744	2.379	1.378
(F) SCFlow	0.694	0.403	4.689	2.613	2.387	1.391	4.326	2.380	3.024	1.696
(G) HiST-SFlow	0.556	0.364	2.212	1.365	1.404	0.998	1.911	1.179	1.521	0.976
(H) Flowformer-Spike	0.533	0.284	1.975	1.179	1.214	0.748	1.668	0.972	1.348	0.796
(I) SIV (ours)	0.370	0.254	1.687	1.080	1.071	0.722	1.449	0.912	1.144	0.742

Table 2: EPE (\downarrow) results on PSSD dataset - Problem 2. The best results are marked in bold.

Methods	Channel		Isotropic		Mhd		Mixing		Average	
	$\Delta t=21$	$\Delta t=11$	$\Delta t=21$	$\Delta t=11$	$\Delta t=21$	$\Delta t=11$	$\Delta t=21$	$\Delta t=11$	$\Delta t=21$	$\Delta t=11$
(A) coarse-to-fine HS	7.908	1.044	7.924	2.706	8.345	3.130	7.854	2.716	8.008	2.399
(B) RAFT-PIV-Image	0.447	0.244	2.694	1.495	1.447	0.888	2.311	1.271	1.725	0.974
(C) RAFT-PIV-Spike	0.320	0.200	1.203	0.740	0.798	0.530	0.946	0.576	0.817	0.512
(D) PIV-NetS-Image	0.667	0.343	2.786	1.602	1.641	0.956	2.393	1.378	1.872	1.070
(E) PIV-NetS-Spike	0.476	0.247	1.420	0.856	1.157	0.716	1.342	0.774	1.099	0.648
(F) SCFlow	0.462	0.269	1.850	0.837	1.300	0.774	1.972	0.769	1.396	0.662
(G) HiST-SFlow	0.329	0.180	1.179	0.640	0.832	0.452	0.960	0.476	0.825	0.437
(H) Flowformer-Spike	0.269	0.145	1.052	0.620	0.716	0.429	0.813	0.481	0.713	0.419
(I) SIV (ours)	0.234	0.155	0.880	0.573	0.598	0.404	0.677	0.438	0.598	0.392

Table 3: EPE (\downarrow) results on PSSD dataset - Problem 3. The best results are marked in bold.

effectiveness of DPHT, GE, MSVR, and L_{grad} . In Table 4, comparisons between experiments (A) and (B) show the effectiveness of DPHT, while comparisons between (B) and (C), and (D) and (E), validate the contribution of the Graph Encoder (GE). Additionally, comparisons between (B) and (D), and (C) and (E), highlight the impact of MSVR.

Ablation of β . Table 5 summarizes the ablation study on the weighting coefficient β in the proposed loss function. The results show that $\beta = 0.3$ yields the best performance for the current model and dataset.

#	Settings			Problem 1 $\Delta t=21$	Problem 2 $\Delta t=21$	Problem 3 $\Delta t=21$
	DPHT	GE	MSVR			
(A)	×	×	×	0.805	1.450	0.776
(B)	✓	×	×	0.701	1.299	0.659
(C)	✓	✓	×	0.625	1.167	0.616
(D)	✓	×	✓	0.644	1.268	0.635
(E)	✓	✓	✓	0.607	1.144	0.598

Table 4: Ablation study of proposed modules, showing results for EPE (\downarrow). ✓ indicates module presence, × indicates absence. Setting (E) is our final model.

Conclusion

We propose a novel approach for fluid motion estimation using spike cameras and introduce a specialized network architecture, Spike Imaging Velocimetry (SIV), designed to address PIV tasks while adapting to the specific characteristics of fluid scenes. We develop the Detail-Preserving Hier-

#	β	Problem 1	Problem 2	Problem 3
(A)	0.0	0.620	1.169	0.615
(B)	0.1	0.617	1.163	0.605
(C)	0.2	0.615	1.157	0.603
(D)	0.3	0.607	1.144	0.598
(E)	0.4	0.613	1.146	0.610
(F)	0.5	0.619	1.141	0.617

Table 5: Ablation study of β , showing results for EPE (\downarrow).

archical Transform (DPHT) module to preserve particle signal information during spike stream preprocessing, a Graph Encoder (GE) to extract contextual features from highly unstructured and turbulent flows, and a Multi-scale Velocity Refinement (MSVR) module to refine the velocity field. We also introduce the challenging and realistic Particle Scenes with Spike and Displacement (PSSD) dataset. Experimental results demonstrate that SIV achieves state-of-the-art performance on the PSSD dataset, accurately estimating velocities in small-scale yet critical flow structures such as shear layers and vortices. In summary, this study highlights the potential of spike cameras for fluid velocity measurement and offers new perspectives and directions for PIV network design.

This paper aims to demonstrate that spike cameras are a promising tool for PIV estimation. However, given the complexity of fluid flows, there remains substantial room for improvement: (i) although we have generated a large dataset, its scale and diversity should be further increased; (ii) noise should be modeled more comprehensively to achieve stronger performance on experimental data; and (iii) improving inference efficiency to enable real-time prediction would benefit both laboratory and industrial applications.

Acknowledgements

This work is supported by the Postgraduate Research & Practice Innovation Program of Jiangsu Province (No. KYCX25_0173), the National Natural Science Foundation of China under Grants 62371006, 62302016, U24B20140, the Jiangsu Association for Science and Technology (JAST) Young Elite Scientists Sponsorship Program (No. JSTJ-2025-730), and the Beijing Natural Science Foundation (No. 3242008).

References

- Baker, S.; Scharstein, D.; Lewis, J. P.; Roth, S.; Black, M. J.; and Szeliski, R. 2011. A database and evaluation methodology for optical flow. *International journal of computer vision*, 92: 1–31.
- Cai, S.; Liang, J.; Gao, Q.; Xu, C.; and Wei, R. 2019a. Particle image velocimetry based on a deep learning motion estimator. *IEEE Transactions on Instrumentation and Measurement*, 69(6): 3538–3554.
- Cai, S.; Mémin, E.; Dérian, P.; and Xu, C. 2018. Motion estimation under location uncertainty for turbulent fluid flows. *Experiments in Fluids*, 59(1): 8.
- Cai, S.; Zhou, S.; Xu, C.; and Gao, Q. 2019b. Dense motion estimation of particle images via a convolutional neural network. *Experiments in Fluids*, 60: 1–16.
- Chen, B.; Li, X.; and Zhu, Z. 2022. Investigations of energy distribution and loss characterization in a centrifugal impeller through PIV experiment. *Ocean Engineering*, 247: 110773.
- Chen, S.; Yu, Z.; and Huang, T. 2023. Self-supervised joint dynamic scene reconstruction and optical flow estimation for spiking camera. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 350–358.
- Dong, S.; Huang, T.; and Tian, Y. 2021. Spike camera and its coding methods. *arXiv preprint arXiv:2104.04669*.
- Dosovitskiy, A.; Fischer, P.; Ilg, E.; Hausser, P.; Hazirbas, C.; Golkov, V.; Van Der Smagt, P.; Cremers, D.; and Brox, T. 2015. FlowNet: Learning optical flow with convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, 2758–2766.
- Gao, Q.; Lin, H.; Tu, H.; Zhu, H.; Wei, R.; Zhang, G.; and Shao, X. 2021. A robust single-pixel particle image velocimetry based on fully convolutional networks with cross-correlation embedded. *Physics of Fluids*, 33(12).
- Han, Y.; and Wang, Q. 2023. An attention-mechanism incorporated deep recurrent optical flow network for particle image velocimetry. *Physics of Fluids*, 35(7).
- Hu, L.; Zhao, R.; Ding, Z.; Ma, L.; Shi, B.; Xiong, R.; and Huang, T. 2022. Optical flow estimation for spiking camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17844–17853.
- Huang, T. 2019. Method and device for encoding space-time signals. US Patent 10,523,972.
- Huang, Z.; Shi, X.; Zhang, C.; Wang, Q.; Cheung, K. C.; Qin, H.; Dai, J.; and Li, H. 2022. Flowformer: A transformer architecture for optical flow. In *European conference on computer vision*, 668–685. Springer.
- Hui, T.-W.; Tang, X.; and Loy, C. C. 2018. LiteflowNet: A lightweight convolutional neural network for optical flow estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 8981–8989.
- Ji, K.; An, Q.; and Hui, X. 2024. Cross-correlation-based convolutional neural network with velocity regularization for high-resolution velocimetry of particle images. *Physics of Fluids*, 36(7).
- Ji, K.; Hui, X.; and An, Q. 2024. Effects of velocity regularization on neural network performance in processing particle images. *Physics of Fluids*, 36(10).
- Jiang, B.; Xiong, B.; Qu, B.; Salman Asif, M.; Zhou, Y.; and Ma, Z. 2024. Edformer: Transformer-based event denoising across varied noise levels. In *European Conference on Computer Vision*, 200–216. Springer.
- Kompenhans, J.; Raffel, M.; Dieterle, L.; Dewhurst, T.; Vollmers, H.; Ehrenfried, K.; Willert, C.; Pengel, K.; Kähler, C.; Schröder, A.; et al. 2000. Particle image velocimetry in aerodynamics: Technology and applications in wind tunnels. *Journal of Visualization*, 2(3-4): 229–244.
- Lagemann, C.; Lagemann, K.; Mukherjee, S.; and Schröder, W. 2021. Deep recurrent optical flow learning for particle image velocimetry data. *Nature Machine Intelligence*, 3(7): 641–651.
- Lee, C.; Kosta, A. K.; Zhu, A. Z.; Chaney, K.; Daniilidis, K.; and Roy, K. 2020. Spike-flowNet: event-based optical flow estimation with energy-efficient hybrid neural networks. In *European conference on computer vision*, 366–382. Springer.
- Lee, Y.; Yang, H.; and Yin, Z. 2017. PIV-DCNN: cascaded deep convolutional neural networks for particle image velocimetry. *Experiments in Fluids*, 58: 1–10.
- Li, Y.; Perlman, E.; Wan, M.; Yang, Y.; Meneveau, C.; Burns, R.; Chen, S.; Szalay, A.; and Eyink, G. 2008. A public turbulence database cluster and applications to study Lagrangian evolution of velocity increments in turbulence. *Journal of Turbulence*, 9(N31).
- Lichtsteiner, P.; Posch, C.; and Delbruck, T. 2008. A 128×128 120 dB 15 μs Latency Asynchronous Temporal Contrast Vision Sensor. *IEEE Journal of Solid-State Circuits*, 43(2): 566–576.
- Lourenco, L.; and Krothapalli, A. 1995. On the accuracy of velocity and vorticity measurements with PIV. *Experiments in fluids*, 18(6): 421–428.
- Manickathan, L.; Mucignat, C.; and Lunati, I. 2023. A lightweight neural network designed for fluid velocimetry. *Experiments in Fluids*, 64(10): 161.
- MoayeriKashani, M.; Lai, S.; Ibrahim, S.; Sulaiman, N.; and Teo, F. 2016. Tracking the hydrodynamic behavior of fine sediment using particle image velocimetry. *Environmental Earth Sciences*, 75: 1–13.
- Moeys, D. P.; Corradi, F.; Li, C.; Bamford, S. A.; Longinotti, L.; Voigt, F. F.; Berry, S.; Taverni, G.; Helmchen, F.; and Delbruck, T. 2017. A sensitive dynamic and active pixel vision sensor for color or neural imaging applications. *IEEE transactions on biomedical circuits and systems*, 12(1): 123–136.

- Nakagawa, M.; Kallweit, S.; Michaux, F.; and Hojo, T. 2016. Typical velocity fields and vortical structures around a formula one car, based on experimental investigations using particle image velocimetry. *SAE International Journal of Passenger Cars-Mechanical Systems*, 9(2016-01-1611): 754–771.
- Nobach, H.; and Honkanen, M. 2005. Two-dimensional Gaussian regression for sub-pixel displacement estimation in particle image velocimetry or particle position estimation in particle tracking velocimetry. *Experiments in fluids*, 38: 511–515.
- Posch, C.; Matolin, D.; and Wohlgenannt, R. 2010. A QVGA 143 dB dynamic range frame-free PWM image sensor with lossless pixel-level video compression and time-domain CDS. *IEEE Journal of Solid-State Circuits*, 46(1): 259–275.
- Raschi, M.; Mut, F.; Byrne, G.; Putman, C. M.; Tatehima, S.; Viñuela, F.; Tanoue, T.; Tanishita, K.; and Cebal, J. R. 2012. CFD and PIV analysis of hemodynamics in a growing intracranial aneurysm. *International journal for numerical methods in biomedical engineering*, 28(2): 214–228.
- Rebecq, H.; Ranftl, R.; Koltun, V.; and Scaramuzza, D. 2019. High speed and high dynamic range video with an event camera. *IEEE transactions on pattern analysis and machine intelligence*, 43(6): 1964–1980.
- Shiba, S.; Aoki, Y.; and Gallego, G. 2022. Secrets of event-based optical flow. In *European Conference on Computer Vision*, 628–645. Springer.
- Su, C.; Ye, Z.; Xiao, Y.; Zhou, Y.; Cheng, Z.; Xiong, B.; Yu, Z.; and Huang, T. 2024. Intensity-robust autofocus for spike camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 25018–25027.
- Sun, D.; Yang, X.; Liu, M.-Y.; and Kautz, J. 2018. Pwcnnet: Cnns for optical flow using pyramid, warping, and cost volume. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 8934–8943.
- Teed, Z.; and Deng, J. 2020. Raft: Recurrent all-pairs field transforms for optical flow. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, 402–419. Springer.
- Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; and Bengio, Y. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903*.
- Willert, C. E. 2023. Event-based imaging velocimetry using pulsed illumination. *Experiments in Fluids*, 64(5): 98.
- Willert, C. E.; and Klinner, J. 2022. Event-based imaging velocimetry: an assessment of event-based cameras for the measurement of fluid flows. *Experiments in Fluids*, 63(6): 101.
- Xia, L.; Ding, Z.; Zhao, R.; Zhang, J.; Ma, L.; Yu, Z.; Huang, T.; and Xiong, R. 2024. Unsupervised optical flow estimation with dynamic timing representation for spike camera. *Advances in Neural Information Processing Systems*, 36.
- Xiong, B.; Su, C.; Lin, Z.; Chen, Y.; Zhou, Y.; Cheng, Z.; Yu, Z.; and Huang, T. 2024. Real-time parameter evaluation of high-speed microfluidic droplets using continuous spike streams. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 6833–6841.
- Yu, C.; Bi, X.; Fan, Y.; Han, Y.; and Kuai, Y. 2021. Light-PIVNet: An effective convolutional neural network for particle image velocimetry. *IEEE Transactions on Instrumentation and Measurement*, 70: 1–15.
- Zhang, M.; and Piggott, M. D. 2020. Unsupervised learning of particle image velocimetry. In *High Performance Computing: ISC High Performance 2020 International Workshops, Frankfurt, Germany, June 21–25, 2020, Revised Selected Papers 35*, 102–115. Springer.
- Zhang, W.; Dong, X.; Sun, Z.; and Xu, S. 2023a. An unsupervised deep learning model for dense velocity field reconstruction in particle image velocimetry (PIV) measurements. *Physics of Fluids*, 35(7).
- Zhang, W.; Nie, X.; Dong, X.; and Sun, Z. 2023b. Pyramidal deep-learning network for dense velocity field reconstruction in particle image velocimetry. *Experiments in Fluids*, 64(1): 12.
- Zhao, R.; Xiong, R.; Zhang, J.; Zhang, X.; Yu, Z.; and Huang, T. 2024. Optical Flow for Spike Camera with Hierarchical Spatial-Temporal Spike Fusion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 7496–7504.
- Zhao, R.; Xiong, R.; Zhao, J.; Yu, Z.; Fan, X.; and Huang, T. 2022. Learning optical flow from continuous spike streams. *Advances in Neural Information Processing Systems*, 35: 7905–7920.
- Zheng, Z.; Nie, N.; Ling, Z.; Xiong, P.; Liu, J.; Wang, H.; and Li, J. 2022. Dip: Deep inverse patchmatch for high-resolution optical flow. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8925–8934.