UNIVERSITY OF CAMBRIDGE

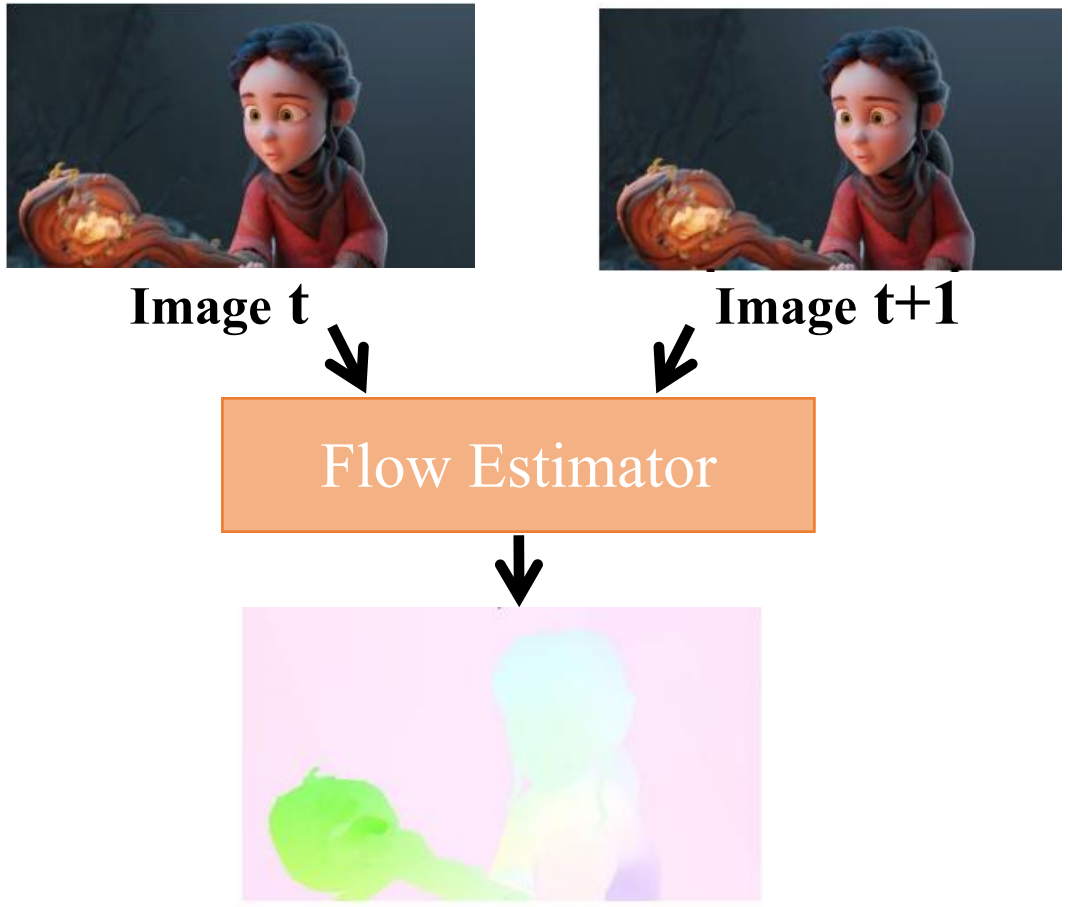
ARFlow: Auto-regressive Optical Flow Estimation for Arbitrary -Length Videos via Progressive Next-Frame Forecasting

Jiuming Liu*, Mengmeng Liu*, Siting Zhu, Yunpeng Zhang, Jiangtao Li, Michael Ying Yang, Francesco Nex, Hao Cheng, Hesheng Wang

Division F: Information Engineering

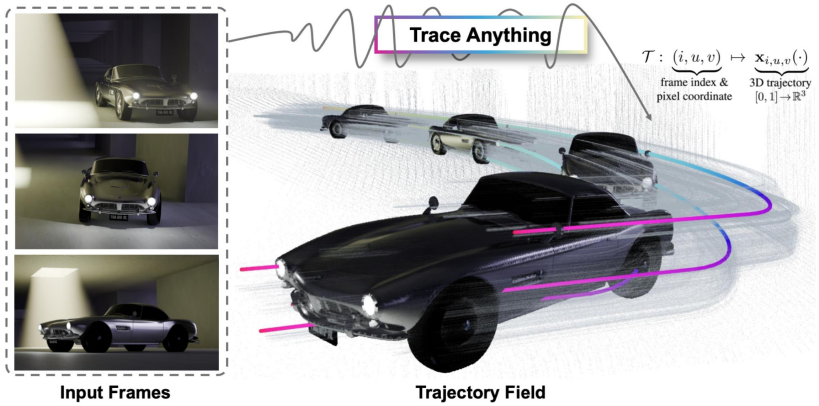
Liu, J., Liu, M., Zhu, S., Zhang, Y., Li, J., Yang, M. Y., ... & Wang, H. ARFlow: Auto-regressive Optical Flow Estimation for Arbitrary-Length Videos via Progressive Next-Frame Forecasting. In *The Fourteenth International Conference on Learning Representations*.

Task Definition

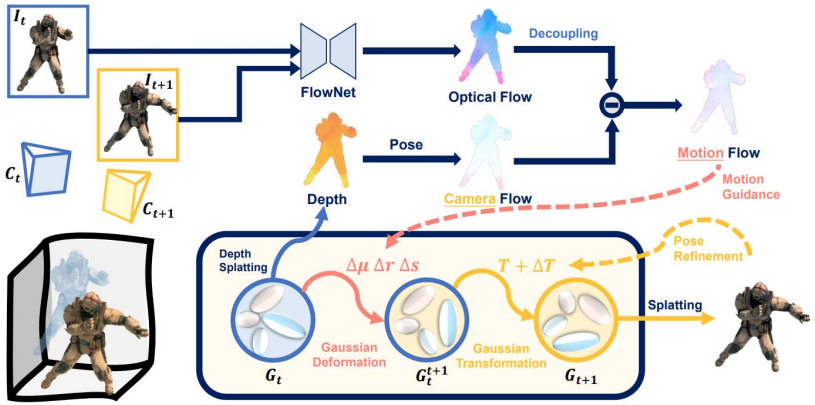


Given a pair of images, the objective of optical flow estimation is to predict per-pixel 2D displacements across two frames.

Applications



motion tracking [1]



dynamic 4D reconstruction [2]

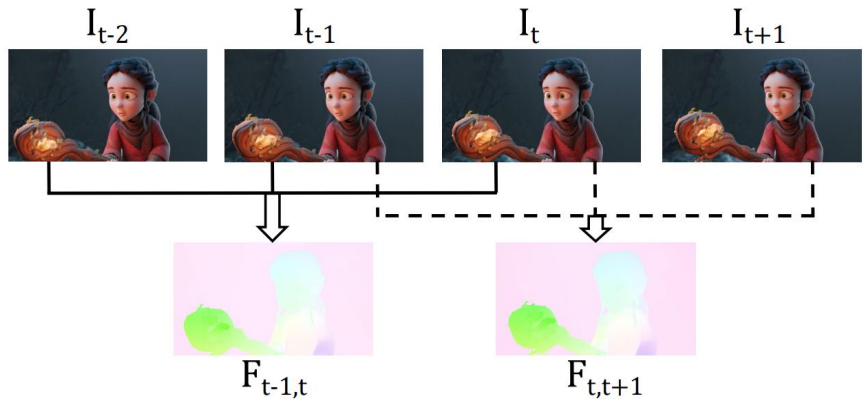
video frame interpolation [3]

video generation [4]

[1] Liu, Xinhang, et al. "Trace Anything: Representing Any Video in 4D via Trajectory Fields." arXiv preprint arXiv:2510.13802 (2025).
[2] Zhu, Ruijie, et al. "Motiongs: Exploring explicit motion guidance for deformable 3d gaussian splatting." Advances in Neural Information Processing Systems 37 (2024): 101790-101817.
[3] Huang, Zhewei, et al. "Real-time intermediate flow estimation for video frame interpolation." *European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2022.
[4] Nam, Hyelin, et al. "Optical-flow guided prompt optimization for coherent video generation." Proceedings of the Computer Vision and Pattern Recognition Conference. 2025.

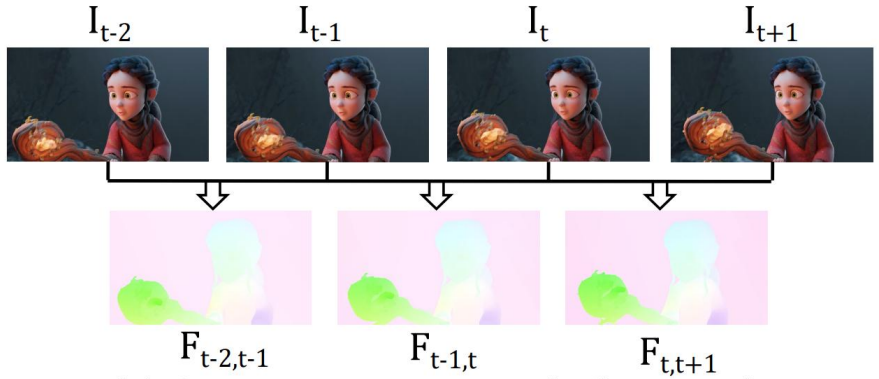
Multi-frame Optical Flow Estimation

- **Motivations:**
  - ✓ **Motion Consistency:** Object motion tends to be consistent across frames.
  - ✓ **Occlusion:** Temporal information facilitates the recovery of occluded objects.
  - ✓ Some recent works [5][6][7] also investigate **memory** mechanism, imitating human-beings.
- **Challenges:**



(A) Many-*to-one multi-frame pipeline*

repeated feature extraction and correlation



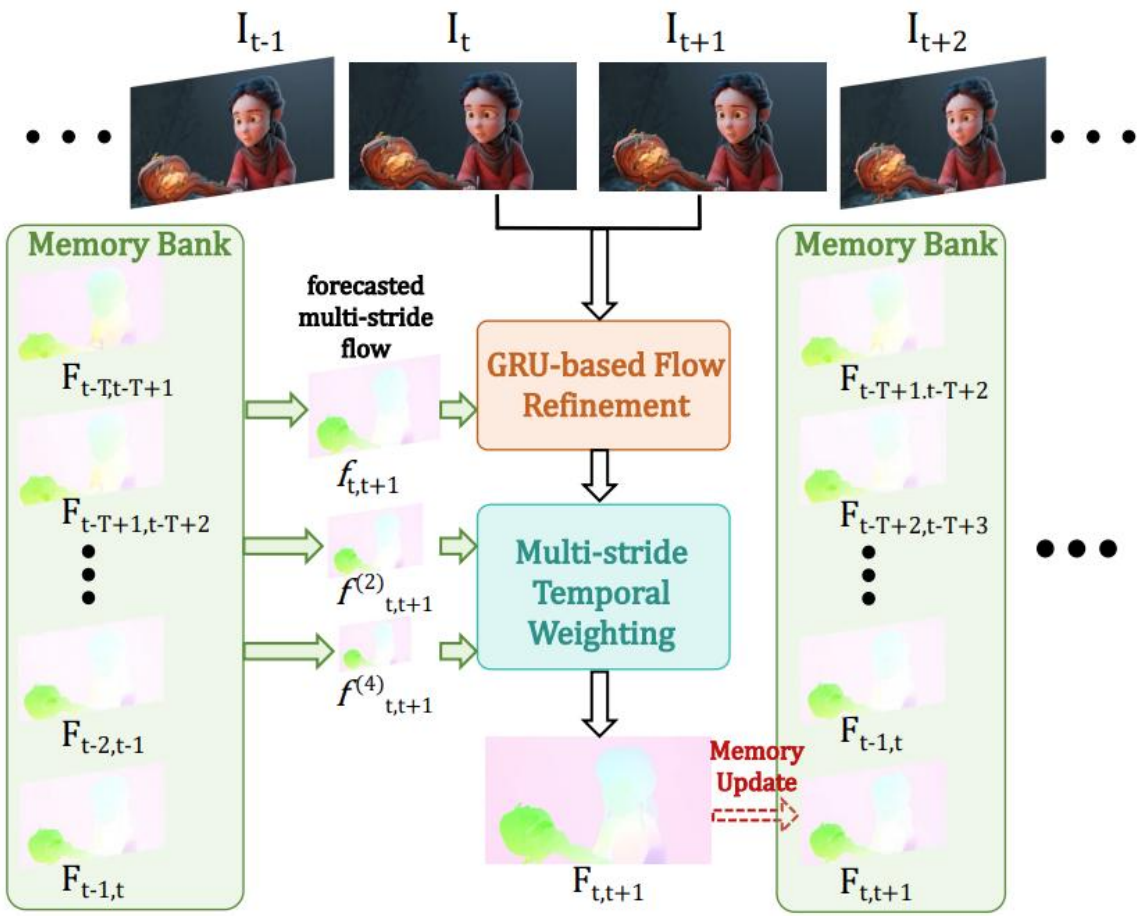
(B) *Sequence-to-sequence multi-frame pipeline*

cannot generalize to long-sequence inputs

(1) limited temporal receptive field!

(2) high computational costs and poor scalability!

Our Method



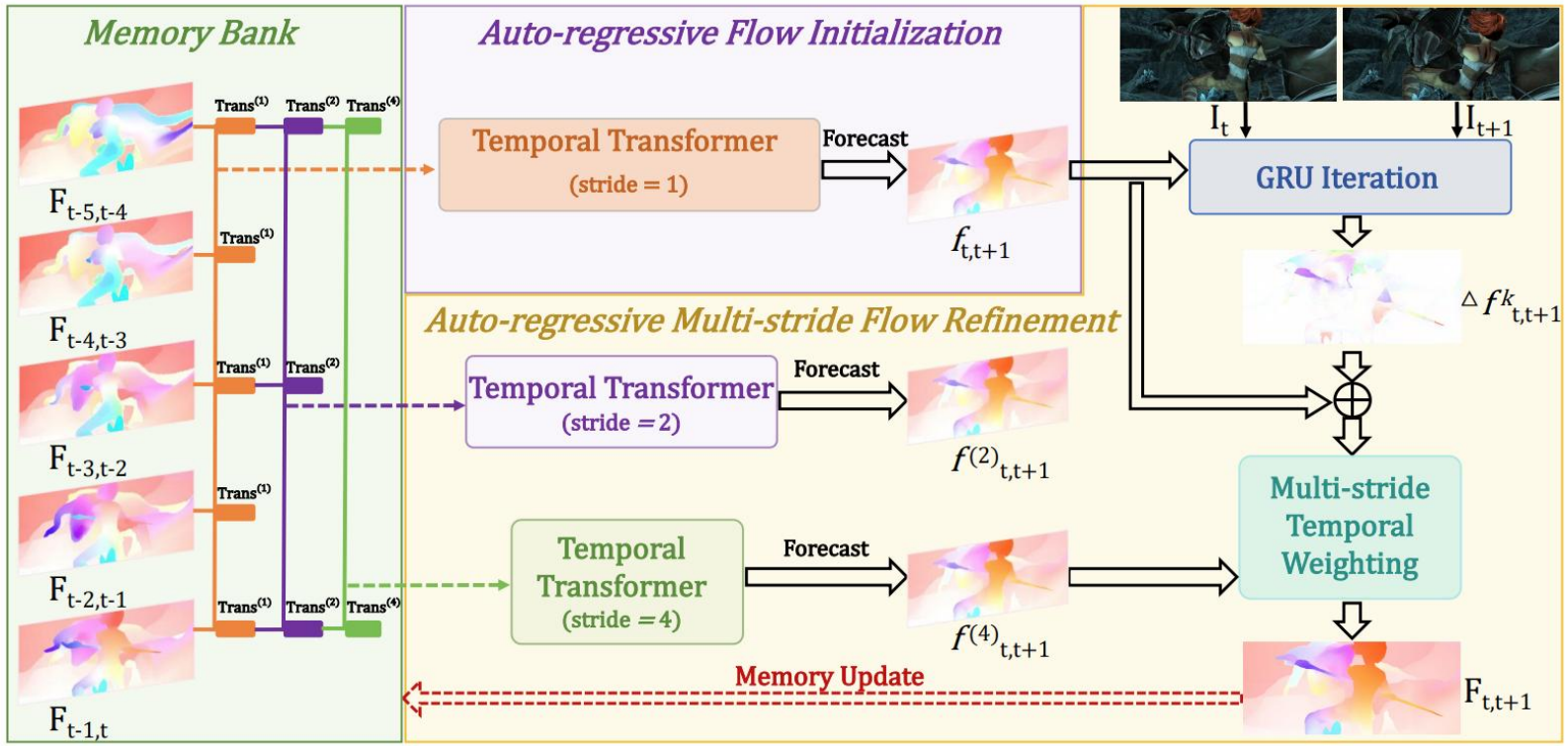
Streamlined paradigm which can scale to arbitrary-length videos with global receptive field

(1) Design a *memory bank* to dynamically store history flow estimates.

(2) Use an *auto-regressive transformer* to forecast the next frame flow.

(3) *Multi-stride forecasted flows* are weighted to consider both long-term and short-term motions.

Pipeline



(1) Auto-regressive Flow Initialization

$$\{\mathbf{Feat}_i^{(1)}\}_{i=t-T}^{t-1} = \mathrm{Trans}^{(1)}\left(\{F_{i,i+1}\}_{i=t-T}^{t-1}\right),$$

$$f_{t,t+1} = \phi(\mathbf{Feat}_{t-1}^{(1)}),$$

(2) Auto-regressive Multi-Stride Flow Refinement

$$\{\mathbf{Feat}_i^{(2)}\}_{i\in\{t-1,\,t-3,\,\dots\}} = \mathrm{Trans}^{(2)}\left(\{\mathbf{Feat}_i^{(1)}\}_{i\in\{t-1,\,t-3,\,\dots\}}\right), \quad f_{t,t+1}^{(2)} = \phi(\mathbf{Feat}_{t-1}^{(2)})$$

$$\{\mathbf{Feat}_i^{(4)}\}_{i\in\{t-1,\,t-5,\,\dots\}} = \mathrm{Trans}^{(4)}\left(\{\mathbf{Feat}_i^{(2)}\}_{i\in\{t-1,\,t-5,\dots\}}\right), \quad f_{t,t+1}^{(4)} = \phi(\mathbf{Feat}_{t-1}^{(4)})$$

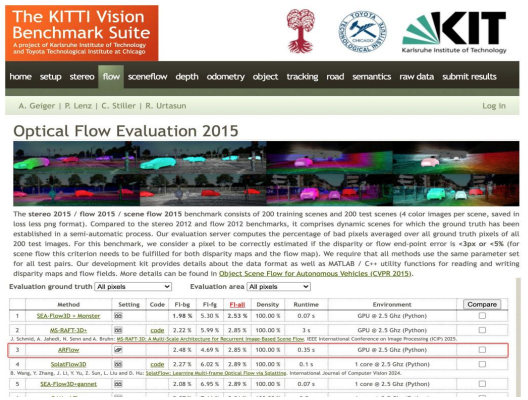
$$\{\mathbf{Feat}_{\text{fuse}}^{(l)}\}_{l\in\{1,\,2,\,4\}} = \mathrm{Trans}^{(f)}\left(\mathbf{Feat}_{t-1}^{(1)}, \mathbf{Feat}_{t-1}^{(2)}, \mathbf{Feat}_{t-1}^{(4)}\right), \quad f_{\text{fuse}} = \phi(\mathbf{Feat}_{\text{fuse}}^{(4)}),$$

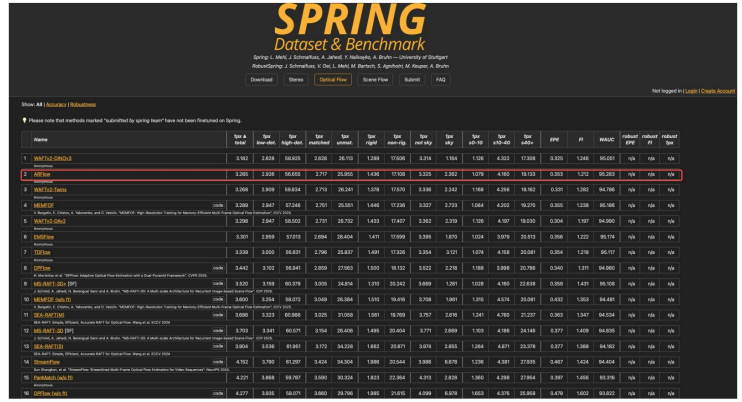
Quantitative Results

	Method	Reference	Sintel Clean			Sintel Final			KITTI-15	
			Mat. ↓	Unm. ↓	All ↓	Mat. ↓	Unm. ↓	All ↓	All ↓	Non-Occ ↓
Pairwise	FlowNet2 (Ilg et al., 2017b)	CVPR'17	1.56	25.40	4.16	2.75	30.11	5.74	10.41	6.94
	PWC-Net (Sun et al., 2018b)	CVPR'18	1.45	23.47	3.86	2.44	27.08	5.04	9.60	6.12
	RAFT (Teed & Deng, 2020)	ECCV'20	0.62	9.65	1.61	1.41	14.68	2.86	5.10	3.07
	GMFlow (Xu et al., 2022)	CVPR'22	0.65	10.56	1.74	1.32	15.80	2.90	9.32	3.80
	FlowFormer (Huang et al., 2022)	CVPR'22	0.42	7.16	1.16	0.96	11.30	2.09	4.68	2.69
	GMFlow+ (Xu et al., 2023b)	TPAMI'23	**0.34**	6.68	1.03	1.10	12.74	2.37	4.49	2.40
	Flowformer++ (Shi et al., 2023b)	CVPR'23	0.39	6.64	1.07	0.88	10.63	1.94	4.52	-
	AnyFlow (Jung et al., 2023)	CVPR'23	0.42	7.68	1.21	1.12	13.37	2.44	4.41	2.69
	CroCoFlow (Weinzaepfel et al., 2023)	ICCV'23	0.39	6.85	1.09	1.21	12.42	2.44	3.64	2.40
	DDVM (Saxena et al., 2023)	NeurIPS'23	0.83	9.26	1.75	1.28	12.20	2.48	3.26	2.24
	FlowDiffuser (Luo et al., 2024)	CVPR'24	0.38	6.23	1.02	0.97	10.67	2.03	4.17	2.82
	SEA-RAFT(L) (Wang et al., 2024b)	ECCV'24	0.44	8.40	1.31	1.20	14.06	2.60	4.30	-
	SAMFlow (Zhou et al., 2024)	AAAI'24	0.38	5.97	1.00	1.04	10.60	2.08	4.49	-
	DPFlow (Morimitsu et al., 2025)	CVPR'25	0.39	6.36	1.04	0.91	10.69	1.97	3.56	2.12
	CEDFlow++ (Zuo et al., 2025)	IJCV'25	-	-	1.37	-	-	2.40	4.78	-
	WAFT (Wang & Deng, 2025)	Arxiv'25	-	-	1.09	-	-	2.34	3.42	2.04
Multi-frame	MFCFlow (Chen et al., 2023)	WACV'23	0.65	8.34	1.49	1.33	12.81	2.58	5.00	-
	TransFlow (Lu et al., 2023)	CVPR'23	0.36	6.77	1.06	0.99	10.96	2.08	4.32	-
	SplatFlow (Wang et al., 2024a)	IJCV'24	0.51	6.06	1.12	1.06	10.29	2.07	4.61	2.96
	StreamFlow (Sun et al., 2024)	NeurIPS'24	0.38	6.42	1.04	0.82	10.44	1.87	4.24	2.45
	MemFlow (Dong & Fu, 2024)	CVPR'24	0.43	6.09	1.05	0.93	9.93	1.91	4.10	2.56
	MEMFOF (Bargatin et al., 2025)	ICCV'25	0.40	**5.59**	**0.96**	0.88	10.30	1.91	2.94	1.97
	ARFlow (Ours)	—	0.39	5.64	**0.96**	**0.81**	**9.79**	**1.78**	**2.85**	**1.91**

	Method	#Frames	Inference Cost (1080p)		Spring (test)			
			Memory, GB	Runtime, ms	1px ↓	EPE ↓	Fl ↓	WAUC ↑
NO FINE-TUNE	Flow1D (Xu et al., 2021)	2	1.34	405	-	-	-	-
	MeFlow (Xu et al., 2023a)	2	1.32	1028	-	-	-	-
	PWC-Net (Sun et al., 2018b)	2	1.41	76	82.265	2.288	4.889	45.670
	FlowNet2 (Ilg et al., 2017b)	2	4.16	167	6.710	1.040	2.823	90.907
	RAFT (Teed & Deng, 2020)	2	7.97	557	6.790	1.476	3.198	90.920
	RAFT3D* (Teed & Deng, 2021)	2	-	-	13.962	2.528	6.889	81.267
	GMA (Jiang et al., 2021)	2	13.26	1185	7.074	0.914	3.079	90.722
	GMFlow (Xu et al., 2022)	2	-	151	10.355	0.945	2.952	82.337
	FlowFormer (Huang et al., 2022)	2	OOM	-	6.510	0.723	2.384	91.679
	RPKNet (Morimitsu et al., 2024)	2	8.49	295	4.809	0.657	1.756	92.638
	Win-Win (Leroy et al., 2024)	2	-	-	5.371	0.475	1.621	92.720
	MS-RAFT+ (Jahedi et al., 2024)	2	-	-	5.724	0.643	2.189	92.888
	MatchAttention (Yan et al., 2025)	2	14.34	755	4.584	0.453	1.505	93.389
	MS-RAFT-3D* (Schmid et al., 2025)	2	-	-	3.520	0.359	1.431	95.108
	M-Fuse (F)* (Mehl et al., 2023a)	3	-	-	20.374	2.948	8.791	76.550
	VideoFlow-BOF (Shi et al., 2023a)	3	17.74	1648	-	-	-	-
	VideoFlow-MOF (Shi et al., 2023a)	5	OOM	-	-	-	-	-
	MemFlow (Dong & Fu, 2024)	3	8.08	885	5.759	0.627	2.114	92.253
	StreamFlow (Sun et al., 2024)	4	18.97	929	5.215	0.606	1.856	93.253
	MEMFOF (Bargatin et al., 2025)	3	2.09	472	3.600	0.432	1.353	94.481
	ARFlow (Ours)	8	2.10	403	**3.587**	**0.428**	**1.313**	**94.501**
FINE-TUNE	CrocoFlow (Weinzaepfel et al., 2023)	2	2.01	6524	4.565	0.498	1.508	93.660
	SEA-RAFT (S) (Wang et al., 2024b)	2	8.15	205	3.904	0.377	1.389	94.182
	SEA-RAFT (M) (Wang et al., 2024b)	2	8.19	286	3.686	0.363	1.347	94.534
	MemFlow (Dong & Fu, 2024)	3	8.08	885	4.482	0.471	1.416	93.855
	StreamFlow (Sun et al., 2024)	4	18.97	929	4.152	0.467	1.424	94.404
	DPFlow (Morimitsu et al., 2025)	2	10.39	990	3.442	**0.340**	1.280	94.663
	MEMFOF (Bargatin et al., 2025)	3	2.09	472	3.289	0.355	1.238	95.186
	ARFlow (Ours)	8	2.10	403	**3.265**	0.353	**1.212**	**95.283**

Ranking the 1st on the KITTI benchmark!
Ranking the 1st on the Spring benchmark!
Ranking the 2nd on the Sintel benchmark!

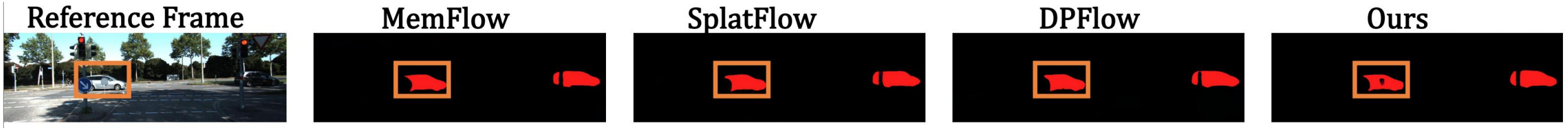




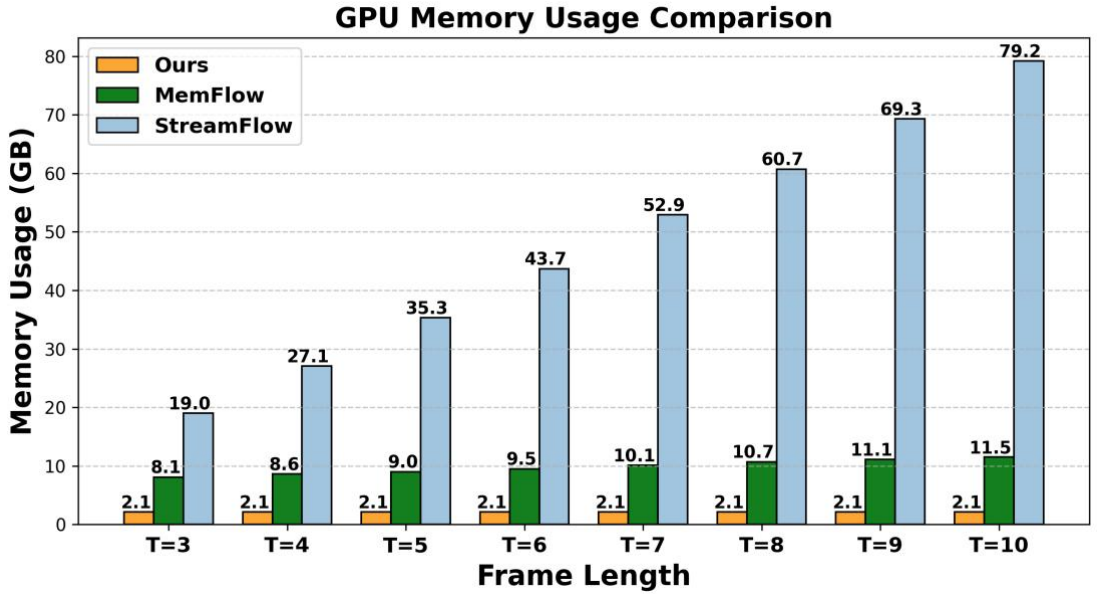
UNIVERSITY OF CAMBRIDGE

Qualitative Results

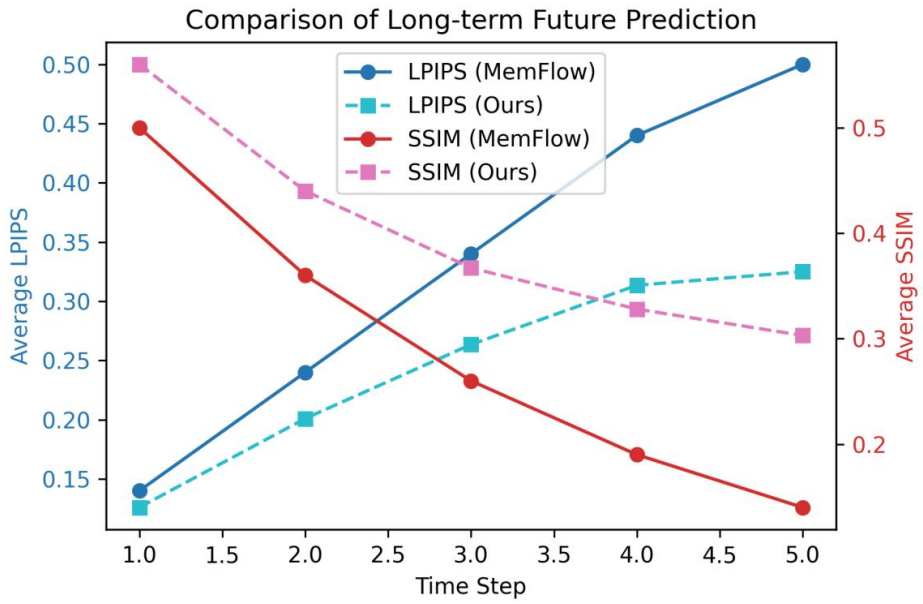
- **Better estimation on occlusions:**



- **Consistent GPU memory usage:**

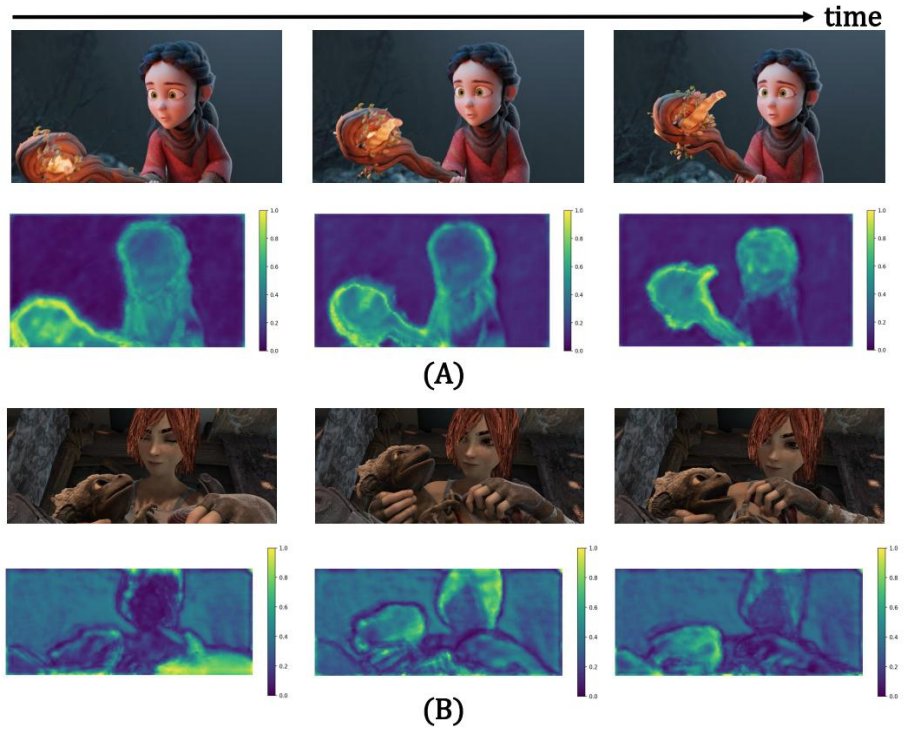


- **Future flow prediction:**



Visualizations

- **Attention weights:**



Higher weights on motion boundaries

- **Zero-shot Generalization on Waymo:**

Visualizations

- **Zero-shot Generalization on nuScenes:**

UNIVERSITY OF CAMBRIDGE

Thanks for watching!

Jiuming Liu*, Mengmeng Liu*, Siting Zhu, Yunpeng Zhang, Jiangtao Li, Michael Ying Yang, Francesco Nex, Hao Cheng, Hesheng Wang

Division F: Information Engineering