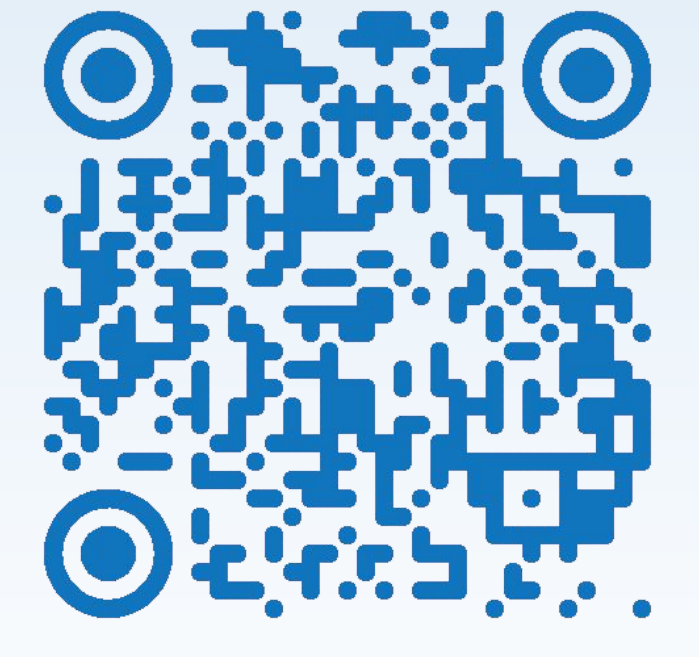


Flow4R: Unifying 4D Reconstruction and Tracking with Scene Flow

Shenhan Qian^{1,2}, Ganlin Zhang^{1,2}, Shangzhe Wu^{3,†}, Daniel Cremers^{1,2,†}

¹Technical University of Munich, ²MCML, ³University of Cambridge



Project Page



Task

Feedforward 4D reconstruction and tracking with a two-view vision transformer.

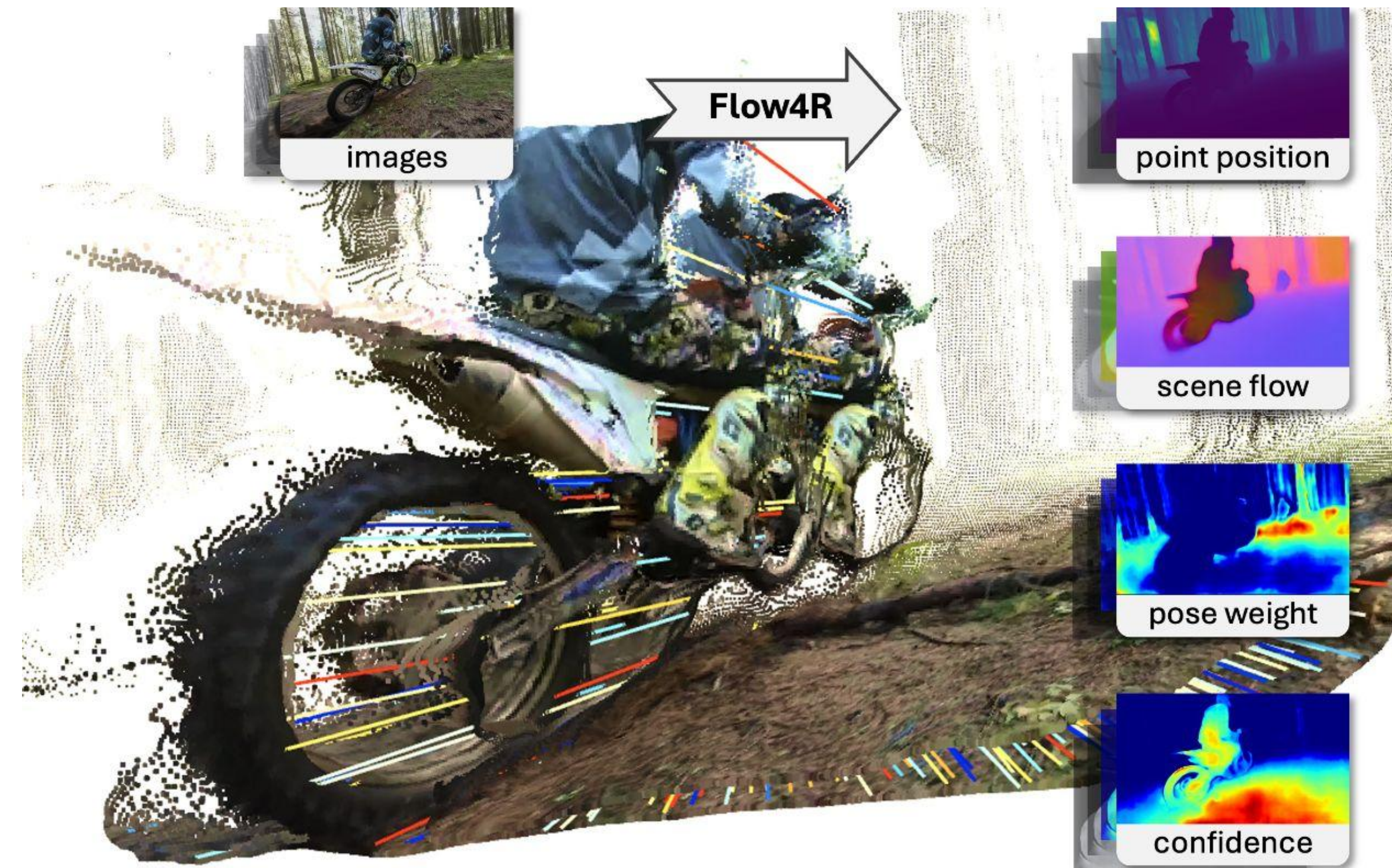
Motivation

Motion is relative — so should the architecture.

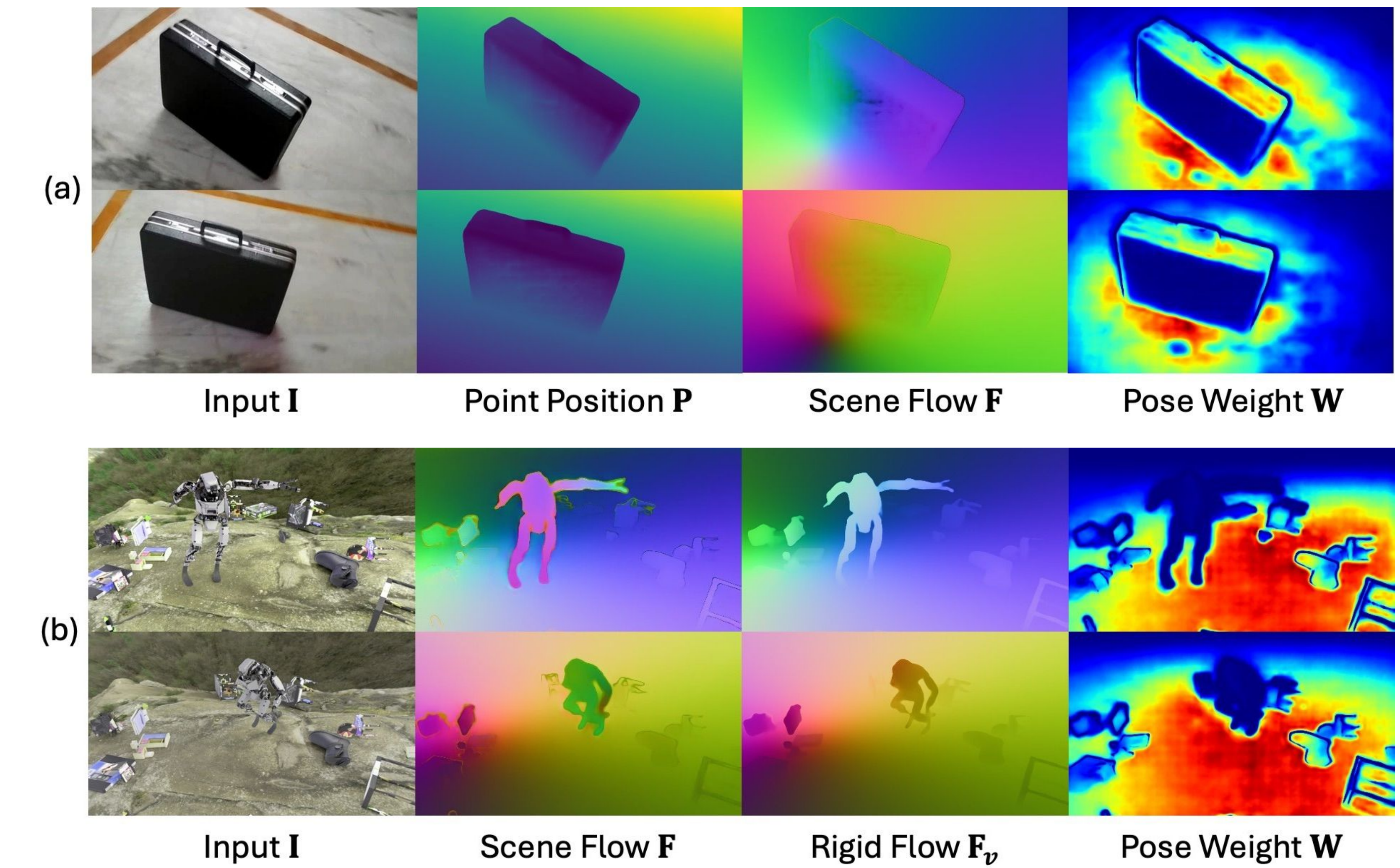
Coordinate system is arbitrary, but scene flow is invariant.

Features

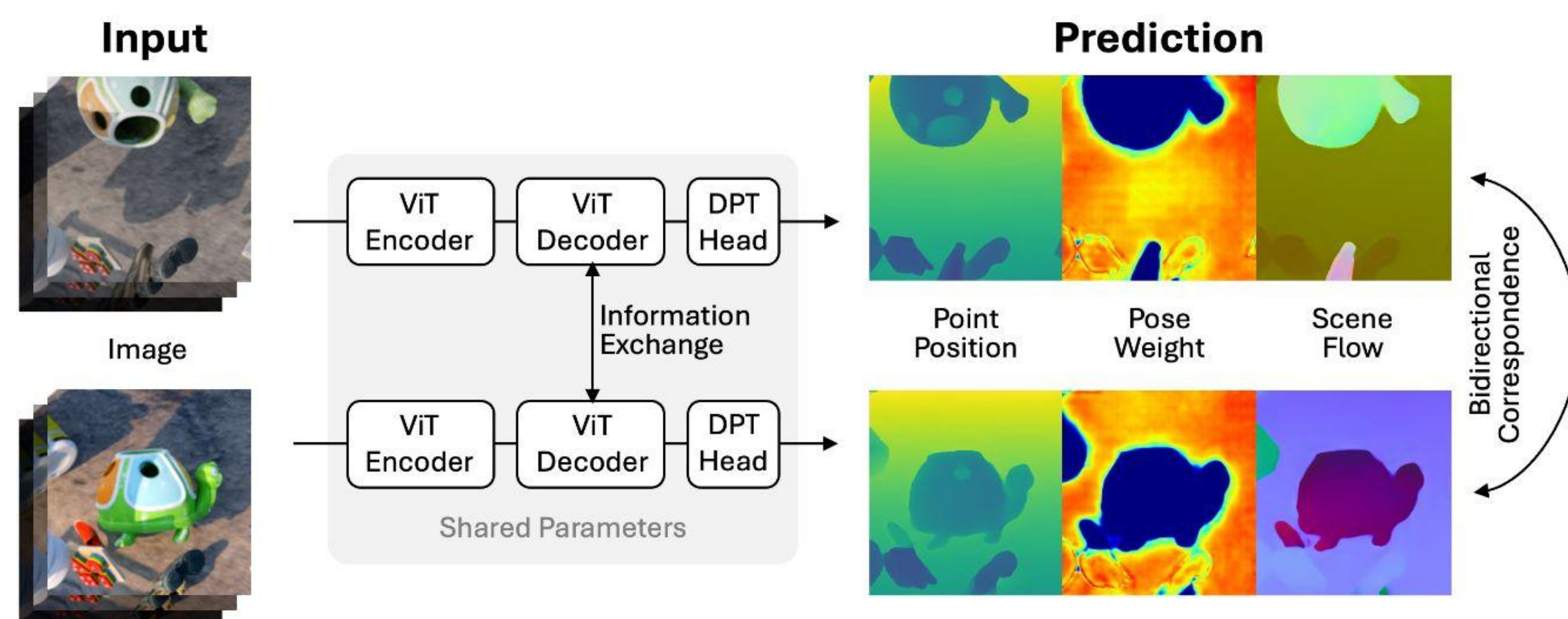
	tracking	bidirectional	symmetry	egocentric
MonST3R, CUT3R				
D ² USt3R, POMATO, St4RTrack	✓			
DPM	✓	✓		
Flow4R	✓	✓	✓	✓



2D Visualization



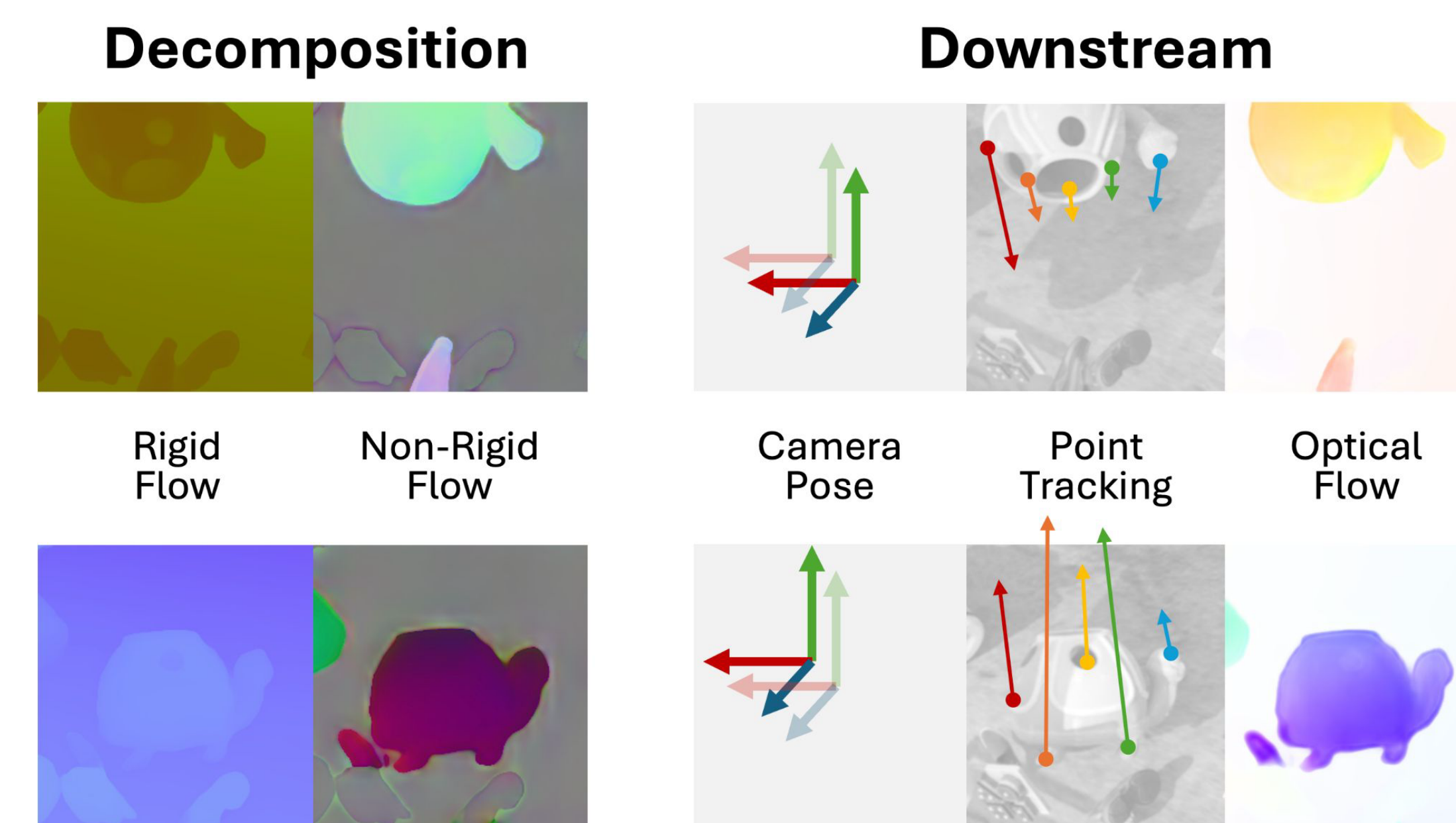
Symmetric Architecture



Local point	$\mathbf{P}$	Confidence	$\mathbf{C}$
Displaced point	$\mathbf{P}_{vt}$	Pose weight	$\mathbf{W}$

Scene flow $\mathbf{F} = \mathbf{P}_{vt} - \mathbf{P}$

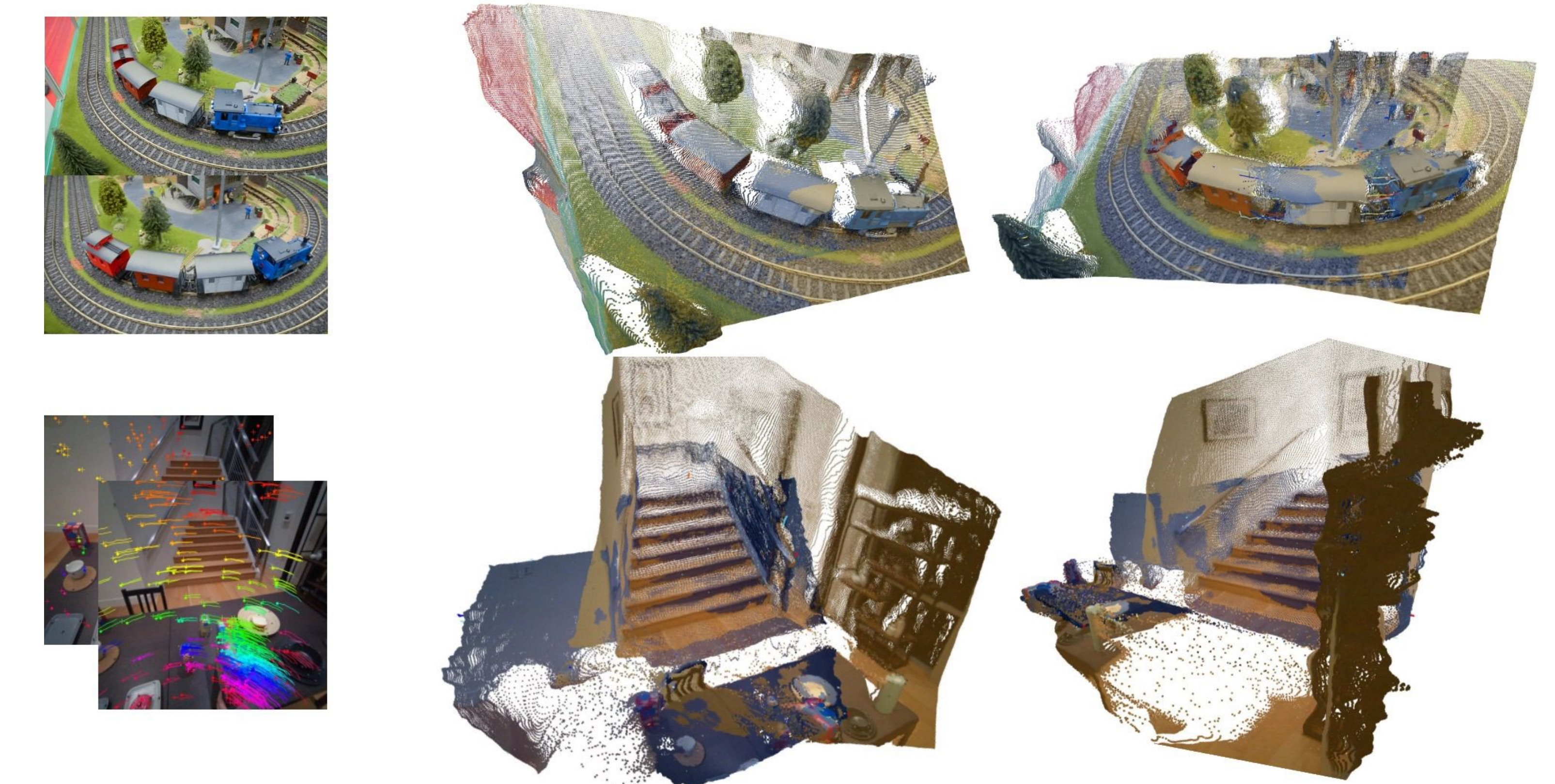
Bidirectional Correspondence



Camera pose $\hat{\mathbf{T}} = \arg \min_{\mathbf{T} \in SE(3)} \sum_{i=1}^{HW} \mathbf{W}^i \|\mathbf{P}_{vt}^i - \mathbf{T}\mathbf{P}^i\|_2$

Reconstruction $\mathbf{P}_v = \hat{\mathbf{T}} \mathbf{P}$
Tracking $\mathbf{P}_t = \hat{\mathbf{T}}^{-1} \mathbf{P}_{vt}$

3D Visualization



3D Point Tracking Performance on WorldTrack

Method	All Points				Dynamic Points			Avg.	# Param. (B)
	ADT	DR	PO	PS	ADT	DR	PO		
MonST3R	74.4	58.1	33.5	51.3	67.9	51.9	39.4	53.8	0.7
SpaTracker	45.7	54.9	38.5	62.6	67.7	58.7	51.2	54.2	0.2
POMATO	57.2	68.4	49.7	<u>64.9</u>	78.1	62.7	58.1	62.7	0.7
St4RTrack	<u>76.0</u>	<u>73.7</u>	<u>68.0</u>	69.7	<u>75.3</u>	<u>68.1</u>	<u>68.7</u>	<u>71.4</u>	0.7
Flow4R	78.6	78.5	71.1	64.3	70.9	77.2	72.9	73.4	<u>0.4</u>

Scene flow as the central representation to bridge geometry and motion.

53%
less VRAM
vs St4RTrack

26.8
pairs / second
with RTX PRO 6000

73.4
average APD
on WorldTrack