DOI: 10.1111/cgf.70274
Pacific Graphics 2025
M. Christie, N. Pietroni, and Y.-S. Wang
(Guest Editors)

COMPUTER GRAPHICS *forum*
Volume 44 (2025), Number 7

FlowCapX: Physics-Grounded Flow Capture with Long-Term Consistency

N. Tao[1], L. Zhang[2], X. Ni[3], M. Chu[†1] and B. Chen[†1]

[1] School of IST & State Key Laboratory of General Artificial Intelligence, Peking University, China
[2] School of Electronics Engineering and Computer Science, Peking University, China
[3] School of CS & State Key Laboratory of General Artificial Intelligence, Peking University, China
[†] Corresponding authors

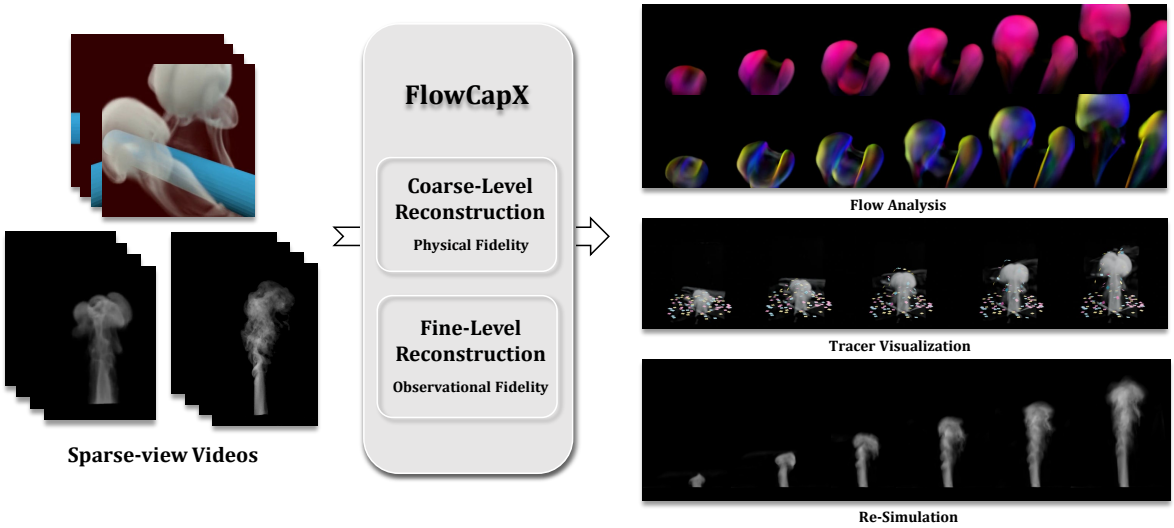

Figure 1: *FlowCapX enables high-fidelity flow reconstruction from sparse video inputs, supporting downstream tasks including (1) Velocity-based flow analysis, (2) Robust scene augmentation with tracer visualization, and (3) Accurate re-simulation via reconstructed velocity.*

Abstract

*We present **FlowCapX**, a physics-enhanced framework for flow reconstruction from sparse video inputs, addressing the challenge of jointly optimizing complex physical constraints and sparse observational data over long time horizons. Existing methods often struggle to capture turbulent motion while maintaining physical consistency, limiting reconstruction quality and downstream tasks. Focusing on velocity inference, our approach introduces a hybrid framework that strategically separates representation and supervision across spatial scales. At the coarse level, we resolve sparse-view ambiguities via a novel optimization strategy that aligns long-term observation with physics-grounded velocity fields. By emphasizing vorticity-based physical constraints, our method enhances physical fidelity and improves optimization stability. At the fine level, we prioritize observational fidelity to preserve critical turbulent structures. Extensive experiments demonstrate state-of-the-art velocity reconstruction, enabling velocity-aware downstream tasks, e.g., accurate flow analysis, scene augmentation with tracer visualization and re-simulation. Our implementation is released at* `https://github.com/taoningxiao/FlowCapX.git`.

CCS Concepts

• ***Computing methodologies*** → ***Physical simulation; Neural networks;***

1 Introduction

Accurate reconstruction of turbulent flows from sparse-view RGB videos remains a pivotal challenge, with critical implications for

applications ranging from aerodynamic analysis [RCDB23] to visual effects [GITH14]. While recent advances in neural reconstruction have significantly improved density and appearance recovery [WTC24; YZG*24; CLZ*22], progress in *physically consistent velocity estimation* remains insufficient, hindering reliable analysis and applications.

One major challenge for velocity reconstruction is the inherent ambiguity in reconstructing turbulent motion from sparse observations. Prior work in experimental settings with known lighting conditions [EUT19; FST21] has shown that long temporal physical consistency is critical to resolve this ambiguity, as it establishes additional temporal correspondences across frames. However, jointly optimizing neural velocity representations across frames to enforce long-term consistency is challenging and often fails to achieve low-error solutions. Neural trajectory representations [WTC24], by contrast, inherently encode temporal correspondence through their formulation. However, their over-constrained representation space tends to filter out essential turbulent phenomena, such as vortex shedding or small-scale eddies. Another challenge stems from the trade-off between enforcing physical laws and maintaining observational fidelity. Strictly enforcing physical laws often leads to the over-smoothing of turbulent details [CLZ*22], while observation-driven methods struggle to maintain physical plausibility [DWD*24]. Optimizing both physical laws and observational data is a known difficulty in methods such as Physics Informed Neural Networks [CDG*22; LCT25]. This compromise leads to sub-optimal velocity estimation that lacks the fidelity required for precise analysis and the robustness necessary for reliable downstream tasks, such as tracer visualization or re-simulation.

To address these challenges, we propose a hybrid framework that strategically splits representation and supervision across spatial scales. At the fine scale, where turbulence is highly complex, we prioritize observational fidelity over strict physical constraints to preserve the critical turbulent characteristics. On the coarse scale, we enforce long temporal physical consistency through an innovative flow transport supervision, which resolves temporal ambiguities by combining multi-frame observation cues and efficiently penalizing accumulated drift rather than only frame-by-frame error. We further complemented it with vorticity-based physical constraints, ensuring better optimization convergence and vorticity preservation. This separation allows the coarse scale to be optimized for physical correctness and convergence stability—unaffected by small-scale turbulence—while the fine scale focuses on capturing high-frequency observational detail only within the physically valid regions defined by the coarse velocity. As a result, by merging the coarse and fine scales, our method yields a velocity field that is accurate, robust, and faithfully turbulent. We further demonstrate the effectiveness of our approach through extensive evaluations, focusing on both reconstruction accuracy and downstream tasks including tracer visualization and re-simulation. Our key contributions are summarized as follows:

- **Hybrid framework:** A spatially split strategy that combines fine-scale observational fidelity with coarse-scale physical consistency.
- **Physical fidelity:** Robust velocity estimation with enhanced long-term consistency and convergence stability via flow transport and vorticity constraints.
- **Downstream task supports:** Improved velocity reconstructions for accurate analysis, tracer visualization and re-simulation, benefiting various physics-based applications.

2 Related Work

In fluid reconstruction, recent advancements have shifted from active sensing with specialized hardware—such as structured light systems [GNG*13; JYY13] and particle imaging velocimetry [Gra97; ESWvO06]—to RGB-video-based techniques [EUT19] and implicit neural representations [CLZ*22].

Many recent methods have achieved impressive results in reconstructing the visual appearance of dynamic phenomena. For instance, Zeng et al. [ZBY*24] proposed an encoder–decoder framework for real-time acquisition and high-quality reconstruction of temporally varying 3D scenes, while Qiu et al. [QCL*24] combined 3D neural transportation fields with 2D CNN-based detail refinement to efficiently reconstruct smoke from multi-view videos. Although these approaches excel at producing visually compelling reconstructions, they give minimal or no emphasis on enforcing the physical constraints imposed by the Navier–Stokes equations.

In constrast, methods like GlobalTrans [FST21] and PICT [WTC24] have made significant strides in robust motion estimation by integrating long-term supervision and enforcing physical consistency. GlobalTrans achieves this through differentiable rendering combined with physics, albeit at the cost of requiring known geometry and lighting conditions, whereas PICT employs a long-term trajectory representation that, while effective, is less adept at modeling turbulence.

In forward fluid simulations, various strategies have been developed to accurately solve the Navier–Stokes equations. Flow map methods [TP11; QZG*19; NWRC22; DYZ*23] maintain the spatiotemporal trajectories of fluid particles, delivering long-term consistency with reduced numerical vorticity dissipation. Vortex methods [Cot00] preserve vortex energy by solving the vorticity formulation, ensuring accurate representation of rotational dynamics. Additionally, frequency-decomposed methods [KTJG08] enhance turbulent synthesis by simulating low-frequency components at coarse resolutions and subsequently integrating high-frequency details through numerical procedures, thereby compensating for truncation errors. Drawing inspiration from these forward simulation techniques, we integrate these principles to achieve high-fidelity velocity estimation in a reconstruction pipeline.

3 Preliminaries

Neural Representations. Taking multi-view images as inputs, NeRF [MST*20] trains a network $\mathcal{F}(\boldsymbol{x}) = (c, \sigma)$ for scene reconstruction, where $\boldsymbol{x}$, c, and σ denote the spatial position, radiance color, and radiance density, respectively. According to the volumetric rendering formulation, the color C for each pixel is computed by sampling n points along the ray cast from the camera as follows:

$$C = \sum_{i=1}^{n} T_i \left(1 - \mathrm{e}^{-\sigma_i \delta_i}\right) c_i, \; T_i = \mathrm{e}^{-\sum_{j=1}^{i-1} \sigma_j \delta_j}, \; \delta_j = h_{j+1} - h_j,$$

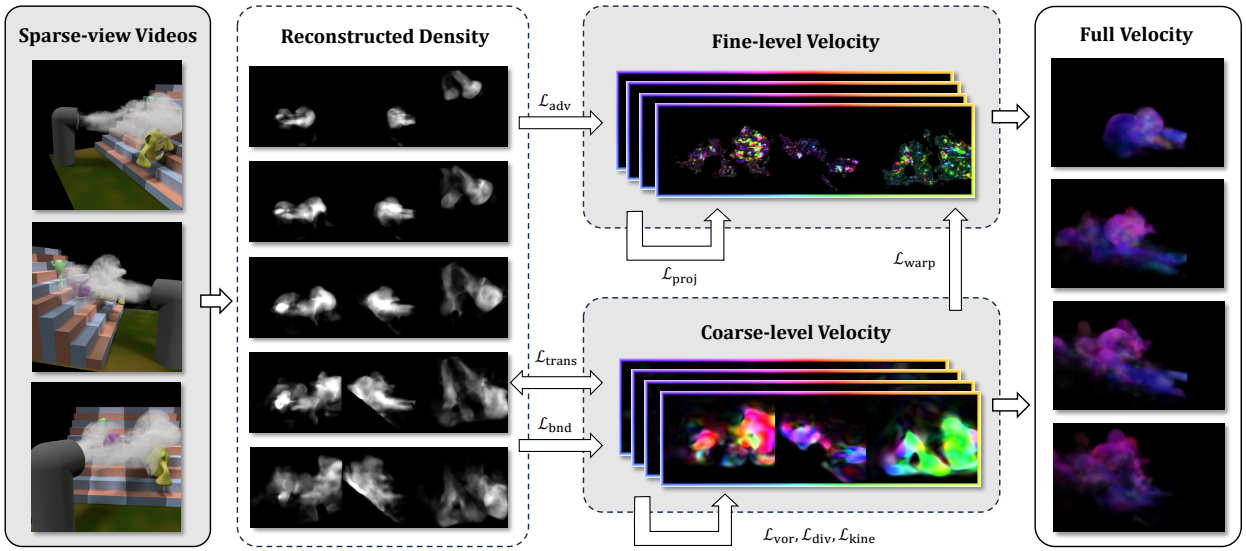

Figure 2: *Method overview. We utilize two distinct neural networks to reconstruct the velocity field at coarse and fine levels. The coarse-level network emphasizes long-term physical consistency, while the fine-level network recovers observational details within the physically valid regions defined by the coarse level. Ultimately, we merge the two into a unified reconstruction that preserves both physical correctness and detailed turbulent motion.*

with h_i being the camera distance of the i-th sampled point. To address the high computational cost of NeRF, iNGP [MESK22] introduces a multi-resolution hash encoding that maps spatial coordinates $\boldsymbol{x}$ to a compact feature vector $\boldsymbol{y}$. This encoding involves hashing the input coordinates and querying feature grids at multiple resolutions, where the range of these resolutions determines the level of detail the reconstruction can capture. A higher range encourages the model to reconstruct finer, high-frequency details, while a lower range focuses more on coarse, low-frequency structures. Subsequently, a lightweight network $m(\boldsymbol{y})$ predicts the radiance color c and density σ, enhancing computational efficiency without sacrificing reconstruction quality.

Extracting surfaces from NeRF is challenging due to the lack of sufficient surface constraints in its representation. To address this, NeuS [WLL*21] introduces a novel volume rendering method to train a neural signed distance function (SDF) representation, which excels at reconstructing high-quality static boundary surfaces.

Chu et al. [CLZ*22] proposed the *SIREN+T* model to improve NeRF's ability on velocity field representation. This model learns $\mathcal{F}(\boldsymbol{x},t) = (\boldsymbol{u})$ for velocity reconstruction, where $\boldsymbol{x}$, t, and $\boldsymbol{u}$ denote the spatial position, time, and flow velocity, respectively. *SIREN+T* uses the MLPs with periodic activation functions proposed in SIREN [SMB*20] instead of ReLU-based MLPs with positional encoding strategies. This design enhances the modeling of continuous derivatives, making it well-suited for representing continuous flow fields.

Physics Constraints The flow we aim to reconstruct is governed by the Navier–Stokes equations (NSEs):

$$\frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{u}\cdot\nabla\boldsymbol{u} = -\frac{1}{\rho_f}\nabla p + \nu\nabla^2\boldsymbol{u} + \boldsymbol{f} \quad \text{and} \quad \nabla\cdot\boldsymbol{u} = 0,$$

where $\boldsymbol{u}$, t, ρ_f, p, ν, $\boldsymbol{f}$ represent the flow velocity, time, flow density, pressure, viscosity coefficient, and external force, respectively. We assume inviscid flow without external forces, following previous methods [CLZ*22; WTC24; YZG*24]. The concentration density ρ satisfies the transport equation:

$$\frac{\partial \rho}{\partial t} + \boldsymbol{u}\cdot\nabla\rho = 0. \tag{1}$$

According to Beer–Lambert law, the concentration density ρ is proportional to the radiance density σ. This relationship allows us to leverage the sparse-view videos to supervise the training of the velocity field.

4 Method

Given the intricate complexity and variability of flow motion, we hierarchically decompose the velocity field into two components: a coarse-level component $\boldsymbol{u}^{\mathrm{c}}$ characterizing the overall flow patterns, and a fine-level component $\boldsymbol{u}^{\mathrm{f}}$ capturing turbulent details [Fri95; Pop00]. As shown in Fig. 2, we employ two separate neural networks to reconstruct them independently, each with a distinct emphasis on physical properties: the coarse-level reconstruction prioritizes long-term physical consistency (§4.1), while the fine-level reconstruction focuses on recovering observational details (§4.2). Finally, we combine both components to obtain our final velocity reconstruction $\boldsymbol{u}^{\mathrm{full}}$. This hierarchical approach effectively integrates strict physical constraints with detailed flow reconstruction.

To supervise the velocity reconstruction, we leverage a density representation inferred from videos via the Navier–Stokes equations, following Wang et al. [WTC24]. Specifically, we employ a *SIREN+T* model for dynamic density and a NeuS model for static boundary reconstruction.

4.1 Coarse-Level Reconstruction

The coarse-level reconstruction aims to establish the fundamental physical fidelity of the flow. Our coarse-level velocity field $\boldsymbol{u}^{\mathrm{c}}$ is represented by a *SIREN+T* model [CLZ*22] to better capture the continuous structure of the flow. We introduce novel supervision: a long-term transport loss optimizing both velocity and density enforcing their consistency over time, A velocity–vorticity formulation loss that ensures accurately compliance with the NSEs while improving convergence stability, a kinetic energy loss suppressing velocity in unsupervised regions, and a boundary loss ensuring velocity constraints at obstacle boundaries for realistic interactions.

Long-Term Transport Loss Most of the previous methods [CLZ*22; YZG*24] employ PDE-based constraints according to Eq. (1) for velocity field learning, which often suffers from localized constraint enforcement while neglecting long-term error accumulation. Wang et al. [WTC24] proposed a long-term supervision framework, however, its velocity is represented by first-order differentiation of neural networks, resulting in significant computational overhead. Inspired by flow map methodologies [DYZ*23], we propose a novel long-term constraint scheme that eliminates differentiation requirements while maintaining neural network compatibility.

Given both the neural network-predicted density field ρ_t at time t and subsequent velocity fields $\boldsymbol{u}^{\mathrm{c}}_t, \cdots, \boldsymbol{u}^{\mathrm{c}}_{t+k-1}$ over k time steps, we can derive the density field $\hat{\rho}_{t+k}$ at time $t+k$ through recursive advection according to Eq. (1):

$$\hat{\rho}_{t+k} = \mathcal{A}(\mathcal{A}(\mathcal{A}(\rho_t, \boldsymbol{u}^{\mathrm{c}}_t), \boldsymbol{u}^{\mathrm{c}}_{t+1}) \cdots, \boldsymbol{u}^{\mathrm{c}}_{t+k-1}), \tag{2}$$

where $\mathcal{A}(\rho, \boldsymbol{u}^{\mathrm{c}})$ denotes a second-order transport scheme for density field ρ via velocity field $\boldsymbol{u}^{\mathrm{c}}$.

As a result, the complete long-term transport loss $\mathcal{L}_{\mathrm{trans}}$ can be formulated as a temporally weighted summation:

$$\mathcal{L}_{\mathrm{trans}} = \sum_{i=1}^{k} \beta^{i-1} \|\hat{\rho}_{t+i} - \rho_{t+i}\|_2^2, \tag{3}$$

where $\beta \in (0, 1]$ serves as a discount factor regulating error propagation across time steps.

Velocity–Vorticity Formulation Loss Previous methods [CLZ*22; YZG*24; WTC24] optimize velocity fields using simplified Navier–Stokes equations, incorporating velocity loss $\mathcal{L}_{\mathrm{vel}}$ and divergence loss $\mathcal{L}_{\mathrm{div}}$ as follows:

$$\mathcal{L}_{\mathrm{vel}} = \left\| \frac{\partial \boldsymbol{u}^{\mathrm{c}}}{\partial t} + \boldsymbol{u}^{\mathrm{c}} \cdot \nabla \boldsymbol{u}^{\mathrm{c}} \right\|_2^2 \quad \text{and} \quad \mathcal{L}_{\mathrm{div}} = \left\| \nabla \cdot \boldsymbol{u}^{\mathrm{c}} \right\|_2^2. \tag{4}$$

However, these methods neglect the pressure projection term in velocity loss $\mathcal{L}_{\mathrm{vel}}$, causing conflicting optimization directions between velocity loss $\mathcal{L}_{\mathrm{vel}}$ and divergence loss $\mathcal{L}_{\mathrm{div}}$. To address this, we introduce a vorticity loss $\mathcal{L}_{\mathrm{vor}}$ as a replacement for velocity loss $\mathcal{L}_{\mathrm{vel}}$, based on the velocity–vorticity formulation of the Navier–Stokes equations:

$$\mathcal{L}_{\mathrm{vor}} = \left\| \frac{\partial \omega^{\mathrm{c}}}{\partial t} + \boldsymbol{u}^{\mathrm{c}} \cdot \nabla \omega^{\mathrm{c}} - \omega^{\mathrm{c}} \cdot \nabla \boldsymbol{u}^{\mathrm{c}} \right\|_2^2, \tag{5}$$

where vorticity $\omega^{\mathrm{c}} = \nabla \times \boldsymbol{u}^{\mathrm{c}}$.

By enforcing vorticity loss $\mathcal{L}_{\mathrm{vor}}$, our method strictly adheres to the Navier–Stokes constraints while avoiding conflicts with divergence loss $\mathcal{L}_{\mathrm{div}}$, leading to improved convergence stability.

Kinetic Energy Loss Existing methods [CLZ*22; YZG*24; WTC24] overlook velocity reconstruction in regions where the smoke density is zero. Physically, velocities should remain nonzero near the smoke and decay to zero farther away. However, current approaches either ignore this issue or simply apply a mask to the reconstructed velocity based on whether the smoke density is zero, leading to physically inaccurate results. To address this, we introduce an kinetic energy loss $\mathcal{L}_{\mathrm{kine}}$ to obtain the minimum kinetic energy solution while satisfying other constraints:

$$\mathcal{L}_{\mathrm{kine}} = \sum \left\| \boldsymbol{u}^{\mathrm{c}} \right\|_2^2. \tag{6}$$

This naturally enforces a suitable mask on the velocity.

Boundary Loss We enforce the no-slip boundary condition on the reconstructed velocity using the immersed boundary method. To achieve this, we introduce a boundary loss $\mathcal{L}_{\mathrm{bnd}}$ to penalize velocities inside or on the boundary:

$$\mathcal{L}_{\mathrm{bnd}} = \sum_{\|S(\boldsymbol{x})\| \le 0} \left\| \boldsymbol{u}^{\mathrm{c}}(\boldsymbol{x}) \right\|_2^2, \tag{7}$$

where $S(\boldsymbol{x})$ is the SDF reconstructed by NeuS. $S(\boldsymbol{x}) \le 0$ indicates that position $\boldsymbol{x}$ lies inside or on the boundary.

All the aforementioned loss terms are computed via auto-differentiation [PGC*17]. The overall loss function for coarse-level reconstruction is formulated as:

$$\mathcal{L}_{\mathrm{coarse}} = \mathcal{L}_{\mathrm{trans}} + \lambda_{\mathrm{vor}} \mathcal{L}_{\mathrm{vor}} + \lambda_{\mathrm{div}} \mathcal{L}_{\mathrm{div}} + \lambda_{\mathrm{kine}} \mathcal{L}_{\mathrm{kine}} + \lambda_{\mathrm{bnd}} \mathcal{L}_{\mathrm{bnd}},$$

where all the λs are loss weights.

4.2 Fine-Level Reconstruction

Fine-level reconstruction focuses on capturing the intricate details of the velocity field from the reconstructed density. This is crucial for preserving turbulence characteristics, which are often essential for accurate velocity estimation and downstream applications. To achieve this, we designed an iNGP network to represent the fine-level velocity field $\boldsymbol{u}^{\mathrm{f}}$, with encoding resolutions set to a higher range, encouraging the network to capture high-frequency details.

Since we focus on local details now, we train the fine-level velocity field $\boldsymbol{u}^{\mathrm{f}}$ using a short-term PDE-based advection loss $\mathcal{L}_{\mathrm{adv}}$:

$$\mathcal{L}_{\mathrm{adv}} = \left\| \frac{\partial \rho}{\partial t} + (\boldsymbol{u}^{\mathrm{c}} + \boldsymbol{u}^{\mathrm{f}}) \cdot \nabla \rho \right\|_2^2. \tag{8}$$

Enforcing the full physics at the fine-level, as we do at the coarse-level, is not effective due to the lack of accurate detailed information at this scale. Direct physical constraints would be impractical

and may disrupt the turbulent flow dynamics. Therefore, we use the coarse-level reconstructed velocity field $\boldsymbol{u}^{\mathrm{c}}$ to introduce the warp loss $\mathcal{L}_{\mathrm{warp}}$ and projection loss $\mathcal{L}_{\mathrm{proj}}$ as simplified constraints based on the Navier–Stokes equations:

$$\mathcal{L}_{\mathrm{warp}} = \left\| \frac{\partial \boldsymbol{u}^{\mathrm{f}}}{\partial t} + \boldsymbol{u}^{\mathrm{c}} \cdot \nabla \boldsymbol{u}^{\mathrm{f}} \right\|_2^2 \quad \text{and} \tag{9}$$

$$\mathcal{L}_{\mathrm{proj}} = \left\| \boldsymbol{u}^{\mathrm{f}} - \boldsymbol{u}^{\mathrm{f}}_p \right\|_2^2, \tag{10}$$

where the warp loss $\mathcal{L}_{\mathrm{warp}}$ promotes consistent flow reconstruction across different spatial scales, inspired by Kim et al. [KTJG08], while the projection loss $\mathcal{L}_{\mathrm{proj}}$ serves as a weak constraint enforcing the fine-level velocity field $\boldsymbol{u}^{\mathrm{f}}$ to be divergence-free. Here, $\boldsymbol{u}^{\mathrm{f}}_p$ denotes the pressure-projected velocity field of $\boldsymbol{u}^{\mathrm{f}}$, computed using the pressure projection solver described by Yu et al. [YZG*24].

As a result, the overall loss function for fine-level reconstruction is formulated as:

$$\mathcal{L}_{\mathrm{fine}} = \mathcal{L}_{\mathrm{adv}} + \lambda_{\mathrm{warp}} \mathcal{L}_{\mathrm{warp}} + \lambda_{\mathrm{proj}} \mathcal{L}_{\mathrm{proj}},$$

where the λ terms serve as weighting coefficients.

4.3 Velocity Field Combination

Finally, we obtain the full velocity field $\boldsymbol{u}^{\mathrm{full}}$ by combining the coarse-level velocity field $\boldsymbol{u}^{\mathrm{c}}$ and the fine-level velocity field $\boldsymbol{u}^{\mathrm{f}}$. The coarse level captures the global low-frequency structure, while the fine level represents high-frequency details. Their combination yields a representation covering the complete range of flow features.

Formally, the full velocity field $\boldsymbol{u}^{\mathrm{full}}$ is computed as

$$\boldsymbol{u}^{\mathrm{full}} = \boldsymbol{u}^{\mathrm{c}} + \alpha \boldsymbol{u}^{\mathrm{f}}, \quad \text{with} \quad \alpha = \min\{(\|\boldsymbol{u}^{\mathrm{c}}\|_2 / m)^5, 1\}, \tag{11}$$

where m denotes twice the average norm of the coarse-level velocity field $\boldsymbol{u}^{\mathrm{c}}$ at each time step.

The scale factor α is introduced because the fine-level velocity field $\boldsymbol{u}^{\mathrm{f}}$ is not explicitly constrained by the kinetic energy loss L_{kine} and the boundary loss L_{bnd}, it may produce artifacts in boundary regions or in regions with zero density. Rather than imposing the same losses—which would increase the complexity of training—we leverage the already-trained coarse velocity field $\boldsymbol{u}^{\mathrm{c}}$ as a mask, achieving a similar regularizing effect in a simpler manner.

5 Experiments

We evaluate our method against state-of-the-art neural velocity reconstruction approaches, including PINF [CLZ*22], PICT [WTC24], and HyFluid [YZG*24]. To ensure a comprehensive comparison, we conduct experiments on three datasets: two synthetic and one real-captured. The first is the Cylinder scene proposed by Wang et al. [WTC24], a hybrid synthetic dataset that includes obstacles. The second is ScalarSyn, a fully synthetic dataset derived from ScalarFlow [EUT19]. The third is the real captured ScalarFlow dataset [EUT19]. It is important to note that HyFluid supports only scenes without obstacles and is therefore excluded from the evaluation on the Cylinder dataset.

For evaluation, we begin with an analysis of the reconstructed velocity fields through both qualitative visualizations and quantitative metrics (§5.1). To further assess the effectiveness of our method, we apply the reconstructed velocity to a series of downstream tasks, including tracer visualization (§5.2) and re-simulation (§5.3). Finally, ablation studies are presented in §5.4. Additional results, as well as implementation details, are provided in the supplementary material for completeness.

Across these evaluations, our method achieves superior performance compared to all baselines, both qualitatively and quantitatively. We attribute this advantage to our hybrid framework design, in which the coarse-phase stage produces clear boundaries and plausible large-scale structures, while the fine-phase stage captures high-frequency details. By combining these components, our framework reconstructs velocity fields more accurately, leading to improved results in downstream tasks such as tracer visualization and re-simulation.

5.1 Velocity field Analysis

To evaluate the effectiveness of our reconstructed velocity field, we conduct both qualitative and quantitative analyses.

For the qualitative evaluation, we first visualize the velocity field

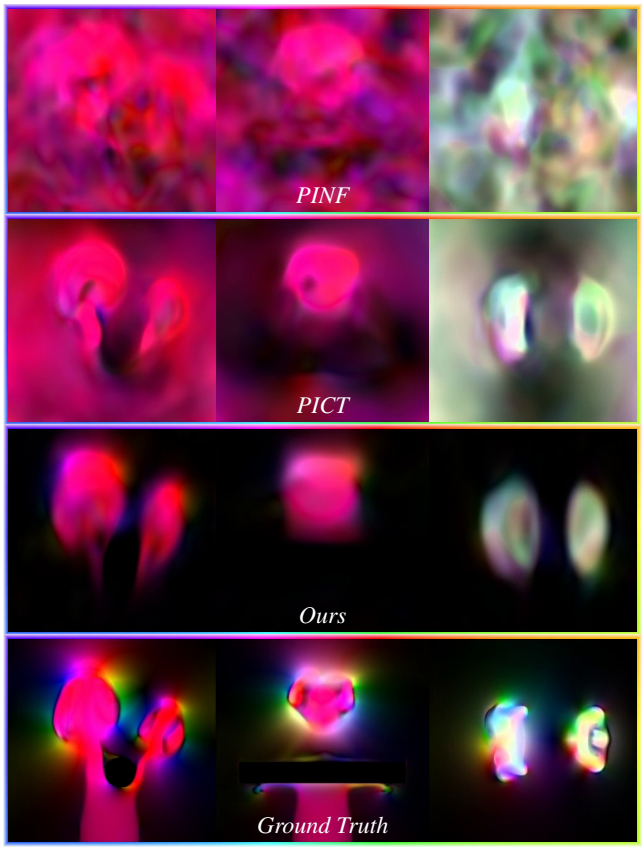

Figure 3: *Velocity visualization on the Cylinder scene. The results show that our method reconstructs a velocity field closer to the ground truth compared to PINF and PICT.*

Table 1: *Quantitative comparisons on Cylinder, ScalarSyn and ScalarFlow. We evaluate the l_2 errors of divergence (towards zero), and the velocity and vorticity fields (against ground truth), restricted to regions where the ground truth density is non-zero. The Cylinder scene excludes HyFluid since it does not support scenes with obstacles. For the real-captured ScalarFlow dataset, ground truth velocity and vorticity are not available, so these metrics are not evaluated. Our method consistently achieves results closest to the ground truth across all scenes.*

Model	Cylinder			ScalarSyn			ScalarFlow		
	divergence↓	velocity↓	vorticity↓	divergence↓	velocity↓	vorticity↓	divergence↓	velocity↓	vorticity↓
PINF	0.0005040	0.1244	0.004989	0.004024	0.1465	0.01302	0.003482	–	–
PICT	0.0004526	0.1146	0.004811	0.002380	0.1166	0.01250	0.002177	–	–
HyFluid	–	–	–	0.3268	0.4033	0.07176	0.003058	–	–
Ours	**0.0001801**	**0.1103**	**0.004571**	**0.002241**	**0.05896**	**0.01189**	**0.0004503**	–	–

using the middle slices of the front, side, and top views, following the approach of previous methods [CLZ*22; WTC24]. As shown in Fig. 3 and Fig. 4, it is evident that our method reconstructs the overall structure and finer details of the flow more accurately than the others. Notably, our method shows a clear advantage in reconstructing the background regions, where other methods often produce spurious non-zero noise. This improvement is largely attributed to the incorporation of the kinetic energy loss $\mathcal{L}_{\text{kine}}$ and boundary loss $\mathcal{L}_{\text{bnd}}$ in our framework.

While the qualitative results highlight the superior background reconstruction of our method, we note that this is not the sole source of improvement. To ensure a fairer evaluation, our quantitative analysis focuses only on regions where the ground truth density is non-zero, excluding the background. As shown in Table 1, our method still achieves lower errors in velocity and vorticity, as well as reduced divergence, demonstrating superior physical accuracy across the entire flow domain.

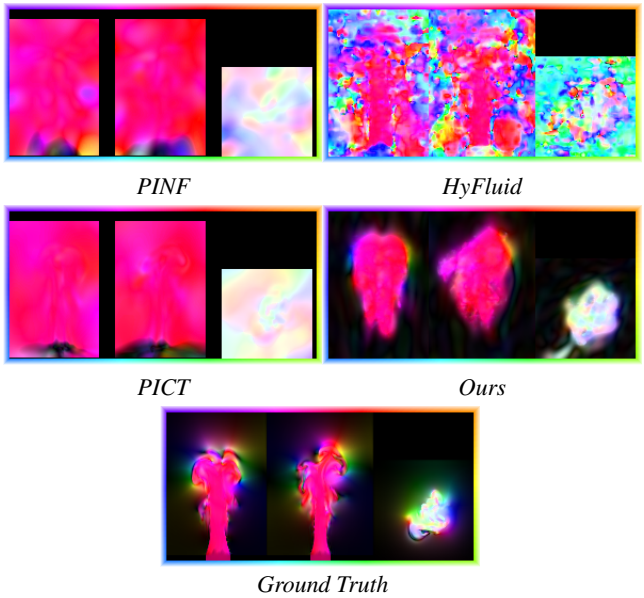

Figure 4: *The velocity visualization on the ScalarSyn scene. The results demonstrate that the velocity field reconstructed by our method is closer to the overall structure of the ground truth, while also preserving the corresponding turbulent details.*

5.2 Tracer Visualization

In this task, we simulate the motion of virtual paper pieces gently laid across a plane near the inflow region. These imaginary tracers are advected by the reconstructed velocity field to visualize the flow dynamics. Both the reconstructed smoke and the paper pieces are rendered in Blender [Ble]. As illustrated in Fig. 8, Hyfluid produces non-physical results, causing the paper to move chaotically in an unrealistic manner. PINF and PICT, on the other hand, also push paper pieces that are supposed to remain stationary, which introduces deviations from the expected behavior. In contrast, our method yields motion that closely aligns with the ground truth.

5.3 Re-Simulation

We conducted the re-simulation task following Yu et al. [YZG*24]. Specifically, we take the first-frame density from the reconstruction and advect it using the reconstructed velocity field until the final frame. The advection is implemented using the MacCormack method [SFK*08]. As shown in Fig. 5, Fig. 6, and Fig. 7, our re-simulated smoke exhibits finer details and better alignment with the ground truth, indicating a more accurate velocity field reconstruction than all baseline methods. Quantitatively, our method also achieves the highest peak signal-noise ratio (PSNR) score, further validating its superior performance.

5.4 Abalation Study

Since our main contributions lie in the hybrid framework that combines coarse-level and fine-level representations, as well as the introduction of the vorticity loss $\mathcal{L}_{\text{vor}}$, transport loss $\mathcal{L}_{\text{trans}}$, kinetic loss $\mathcal{L}_{\text{kine}}$ and boundary loss $\mathcal{L}_{\text{bnd}}$, we conduct an ablation study to evaluate the effectiveness of these components.

As shown in Table 2, both vorticity loss $\mathcal{L}_{\text{vor}}$ and transport loss $\mathcal{L}_{\text{trans}}$ help produce reconstructions that are closer to the ground truth. Furthermore, since vorticity loss $\mathcal{L}_{\text{vor}}$ does not conflict with the divergence loss $\mathcal{L}_{\text{div}}$, it also leads to more stable convergence, as demonstrated in Fig. 12.

As illustrated in Fig. 9, we reconstruct the velocity field on the

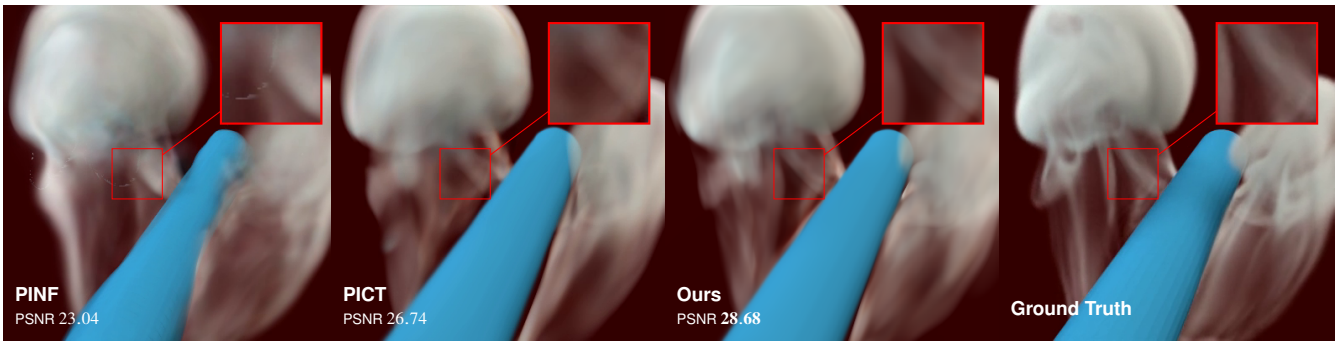

Figure 5: *Visualization of re-simulation results on the Cylinder scene. Compared to PINF and PICT, our method achieves finer details and better alignment with the ground truth.*

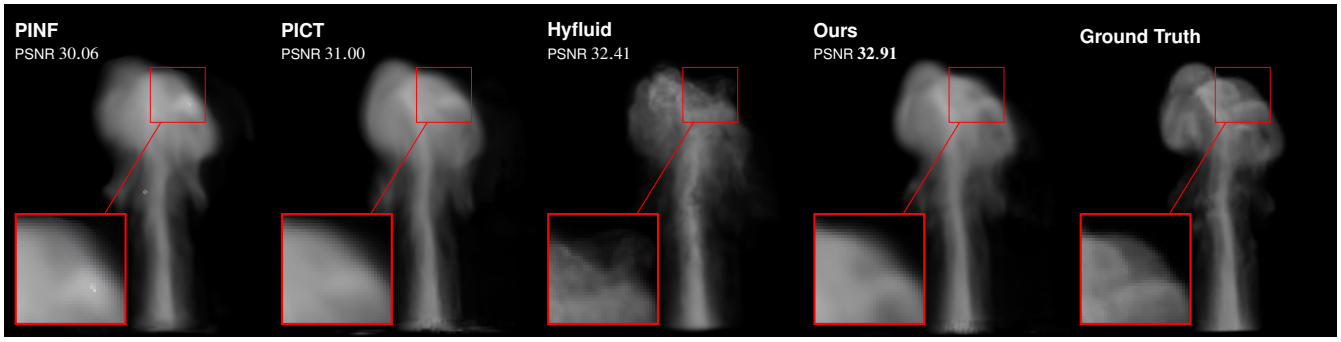

Figure 6: *Visualization of re-simulation results on the ScalarSyn scene. Our method better reproduces the fine high-frequency structures of smoke compared to PINF and PICT, while also avoiding the introduction of unphysical noise compared to HyFluid.*

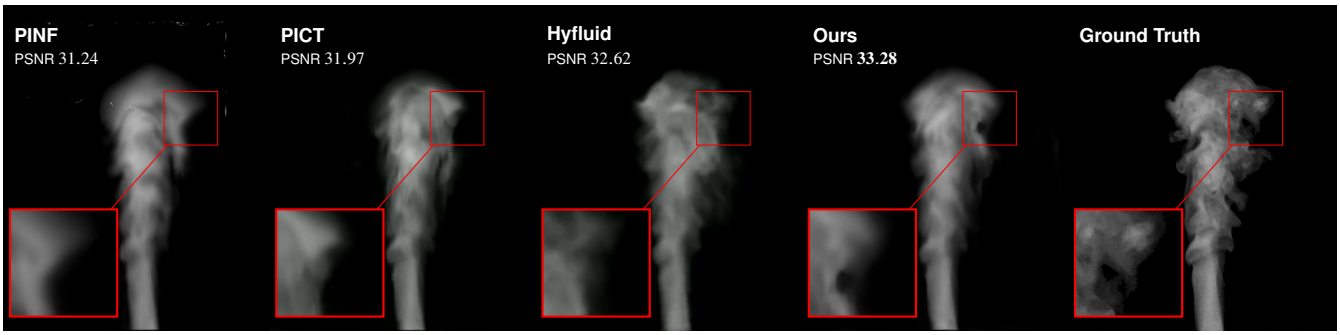

Figure 7: *Visualization of re-simulation results on the ScalarFlow dataset. Our method more effectively prevents smoke streaking compared to PINF and PICT, as it better captures turbulent details by accurately modeling both coarse-level and fine-level velocity. Compared to HyFluid, our method captures high-frequency information with greater physical accuracy, allowing us to faithfully reproduce the smoke details of the ground truth, as shown in the red box.*

Cylinder scene without the kinetic loss $\mathcal{L}_{\text{kine}}$ and the boundary loss $\mathcal{L}_{\text{bnd}}$, respectively. It is clear that the kinetic loss $\mathcal{L}_{\text{kine}}$ contributes to a better reconstruction by encouraging the background velocity to approach zero. In contrast, the boundary loss $\mathcal{L}_{\text{bnd}}$ enforces velocities at the boundaries to zero, ensuring that the boundary conditions are satisfied.

To assess the impact of fine-level reconstruction, we visualize the reconstructed velocity fields and vorticity fields using volume rendering under different scenes, as shown in Fig. 10 and Fig. 11. The results demonstrate that our fine-level reconstruction adaptively adds high-frequency details based on the scene characteristics. In smooth cases like Cylinder, where the coarse-level output is already accurate, it avoids introducing unnecessary noise. In contrast, for turbulent scenes such as ScalarFlow, it supplements missing fine-scale structures that the coarse level fails to capture. As shown in Fig. 13, these added details lead to re-simulations that more closely match the ground truth, confirming their physical relevance.

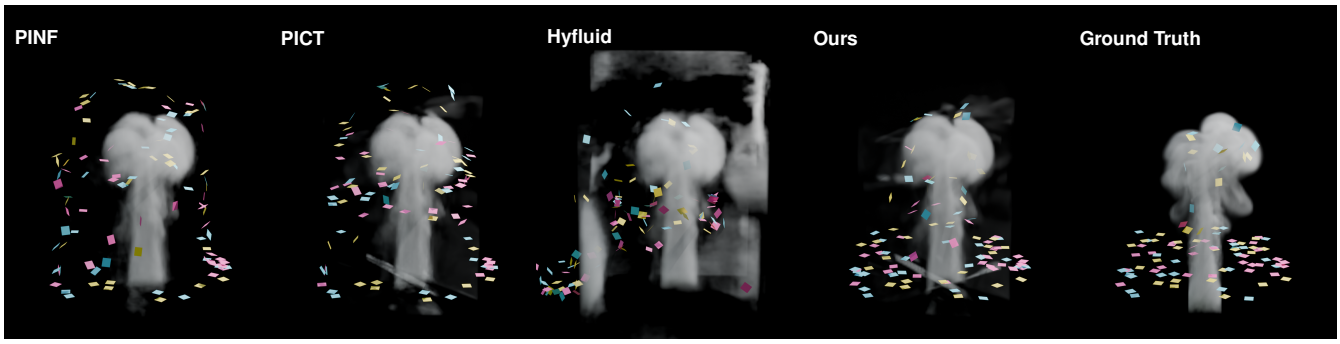

Figure 8: *Tracer visualization results on ScalarSyn. As shown in the visualization, all baseline methods—PINF, PICT, and HyFluid—incorrectly lift the paper pieces that should remain stationary. Moreover, HyFluid produces chaotic velocity fields that are physically implausible. In contrast, our method accurately reconstructs the motion, closely matching the ground truth.*

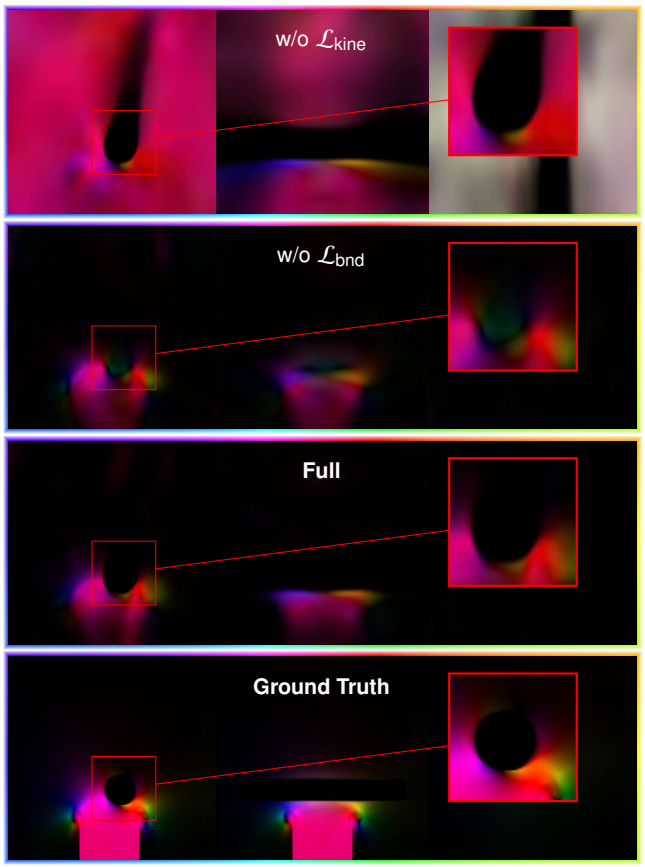

Figure 9: *Ablation study of the kinetic loss $\mathcal{L}_{kine}$ and boundary loss $\mathcal{L}_{bnd}$ on the Cylinder scene. The kinetic loss $\mathcal{L}_{kine}$ promotes a clean reconstruction of the background velocity, while the boundary loss $\mathcal{L}_{bnd}$ enforces velocities at the boundaries to zero, ensuring that the boundary conditions are satisfied.*

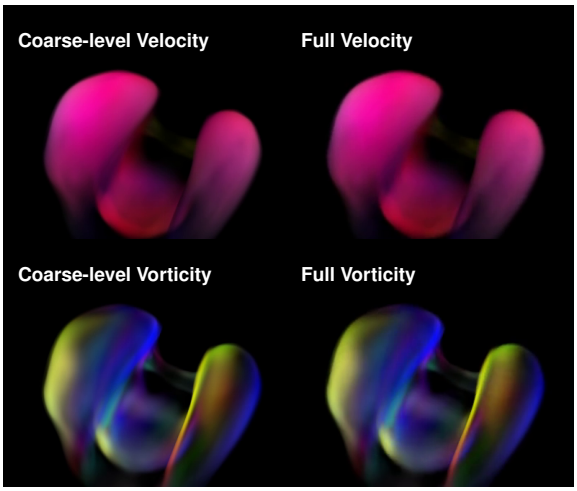

Figure 10: *Volume rendering results of the reconstructed velocity and vorticity fields on Cylinder. Due to the relatively smooth nature of the flow in this scene, the coarse-level reconstruction already captures the essential structures of the velocity and vorticity fields. As a result, the fine-level reconstruction introduces minimal changes, demonstrating that our method adaptively adds fine-scale turbulent details only when necessary.*

6 Conclusion and Discussion

We have presented a novel framework for fluid reconstruction from sparse video inputs that addresses the inherent challenges of accurately capturing turbulent velocity fields while maintaining long-term physical consistency. Our approach introduces a strategic split in supervision across spatial scales, with fine-scale observation fidelity focused on turbulence details and coarse-scale consistency enforcing long-term physical behavior. In this way, our method achieves high-fidelity reconstructions that accurately represent large-scale flow dynamics and fine-scale turbulence.

Nonetheless, accurately recovering the true velocity fields from sparse video inputs remains a challenging task. We believe the primary reason for the discrepancy between the reconstructed velocity

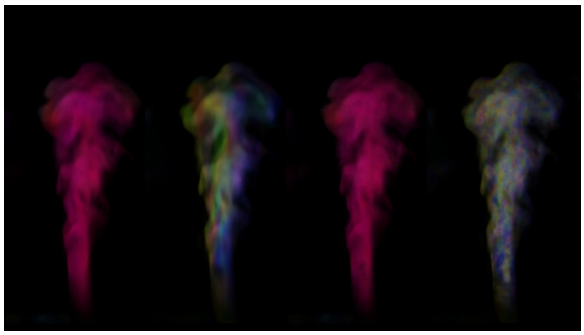

Figure 11: *Volume rendering results of the reconstructed velocity and vorticity fields on ScalarFlow. The four columns show, respectively, the coarse-level velocity field, coarse-level vorticity field, full velocity field and full vorticity field. These results suggest that the fine-level reconstruction enhances turbulent features while maintaining the global structure obtained at the coarse level.*

Model	Temporal Supervision	Physics Constraints
short-**u**	using advection loss $\mathcal{L}_{\text{adv}}$	using velocity loss $\mathcal{L}_{\text{vel}}$
short-ω	using advection loss $\mathcal{L}_{\text{adv}}$	using vorticity loss $\mathcal{L}_{\text{vor}}$
long-**u**	using transport loss $\mathcal{L}_{\text{trans}}$	using velocity loss $\mathcal{L}_{\text{vel}}$
long-ω	using transport loss $\mathcal{L}_{\text{trans}}$	using vorticity loss $\mathcal{L}_{\text{vor}}$

l_2 errors on	divergence↓	velocity↓	vorticity↓
short-**u**	0.0003946	0.1182	0.004697
short-ω	0.0002934	0.1131	0.004656
long-**u**	0.0002186	0.1105	0.004586
long-ω	**0.0001801**	**0.1103**	**0.004571**

Table 2: *The ablation study on the Cylinder scene. The long-ω model is our full method, while others are ablated versions, highlighting that better convergence is achieved by the long temporal loss $\mathcal{L}_{\text{trans}}$ and the vorticity-based physical constraints $\mathcal{L}_{\text{vor}}$.*

field and the ground truth stems from the inaccuracy in the reconstructed density field. Since the density field is evaluated through volume rendering from sparse-view videos, there is significant ambiguity in the solutions found by the neural network — the rendered results may closely match the input videos, but still deviate substantially from the true density, especially under unknown lighting conditions. For example, PINF and PICT tend to produce overly smooth reconstructions, while Hyfluid results are often quite chaotic. Although our hybrid framework incorporates more accurate physical constraints during training, this issue remains only partially resolved. Addressing these reconstruction gaps is an important direction for future work, aiming to further improve the fidelity of inferred velocity fields.

Building on these challenges, our method also encounters certain inherent limitations. A potential direction for future work is to incorporate these factors to enhance the realism and accuracy of the simulation. Additionally, our method focuses solely on gas and does not account for liquids. Since liquids feature free surfaces, they may require additional physical modeling and constraints. Finally, like other NeRF-based neural representations, our optimization process is relatively slow. On an NVIDIA 4090D GPU, our current implementation requires about 12 hours for high-fidelity smoke reconstruction. A promising direction for future work is to integrate our framework with faster reconstruction methods, such as 3DGS [KKLD23], to improve efficiency.

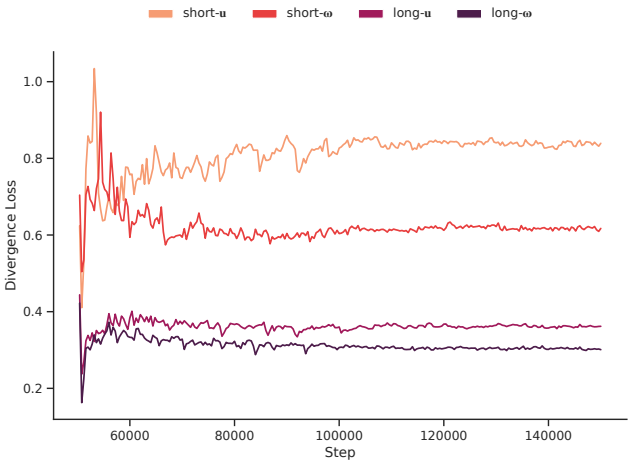

Figure 12: *Convergence curves of the divergence loss $\mathcal{L}_{\text{div}}$ using different loss combinations in the Cylinder scene. It is clear that the velocity-vorticity formulation loss $\mathcal{L}_{\text{vor}}$ is more effective in aiding the convergence of the divergence $\mathcal{L}_{\text{div}}$ than the velocity loss $\mathcal{L}_{\text{vel}}$.*

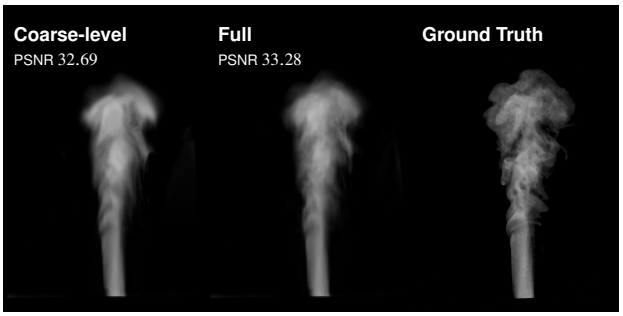

Figure 13: *The ablation study on the ScalarFlow dataset. We visualize the re-simulation results using both coarse-level and full velocity. It is clear that incorporating fine-level velocity helps capture the high-frequency components more effectively.*

References

[Ble] BLENDER ONLINE COMMUNITY. *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam. URL: https://www.blender.org 6.

[CDG*22] CUOMO, SALVATORE, DI COLA, VINCENZO SCHIANO, GIAMPAOLO, FABIO, et al. "Scientific Machine Learning Through Physics–Informed Neural Networks: Where we are and What's Next". *J. Sci. Comput.* 92.3 (Sept. 2022). ISSN: 0885-7474. DOI: 10.1007/s10915-022-01939-z. URL: https://doi.org/10.1007/s10915-022-01939-z 2.

[CLZ*22] CHU, MENGYU, LIU, LINGJIE, ZHENG, QUAN, et al. "Physics informed neural fields for smoke reconstruction with sparse data". *ACM Transactions on Graphics (ToG)* 41.4 (2022), 1–14 2–6.

[Cot00] COTTET, GH. *Vortex Methods: Theory and Practice*. 2000 2.

[DWD*24] DUAN, YUANXING, WEI, FANGYIN, DAI, QIYU, et al. "4d-rotor gaussian splatting: towards efficient novel view synthesis for dynamic scenes". *ACM SIGGRAPH 2024 Conference Papers*. 2024, 1–11 2.

[DYZ*23] DENG, YITONG, YU, HONG-XING, ZHANG, DIYANG, et al. "Fluid simulation on neural flow maps". *ACM Transactions on Graphics (TOG)* 42.6 (2023), 1–21 2, 4.

[ESWvO06] ELSINGA, GERRIT E, SCARANO, FULVIO, WIENEKE, BERNHARD, and van OUDHEUSDEN, BAS W. "Tomographic particle image velocimetry". *Experiments in fluids* 41.6 (2006), 933–947 2.

[EUT19] ECKERT, MARIE-LENA, UM, KIWON, and THUEREY, NILS. "ScalarFlow: a large-scale volumetric data set of real-world scalar transport flows for computer animation and machine learning". *ACM Trans. Graph.* 38.6 (Nov. 2019) 2, 5.

[Fri95] FRISCH, URIEL. *Turbulence: The Legacy of A. N. Kolmogorov*. Cambridge University Press, 1995 3.

[FST21] FRANZ, ERIK, SOLENTHALER, BARBARA, and THUEREY, NILS. *Global Transport for Fluid Reconstruction with Learned Self-Supervision*. 2021. arXiv: 2104.06031 [cs.CV]. URL: https://arxiv.org/abs/2104.06031 2.

[GITH14] GREGSON, JAMES, IHRKE, IVO, THUEREY, NILS, and HEIDRICH, WOLFGANG. "From capture to simulation: connecting forward and inverse problems in fluids". *ACM Transactions on Graphics (TOG)* 33.4 (2014), 1–11 2.

[GNG*13] GU, JINWEI, NAYAR, S.K., GRINSPUN, E., et al. "Compressive Structured Light for Recovering Inhomogeneous Participating Media". *PAMI* 35.3 (2013), 555–567 2.

[Gra97] GRANT, I. "Particle Image Velocimetry: a Review". *Proceedings of the Institution of Mechanical Engineers* 211.1 (1997), 55–76 2.

[JYY13] JI, YU, YE, JINWEI, and YU, JINGYI. "Reconstructing Gas Flows Using Light-Path Approximation". *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2013 2.

[KKLD23] KERBL, BERNHARD, KOPANAS, GEORGIOS, LEIMKÜHLER, THOMAS, and DRETTAKIS, GEORGE. "3d gaussian splatting for real-time radiance field rendering." *ACM Trans. Graph.* 42.4 (2023), 139–1 9.

[KTJG08] KIM, THEODORE, THÜREY, NILS, JAMES, DOUG, and GROSS, MARKUS. "Wavelet turbulence for fluid simulation". *ACM Transactions on Graphics (TOG)* 27.3 (2008), 1–6 2, 5.

[LCT25] LIU, QIANG, CHU, MENGYU, and THUEREY, NILS. "ConFIG: Towards Conflict-free Training of Physics Informed Neural Networks". *The Thirteenth International Conference on Learning Representations*. 2025. URL: https://openreview.net/forum?id=APojAzJQiq 2.

[MESK22] MÜLLER, THOMAS, EVANS, ALEX, SCHIED, CHRISTOPH, and KELLER, ALEXANDER. "Instant neural graphics primitives with a multiresolution hash encoding". *ACM Transactions on Graphics* 41.4 (July 2022), 1–15. ISSN: 1557-7368. DOI: 10.1145/3528223.3530127. URL: http://dx.doi.org/10.1145/3528223.3530127 3.

[MST*20] MILDENHALL, BEN, SRINIVASAN, PRATUL P., TANCIK, MATTHEW, et al. *NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis*. 2020. arXiv: 2003.08934 [cs.CV]. URL: https://arxiv.org/abs/2003.08934 2.

[NWRC22] NABIZADEH, MOHAMMAD SINA, WANG, STEPHANIE, RAMAMOORTHI, RAVI, and CHERN, ALBERT. "Covector fluids". *ACM Transactions on Graphics (TOG)* 41.4 (2022), 1–16 2.

[PGC*17] PASZKE, ADAM, GROSS, SAM, CHINTALA, SOUMITH, et al. "Automatic differentiation in pytorch". (2017) 4.

[Pop00] POPE, STEPHEN B. *Turbulent Flows*. Cambridge University Press, 2000 3.

[QCL*24] QIU, JIAXIONG, CEN, RUIHONG, LI, ZHONG, et al. "NeuSmoke: Efficient Smoke Reconstruction and View Synthesis with Neural Transportation Fields". *SIGGRAPH Asia 2024 Conference Papers*. 2024, 1–12 2.

[QZG*19] QU, ZIYIN, ZHANG, XINXIN, GAO, MING, et al. "Efficient and conservative fluids using bidirectional mapping". *ACM Transactions on Graphics (TOG)* 38.4 (2019), 1–12 2.

[RCDB23] ROSSET, NICOLAS, CORDONNIER, GUILLAUME, DUVIGNEAU, REGIS, and BOUSSEAU, ADRIEN. "Interactive design of 2D car profiles with aerodynamic feedback". *Computer Graphics Forum*. Vol. 42. 2. Wiley Online Library. 2023, 427–437 2.

[SFK*08] SELLE, ANDREW, FEDKIW, RONALD, KIM, BYUNGMOON, et al. "An unconditionally stable MacCormack method". *Journal of Scientific Computing* 35 (2008), 350–371 6.

[SMB*20] SITZMANN, VINCENT, MARTEL, JULIEN, BERGMAN, ALEXANDER, et al. "Implicit neural representations with periodic activation functions". *Advances in neural information processing systems* 33 (2020), 7462–7473 3.

[TP11] TESSENDORF, JERRY and PELFREY, BRANDON. "The characteristic map for fast and efficient vfx fluid simulations". *Computer Graphics International Workshop on VFX, Computer Animation, and Stereo Movies. Ottawa, Canada*. 2011 2.

[WLL*21] WANG, PENG, LIU, LINGJIE, LIU, YUAN, et al. "Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction". *arXiv preprint arXiv:2106.10689* (2021) 3.

[WTC24] WANG, YIMING, TANG, SIYU, and CHU, MENGYU. "Physics-Informed Learning of Characteristic Trajectories for Smoke Reconstruction". *ACM SIGGRAPH 2024 Conference Papers*. 2024, 1–11 2–6.

[YZG*24] YU, HONG-XING, ZHENG, YANG, GAO, YUAN, et al. "Inferring hybrid neural fluid fields from videos". *Advances in Neural Information Processing Systems* 36 (2024) 2–6.

[ZBY*24] ZENG, YIXIN, BI, ZOUBIN, YIN, MINGRUI, et al. "Real-time Acquisition and Reconstruction of Dynamic Volumes with Neural Structured Illumination". *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024, 20186–20195 2.