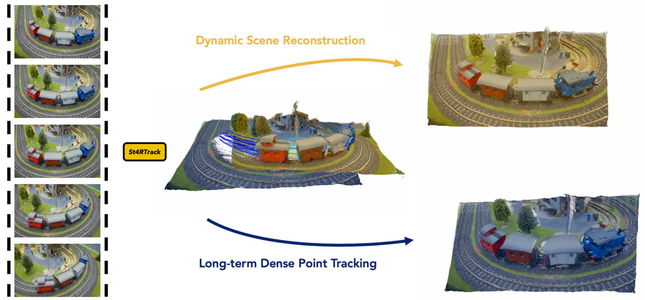
St4RTrack: Simultaneous 4D Reconstruction and Tracking in the World

Haiwen Feng*[1,2], Junyi Zhang*[1], Qianqian Wang[1], Yufei Ye[3], Pengcheng Yu[2], Michael J. Black[2], Trevor Darrell[1], Angjoo Kanazawa[1]
[1]UC Berkeley [2]Max Planck Institute for Intelligent Systems, Tübingen [3]Stanford University
(*: Equal contribution, listed alphabetically)

BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

CVPR Nashville JUNE 11-15, 2025

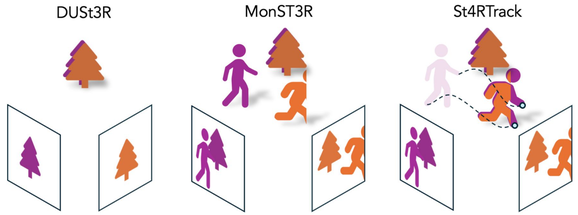
Overview



Given an RGB video capturing dynamic scenes, ***St4RTrack*** simultaneously
1) ***tracks*** the points from the initial frame (in blue) and
2) ***reconstructs*** the geometry of the subsequent frames (in yellow) directly in a *consistent world coordinate*.

Unified 4D Representation

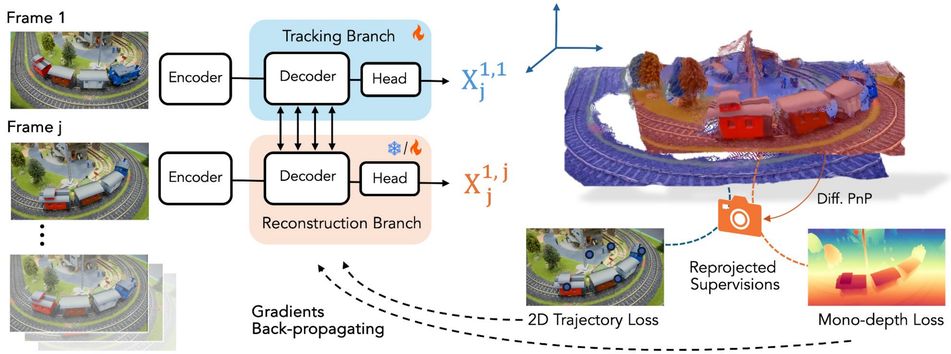
${}^{a}X_{t}^{b}$: the 3D pointmap of physical content from frame b, at time t, expressed in the coordinate system established by frame a.



$f_\theta(I_1, I_j) = {}^{1}X_1^1, {}^{1}X_1^j.$ $f_\theta(I_1, I_j) = {}^{1}X_1^1, {}^{1}X_j^j,$ $f_\theta(I_1, I_j) = {}^{1}X_j^1, {}^{1}X_j^j.$

St4RTrack estimates two pointmaps at the same timestamp.
It predicts how the points in the first frame move to the j frame, and reconstructs the geometry of the j frame.

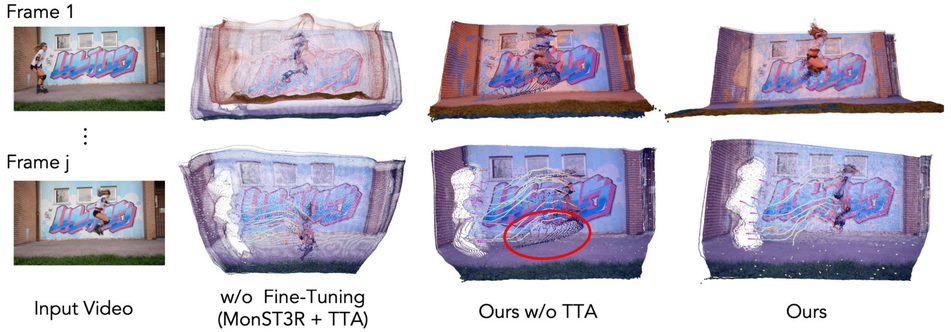
Simultaneous 4D Reconstruction and Tracking



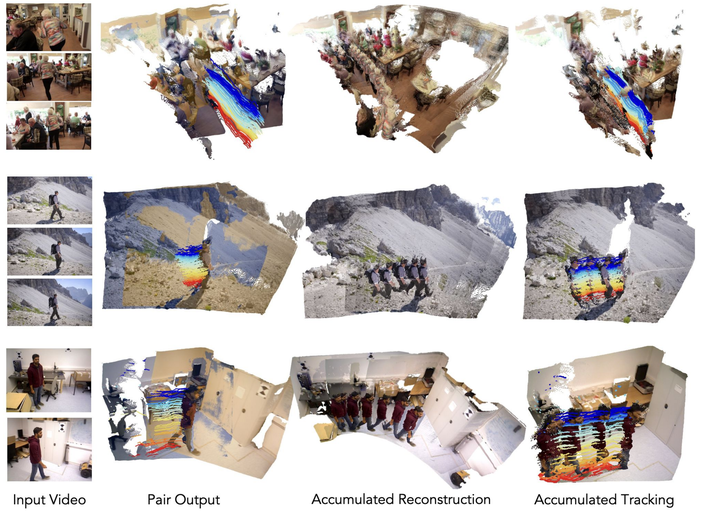
We train our network on three synthetic datasets and found that training on small-scale, sparse labeled, and unrealistic data is sufficient for our network to learn the newly proposed representation.

Adapt to Any Video without 4D Label

However, synthetic training alone still **limits generalization** to complex scenes!
We address this by leveraging the inherent geometric and motion in the representation to adapt to any video without 4D labels, using only reprojected supervision signals.



Qualitative Results



Quantitative Results

Table 1. **World Coordinate 3D Point Tracking.** We report the performance of average points under distance (APD_{3D}) after global median alignment. We evaluate the accuracy of both all points and dynamic points. The best results are **bold**. See Appendix B.2 for more results.

Category	Methods	All Points				Dynamic Points			
		PO	DR	ADT	PStudio	PO	DR	ADT	PStudio
Combinational	SpaTracker+RANSAC-Procrustes	44.03	55.01	50.87	52.05	53.77	58.58	66.49	52.05
	SpaTracker+MonST3R	47.65	55.49	51.95	50.16	58.61	59.21	69.94	50.16
Feed-forward	MonST3R	33.47	58.06	74.35	51.32	39.36	51.86	67.92	51.32
	SpaTracker	38.54	54.85	45.65	62.59	51.20	58.65	67.65	62.59
	St4RTrack (Ours)	**67.95**	**73.74**	**76.01**	**69.67**	**68.72**	**68.13**	**75.34**	**69.67**

Category	Methods	Point Odyssey		TUM-Dynamics	
		EPE↓	APD↑	EPE↓	APD↑
w/ Global Align.	DUSt3R+GA	0.6088	43.90	0.3147	70.49
	MASt3R+GA	0.4030	60.44	0.5186	68.38
	MonST3R+GA	0.2629	72.31	0.3429	63.87
Feed-forward	DUSt3R	0.6386	45.79	0.2891	72.26
	MASt3R	0.4644	56.90	0.5510	66.22
	MonST3R	0.3044	68.25	0.3646	61.38
	St4RTrack (Ours)	**0.2406**	**78.73**	**0.1854**	**83.42**

Category	Methods	Point Odyssey		TUM-Dynamics	
		EPE↓	APD↑	EPE↓	APD↑
w/ Global Align.	DUSt3R+GA	0.3541	62.42	0.2989	69.23
	MASt3R+GA	0.3717	61.31	0.5294	49.81
	MonST3R+GA	**0.2601**	69.31	0.3173	66.00
Feed-forward	DUSt3R	0.4251	56.70	0.3092	67.48
	MASt3R	0.4473	55.09	0.5862	45.43
	MonST3R	0.3462	62.10	0.3508	62.83
	St4RTrack	0.2741	**69.53**	**0.2413**	**74.14**

ArXiv: https://arxiv.org/abs/2504.13152 Webpage: St4RTrack.github.io