Continuous Parametric Optical Flow

Jianqin Luo *, Zhexiong Wan *, Yuxin Mao , Bo Li , Yuchao Dai†

Northwestern Polytechnical University & Shaanxi Key Laboratory of Information Acquisition and Processing

Project Page – CPFlow
https://npucvr.github.io/CPFlow

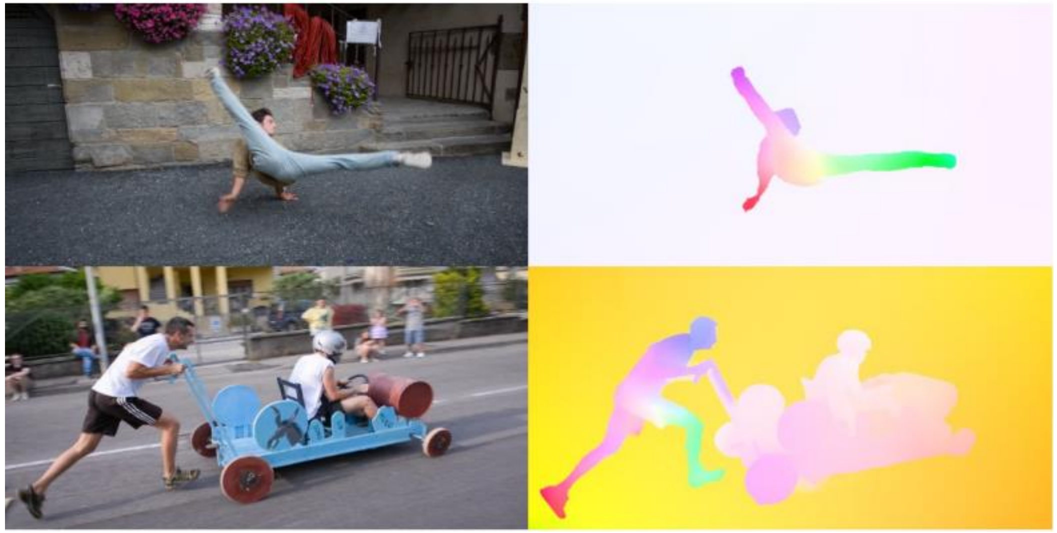
Github Page – CPFlow
https://github.com/LuoRadisher/CPFlow

Introduction

Representation of Fine-grained Motion

- **Optical Flow:**
  - ✓ **Dense & Short-term motion displacements between the specific frame pairs.**
- **Point Tracker (Tracking-Any-Point):**
  - ✓ **Sparse & Long-term motion trajectories during the whole frame sequences.**

** Above description is for one-step inference result with practicle resource



Dense & Short-term Optical Flow [1]

Sparse & Long-term point tracker [2]

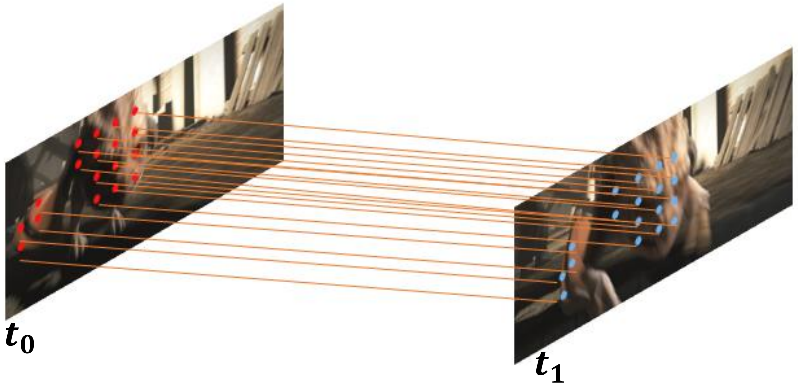
[1] Teed Z, Deng J. Raft: Recurrent all-pairs field transforms for optical flow[C]//Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16. Springer International Publishing, 2020: 402-419.
[2] Harley A W, Fang Z, Fragkiadaki K. Particle video revisited: Tracking through occlusions using point trajectories[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 59-75.

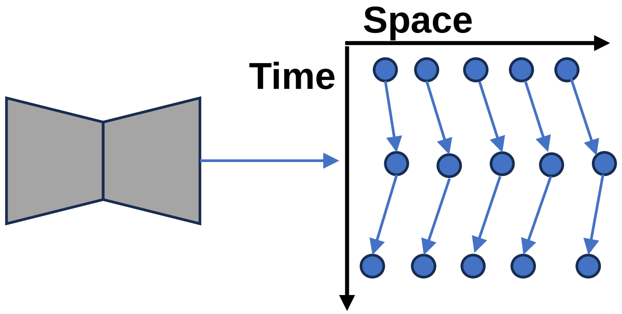
Motivation

Temporal Continuity of Real Motion

Discrete Flow

- **Description:** **discrete** point correspondence.
- **Concept Limitation:** **frame-to-frame** mapping.
- **Implement Limitation:** **implicit prediction.**



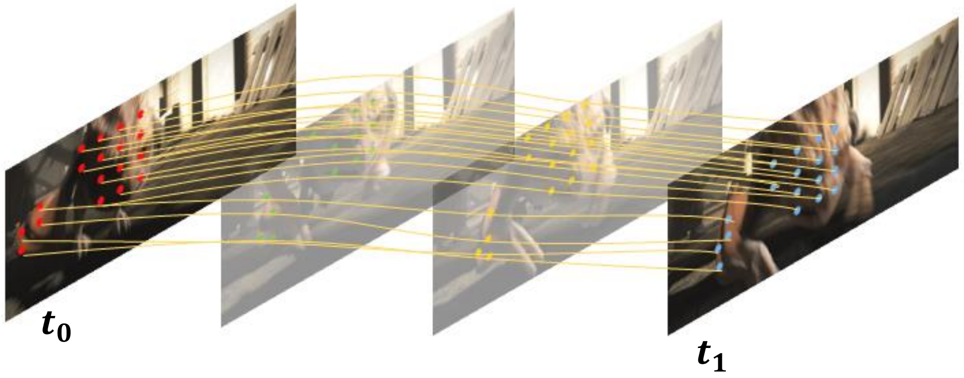


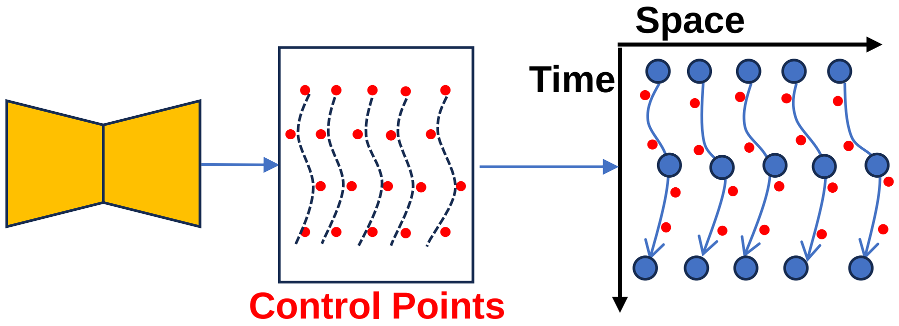
Continuous Parametric Flow

✓ **Description:** **continuous** point trajectories.

✓ **Concept Expansion:** **time-to-time** mapping.

✓ **Implement Optimization:** **parametric constraints with implicit regression.**





Motivation

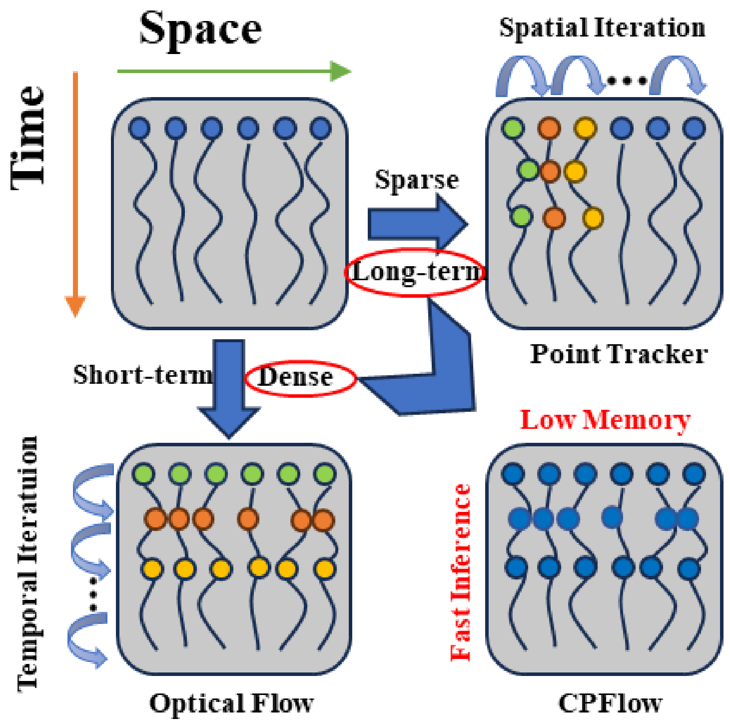
Efficiency of Spatial-Temporal Aggegation

- **Expectation:** dense & long-term motion estimation for a video.

- **Optical Flow:** step-to-step chain needs additional inference time.
- **Point Tracker:** single-point centric model needs large memory for image-scale inference.

✓ **CPFlow: spatially dense** & **temporally continuous** motion prediction for one-step inference.

	Dense?	Long-term?
Optical Flow	✓	✗
Point Tracker	✗	✓
CPFlow	✓	✓



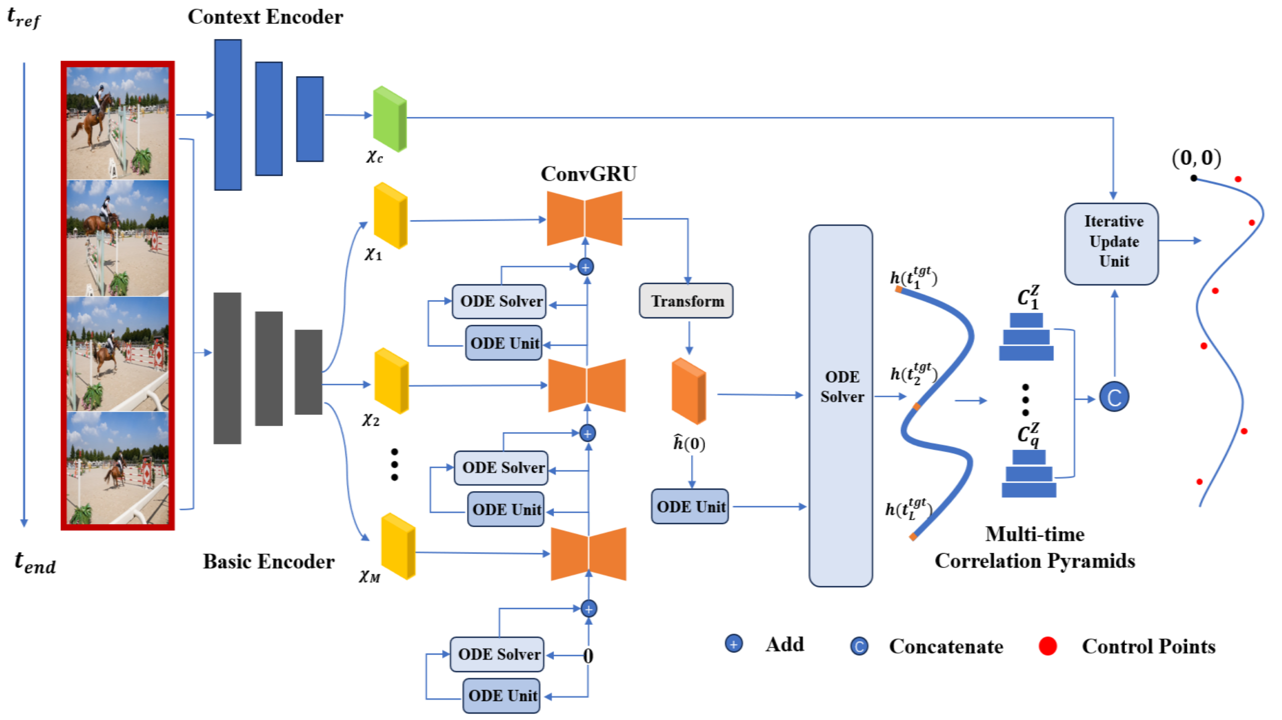
Method

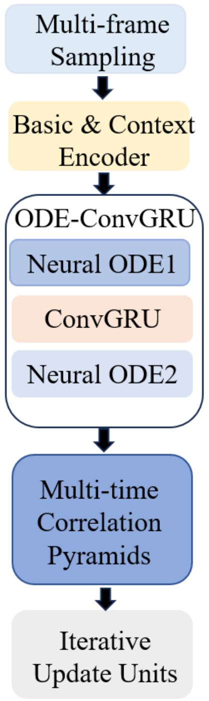
Overall Architechture

- **Continuous Feature Generation:** implicitly construct spatio-temporal representation binded with candidate moment by **Neural ODE** with **ConvGRU.**

- **Continuous Parametric Representation:** explicitly describe continuous flow trajectory by cubic B-splines with flexible control points regressed by our model.

- **Multi-time Correlation & Iterative Update :** effectively connect implicit features with explicit parameter regression.





Method

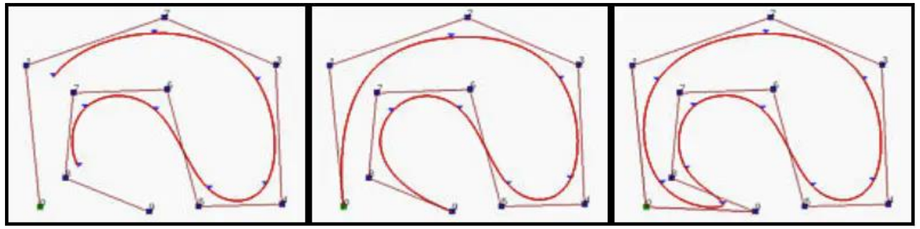
Parametric Model

- **Goal:** select a powerful & flexible curve to describe continuous flow.
- **Comparision:**
  a. Polynomial Curve: fitting limitation, inflexible, hard to pass through occlusion
  b. Bezier Curve: flexible control point, global optimization
  c. B-splines Curve: **flexible control points, local optimization**
- **Solution: Clamp Cubic B-spline**

$$F(t) = \sum_{i=0}^{N-1} \boxed{B_{i,k}(t)} P_i$$

$$B_{i,0}(t) = \begin{cases} 1 & t_i \leq t < t_{i+1} \\ 0 & \text{otherwise} \end{cases},$$

$$B_{i,k}(t) = \frac{t - t_i}{t_{i+k} - t_i} B_{i,k-1}(t) + \frac{t_{i+k+1} - t}{t_{i+k+1} - t_{i+1}} B_{i+1,k-1}(t).$$



Open Clamp Close

Knots List: Repeated Knots adjust the shape of curve

Method

Optimization Framework

- Goal: construct a valid end-to-end model to regress the control points for every pixel
- key: insert spatial-temporal feature representation for specific moment
- Solution: Neural ODE & ConvGRU
  a. Neural ODE: temporal continuity in feature space
  b. ConvGRU: aggregation for spatial-temporal information
- Supervision

$$L = \sum_{i=1}^{S} \gamma^{S-i} \frac{1}{N_{gt}} \sum_{j=1}^{N_{gt}} ||F(t_j^{sup}) - D_{gt}(t_j^{sup})||_1$$

Neural ODE

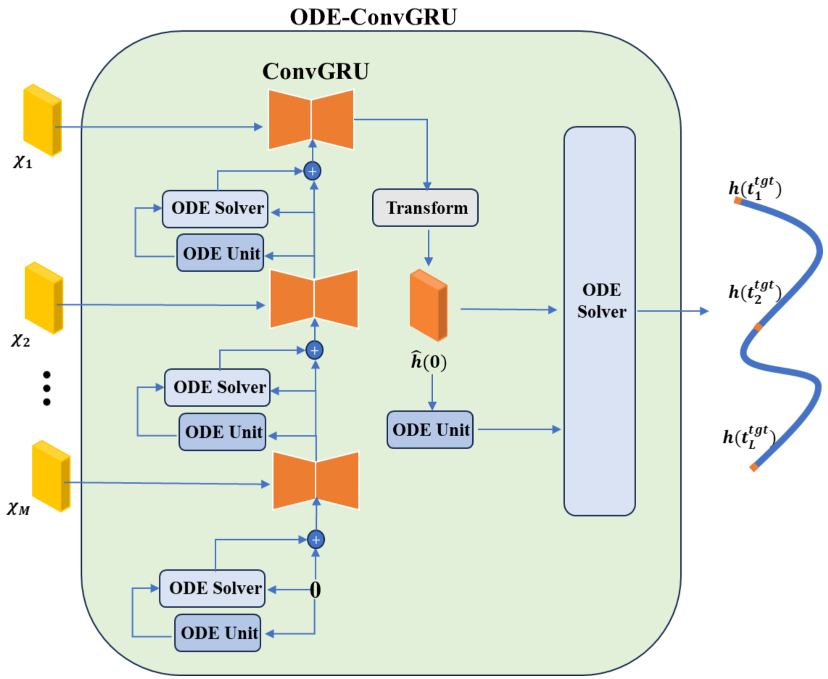
$$h(t) = \mathrm{ODE}(f, t, h_{\mathrm{in}}(t))$$

ODE Unit

$$\frac{dh_{\mathrm{in}}(t)}{dt} = f(h_{\mathrm{in}}(t), t)$$

ODE Solver

$$h(t) = h_{\mathrm{in}}(t_0) + \int_{t_0}^{t} f(h_{\mathrm{in}}(t), \tau) d\tau$$



Step 1: Initial State Update

$$\overline{h}(t_j^{src}) = \mathrm{ODE}(f_\theta, \{t_j^{src}\}, \overline{h}(t_{j-1}^{src}))$$

Step 2: Information Aggregation

$$\tilde{h}(t_j^{src}) = \mathrm{ConvGRU}(\overline{h}(t_j^{src}), \chi_{t_j^{src}})$$

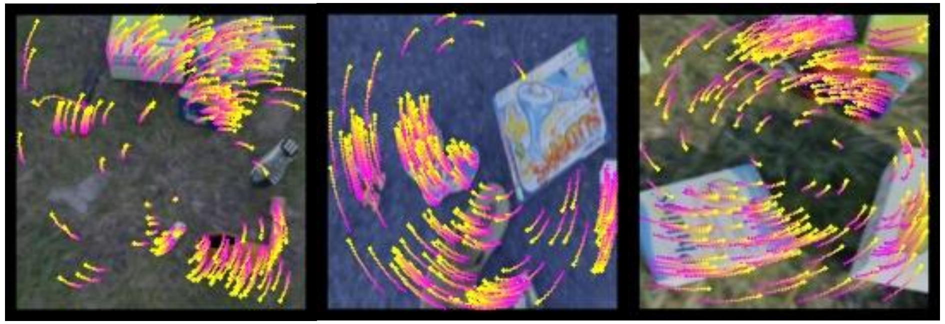
Step 3: Final State Output

$$\{h(t_j^{tgt})\}^L = \mathrm{ODE}(f_\phi, \{t_j^{tgt}\}^L, \hat{h}(0))$$

Experiments

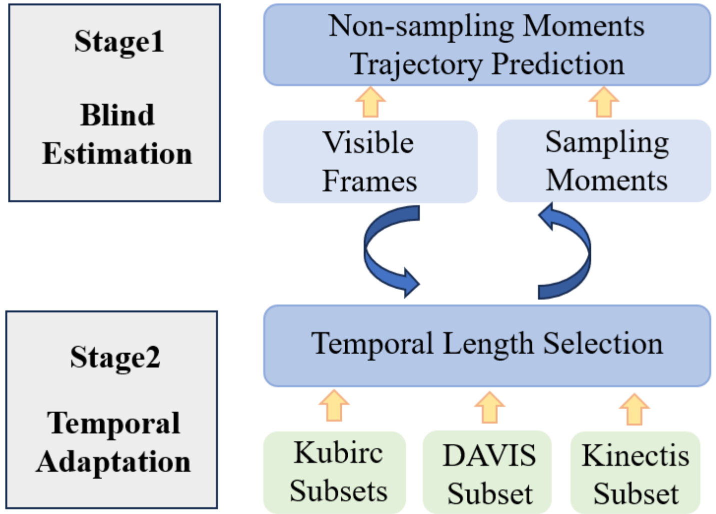
Dataset & Evaluation

☐ Synthetic Dataset



- ✓ A **large-scale** synthetic dataset based on Kubric.
- ✓ **Long-term & Dense (>80%)** GT Annotations
- ✓ **Diverse Temporal Scales.**

☐ Evaluation Perspective



Stage 1

Non-sampling prediction with a fixed timescale

Stage 2

Generalization test for multiple timescales in different dataset

Experiments

Metric & Baseline

- **Metrics**
  - ✓ **Accuracy**

$$ADE = \frac{1}{N_v} * \frac{1}{T} \sum_{\mathbf{x}} \sum ||F(t_k) - F^*(t_k)||_2$$

  - ✓ **Smoothness**

$$TRMSE = \frac{1}{N_v} \sum_{\mathbf{x}} \sqrt{\frac{1}{T} \sum_{a=0}^{T-1} ||F(t_a) - D_{gt}(t_a)||^2}$$

- **Baselines**
  - **RAFT :** optical flow estimator with sub-pixel performance.
  - **PIPs :** recent point-centric tracker with powerful generalization.
- **Process:** Discrete Mapping + Link
  - **RAFT :** chain adjacent flow with **bililnear interpolation** & **explicit motion assumption.**
  - **PIPs :** generate all mapping at sampling moments for **batch of pixels** & **explicit motion assumption.**

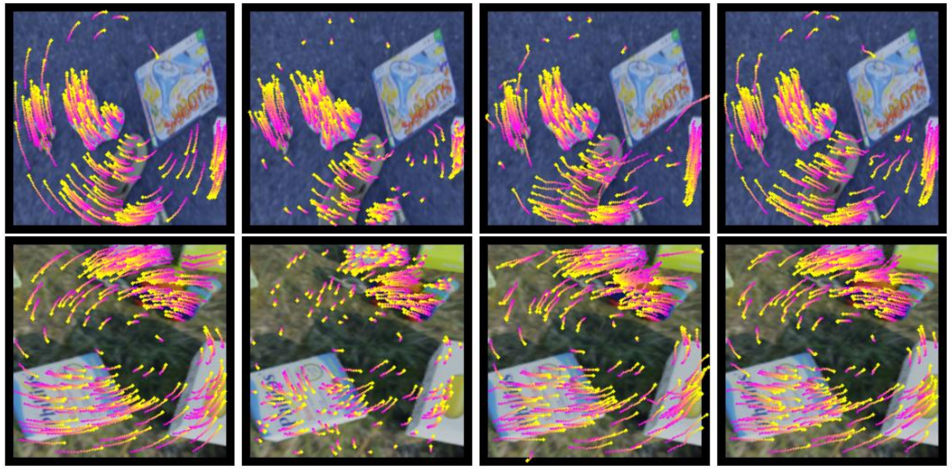
Experiments

Comparision Results on Synthetic Dataset

- Kubric Synthetic Dataset

Method	Metric	Mark	Query-Stride set				Query-First set			
			8f	10f	16f	24f	8f	10f	16f	24f
RAFT	ADE_vis	S	8.18	7.91	13.00	14.43	5.68	7.18	11.12	17.56
		NS	7.13	6.08	10.21	11.31	4.94	5.42	8.47	13.91
	ADE_Occ	S	9.62	12.16	27.73	24.95	6.26	8.60	14.77	24.86
		NS	8.79	9.37	21.98	18.06	5.60	6.43	10.72	19.04
	ADE_All	S	8.32	8.17	14.93	16.37	5.73	7.27	11.44	18.49
		NS	7.39	6.37	12.37	13.06	5.00	5.51	8.75	14.72
	TRMSE	-	8.50	7.82	15.81	16.66	5.86	6.77	10.35	16.85
PIPs	ADE_vis	S	4.66	4.39	8.77	10.13	3.12	4.12	7.81	14,23
		NS	3.97	3.40	6.92	7.92	2.97	3.26	5.97	10.27
	ADE_Occ	S	7.13	7.71	21.54	19.97	4.85	7.28	12.89	22.34
		NS	6.43	5.93	16.83	15.16	4.23	4.03	9.31	16.35
	ADE_All	S	4.74	4.58	10.41	12.09	3.76	4.93	8.35	14.68
		NS	4.13	3.61	8.68	9.86	3.25	3.76	6.42	11.76
	TRMSE	-	4.93	4.49	**11.27**	13.02	3.90	4.72	7.76	13.70
ODE-6spline (Ours)	ADE_vis	S	3.81	3.88	8.08	7.88	2.88	3.72	6.26	12.08
		NS	3.38	3.11	6.47	6.29	2.51	2.86	4.83	9.73
	ADE_Occ	S	6.24	8.09	22.72	19.93	4.70	6.86	11.34	19.61
		NS	5.65	6.35	18.21	15.09	4.17	5.04	8.24	15.19
	ADE_All	S	**4.00**	**4.13**	**9.89**	**10.31**	**2.96**	**3.88**	**6.71**	**13.05**
		NS	**3.60**	**3.37**	**8.54**	**8.66**	**2.61**	**3.02**	**5.26**	**10.64**
	TRMSE	-	**4.26**	**4.17**	11.38	**12.03**	**3.12**	**3.78**	**6.42**	**12.38**

[*] S, NS are respectively referred to as the sampling and non-sampling moments.
[**] f is an abbreviation of frames.



(a) Ground Truth (b) RAFT (c) PIPs (d) Ours

- **Query-Sride:** sample from spatial-temporal space
- **Query-First:** sample from the first frame

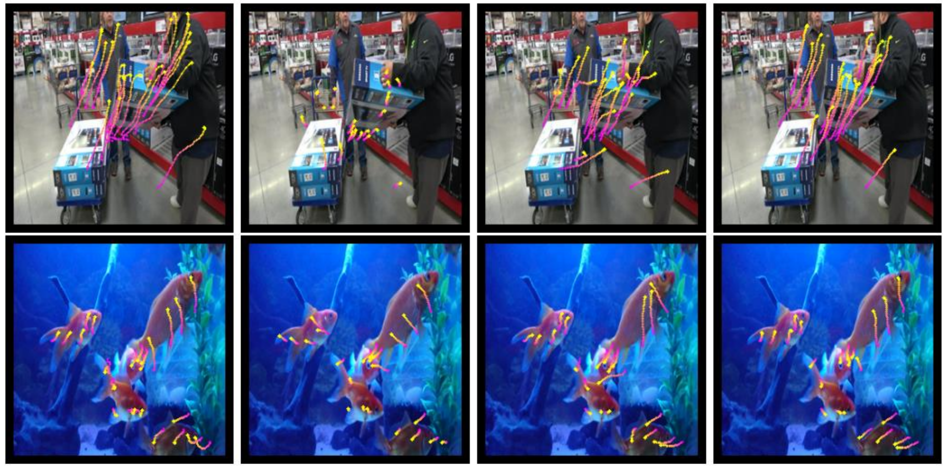
- **Dense Prediction**: improve the accuracy & smoothness for large-scale parallel inference.
- **Favourable Performance:** outperform baselines thoroughly including moments/visibility/timescales.

Experiments

Comparision Results on Synthetic Dataset

☐ Real-World Dataset

Method	Metric	Mark	Vid-DAVIS [8]				Vid-Kinetics [8]			
			20f	24f	28f	32f	36f	48f	128f	250f
RAFT	ADE_Vis	S	16.84	19.72	19.42	23.37	15.97	19.00	30.19	35.93
		NS	12.42	14.71	15.87	18.53	12.98	16.36	24.99	30.63
	ADE_Occ	S	18.58	23.82	27.91	31.35	19.23	21.89	31.98	39.76
		NS	14.67	18.56	20.19	22.53	16.07	18.85	26.11	33.10
	ADE_All	S	17.99	20.25	22.97	25.65	16.74	20.70	31.47	37.81
		NS	14.23	16.07	17.91	19.68	14.90	17.73	25.16	31.49
	TRMSE	-	16.91	18.95	21.28	23.81	16.62	24.51	27.50	34.23
PIPs	ADE_Vis	S	12.96	15.47	15.62	18.73	12.71	16.48	28.34	38.56
		NS	10.39	11.39	12.60	15.27	10.49	13.30	22.31	31.08
	ADE_Occ	S	18.84	26.67	30.60	32.21	17.59	21.08	31.82	42.31
		NS	14,42	20.66	20.88	23.26	15.42	18.56	25.34	35.41
	ADE_All	S	15.18	17.98	22.06	23.31	13.70	17.79	31.76	40.59
		NS	11.94	14.38	16.62	18.49	12.73	13.72	24.67	33.90
	TRMSE	-	14.80	17.38	20.66	22.79	12.51	15.82	28.60	37.25
ODE-6spline (Ours)	ADE_Vis	S	11.37	16.80	16.16	18.99	12.09	15.02	25.25	31.18
		NS	9.32	12.21	13.37	15.22	9.99	12.34	21.35	27.13
	ADE_Occ	S	15.75	20.67	26.71	27.99	17.76	20.50	29.79	37.12
		NS	12.24	15.91	18.00	20.55	15.17	18.00	24.98	32.00
	ADE_all	S	**13.48**	**16.91**	**19.81**	**21.16**	**12.48**	**15.82**	**27.68**	**34.42**
		NS	**10.80**	**13.31**	**15.08**	**16.68**	**10.47**	**13.10**	**22.80**	**29.65**
	TRMSE	-	**12.96**	**15.76**	**18.12**	**20.27**	**12.06**	**14.91**	**25.43**	**32.77**



(a) Ground Truth (b) RAFT [12] (c) PIPs [9] (d) Ours

- **Long-term Tracking**: keep the motion trend to suppress drift for long-term tracking.
- **Generalization Beyond Training Range:** stable tracking with 10% improvement than baselines for ultra-distance(250f) scene.

Conclusion

An Exploration of Continuous Pixel Motion Estimation

- A novel motion representation based on continuous parametric optical flow fusing with implicit feature optimization and explicit parametric modelling.
  - 1) propose **continuous parametric optical flow** to provide spatially dense & temporally continuous pixel displacements simultaneously.
  - 2) implicitly and explicitly fusion of continuous information by ODE-ConvGRU & parametric curves
  - 3) a new simulated dataset & evaluation framework
- **Our work shows that CPFlow is suitable to achieve long-term & continuous tracking for large –scale pixels built with flexible parametric modelling & implicit feature aggregation.**

- Limitation
  - 1) parametric model is easy to fail in complex motion scenes
  - 2) feature aggregation only rely on part of moments.

THANK YOU FOR WATCHING

Project Page – CPFlow
https://npucvr.github.io/CPFlow

Github Page – CPFlow
https://github.com/LuoRadisher/CPFlow