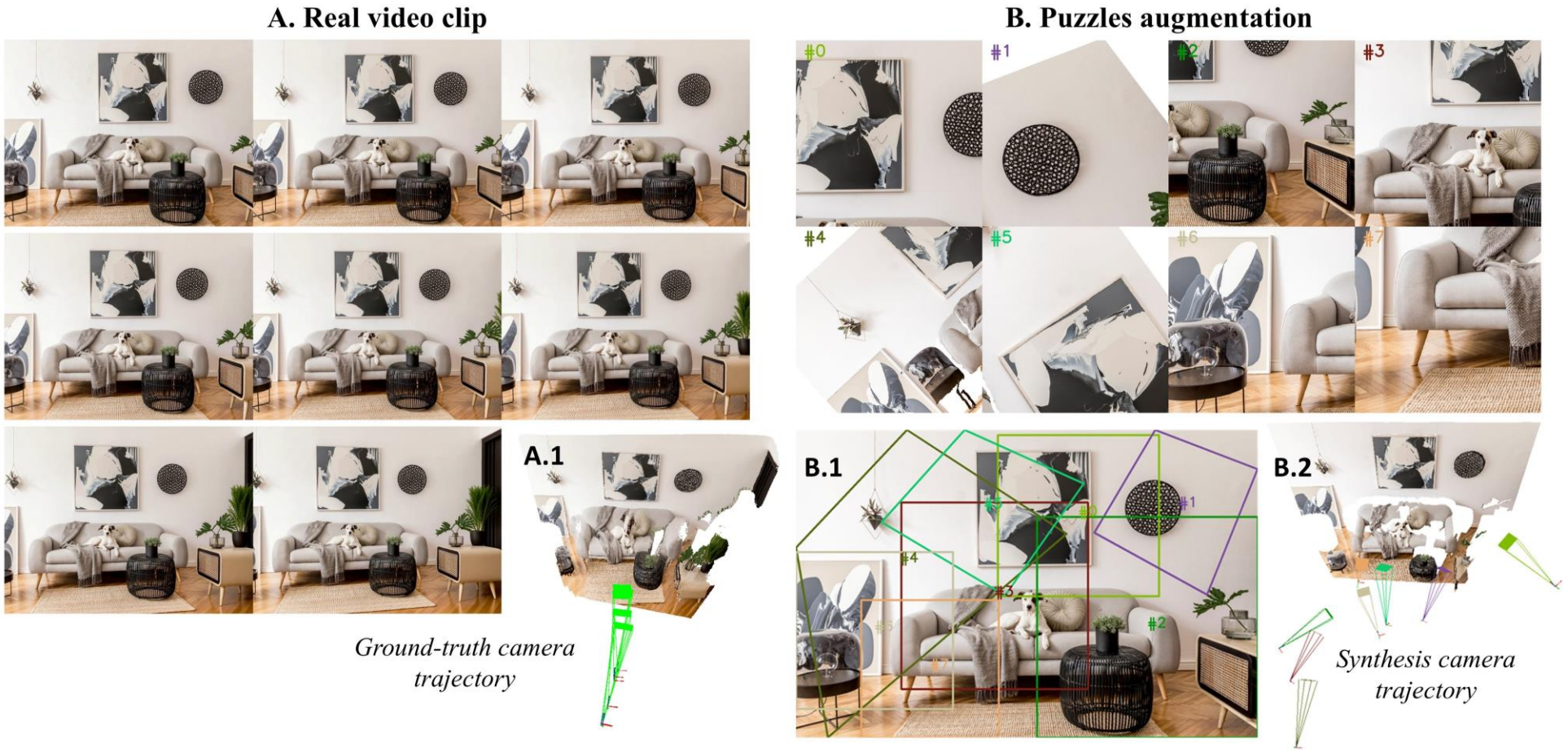
Puzzles: Unbounded Video-Depth Augmentation for Scalable End-to-End 3D Reconstruction

NeurIPS 2025

Jiahao Ma[1,3], Lei Wang[2], Miaomiao Liu[1], David Ahmedt–Aristizabal[3] , Chuong Nguyen[3]

Australian National University[1], Griffith University[2] , CSIRO DATA61[3]

Quick Preview - Research problem & Motivation

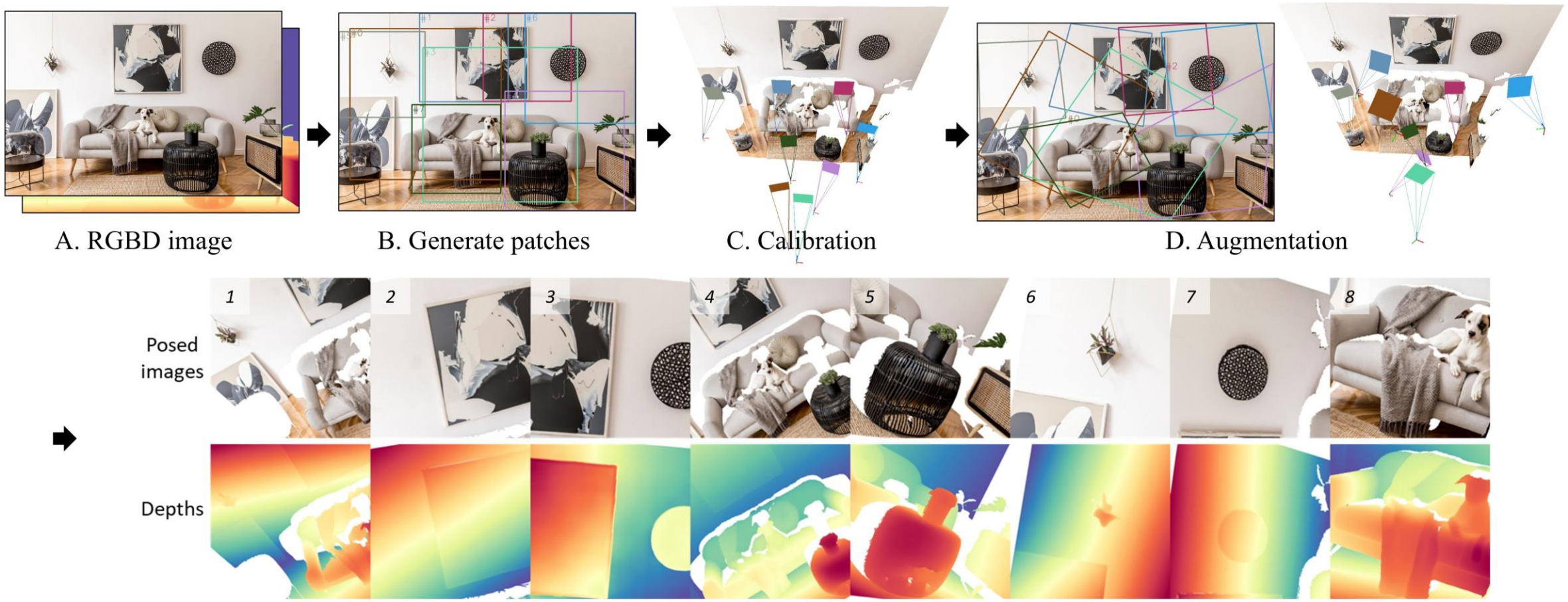


- **Task:** Feedforward 3D reconstruction: given **unposed, uncalibrated images**, predicts **global point maps**.
- **Challenge:** Existing methods need massive posed video; limited **diversity/scale** hurts **robustness/generalization**.
- **Solution (Puzzles):** Plug-and-play augmentation that makes **ordered, overlapping, video-like sub-clips** from images/clips, boosting diversity without changing the model.

Quick Preview

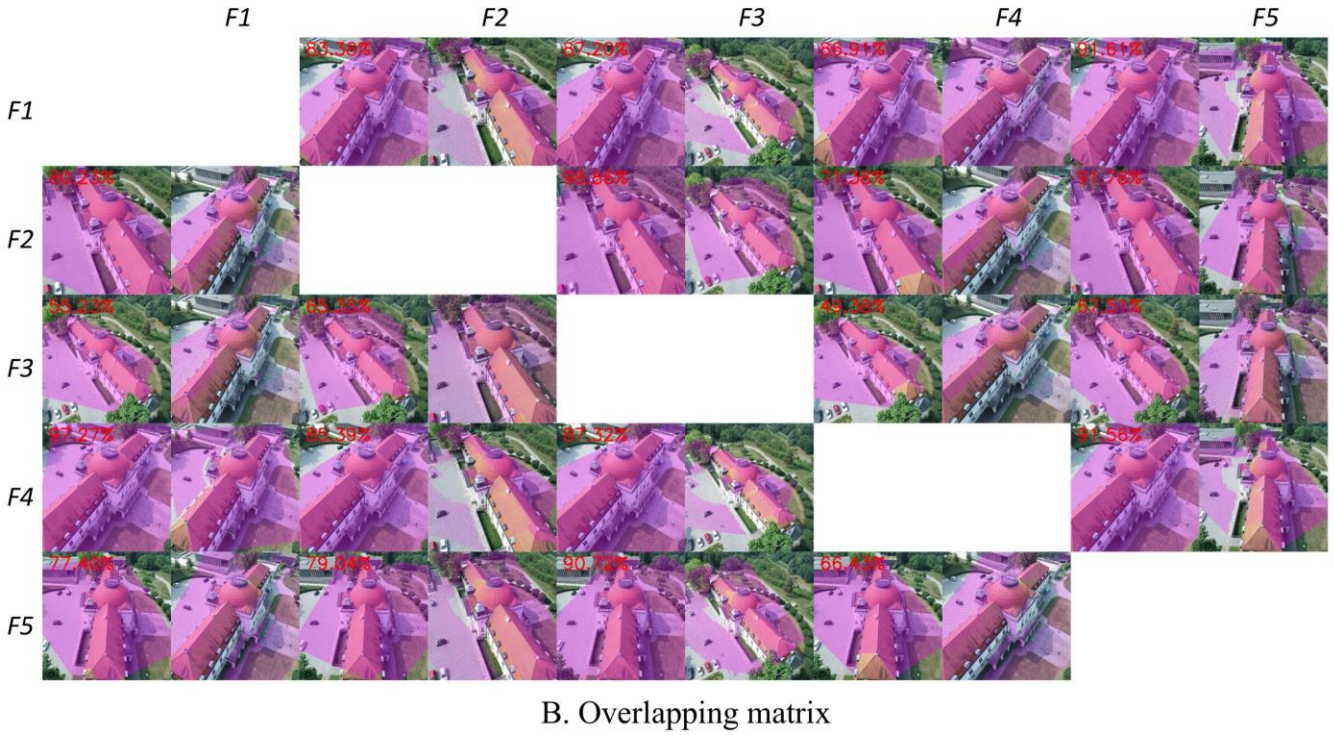
Puzzles: Unbounded Video-Depth Augmentation for Scalable End-to-End 3D Reconstruction

Method – Image-to-Clips

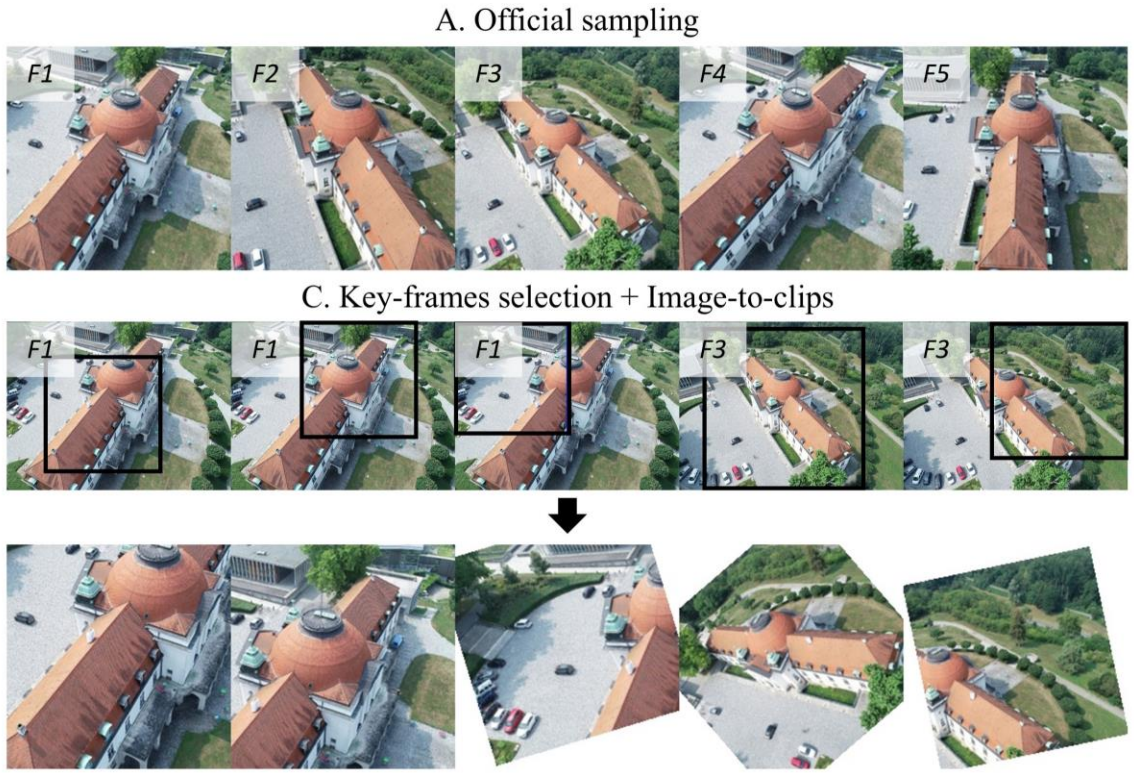


Puzzles: Image-to-Clips. (A) Starting from a single RGB-D image, we (B) partition it into ordered, overlapping patches, (C) simulate diverse viewpoints by calibrating virtual camera poses, and (D) generate augmented, posed images with aligned depth maps for use in 3D reconstruction.

Method – Clips-to-Clips



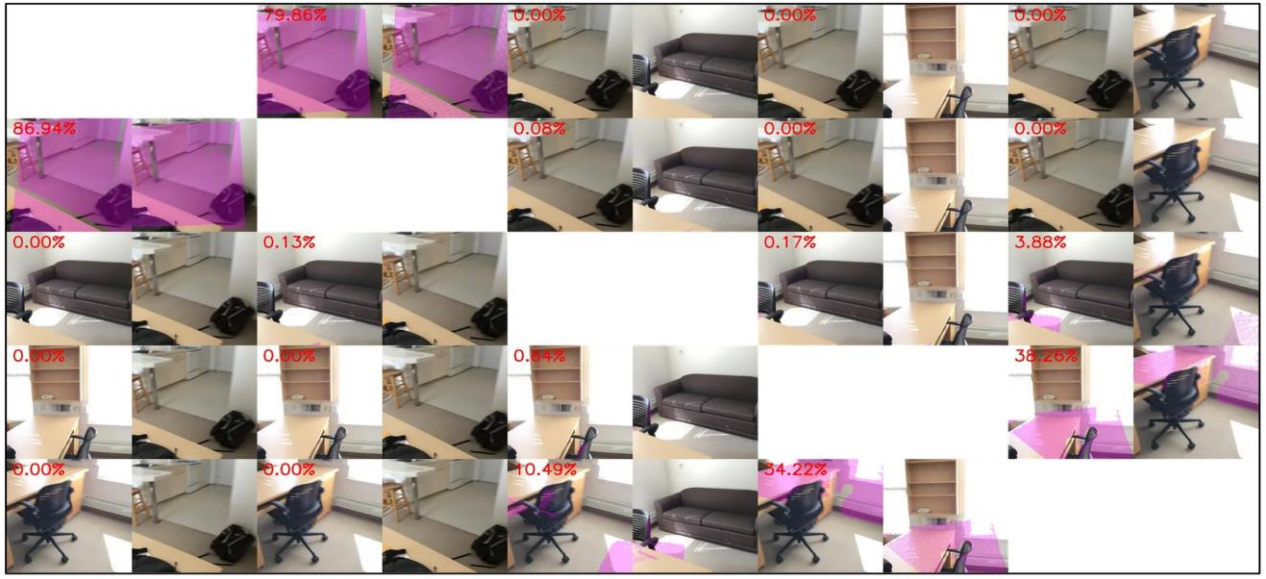
B. Overlapping matrix



Puzzles: Clips- to- Clips. (A) We begin by uniformly sampling frames from a video. (B) A pair-wise overlap matrix is computed to measure frame redundancy, with overlap visualized in purple and overlap ratios annotated in red. (C) Low-redundancy keyframes are then selected, and diverse sub-clips are synthesized from them using the Image-to-Clips method.

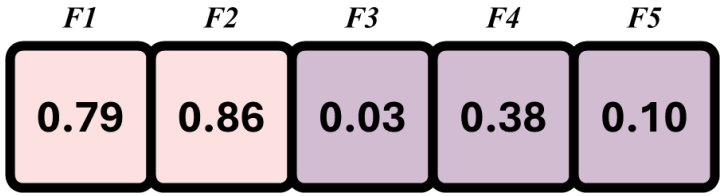
Method – Clips-to-Clips Example

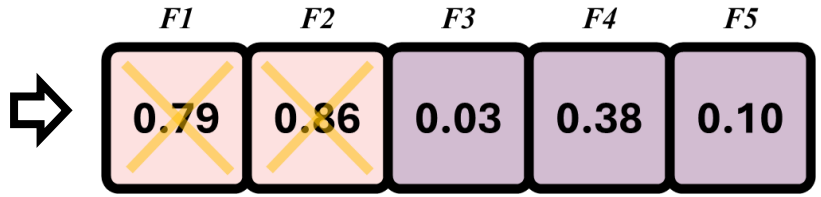
Original sampling



Overlapping matrix

Keyframes selection



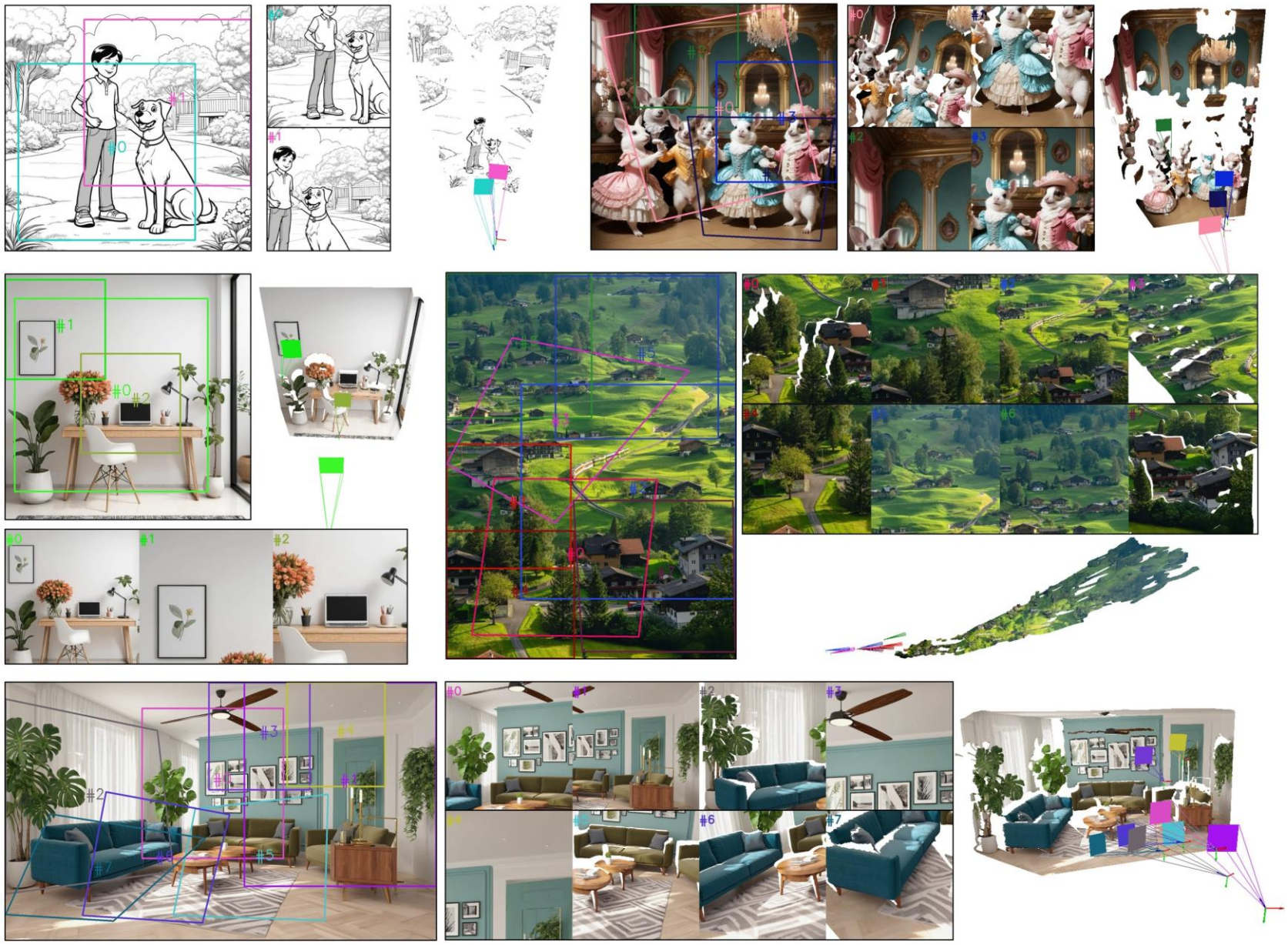


A. Valid frames selection *# Keep frames whose max overlap with any other frame ≥ threshold η → candidates: F1–F5.*

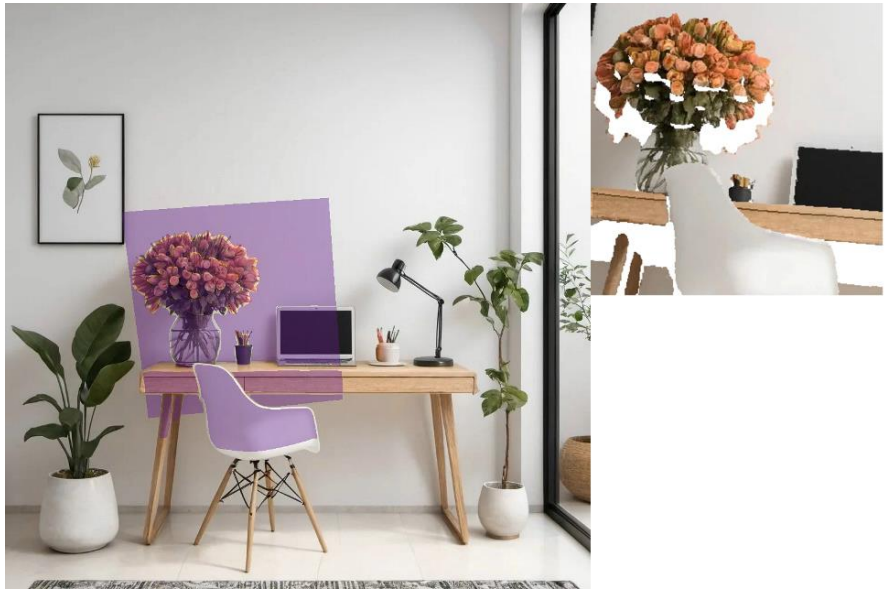
B. Retain the longest set *# From valid frames, choose the largest mutually-overlapping group → pick F3–F5.*

C. Prune redundancy *# Remove frames highly redundant (overlap ≥ ρ =0.01) → final keys: F3–F5.*

Examples – Image-to-Clips



Examples – Image-to-Clips



Examples – Clips-to-Clips

Results

Method	w/ Puzzles	Data	7Scenes				NRGBD				DTU			
			Acc ↓		Comp ↓		Acc ↓		Comp ↓		Acc ↓		Comp ↓	
			Value	Δ(%)	Value	Δ(%)	Value	Δ(%)	Value	Δ(%)	Value	Δ(%)	Value	Δ(%)
Spann3R [9]		full	0.0388		0.0253		0.0686		0.0315		6.2432		3.1259	
	✓	1/10	0.0389	-0.26	0.0248	+1.98	0.0753	-9.79	0.0341	-8.50	4.9832	+20.18	2.5172	+19.47
	✓	full	0.0330	+14.94	0.0224	+11.46	0.0644	+6.00	0.0291	+7.51	5.0004	+19.90	2.5113	+19.66
Fast3R [11]		full	0.0412		0.0275		0.0735		0.0287		4.2961		2.0681	
	✓	1/10	0.0402	+2.30	0.0272	+1.09	0.0772	-5.11	0.0295	-2.78	3.7174	+13.47	1.8941	+8.41
	✓	full	0.0342	+16.99	0.0239	+13.09	0.0684	+6.94	0.0259	+9.75	3.5912	+16.41	1.7379	+15.96
SLAM3R [10]		full	0.0291		0.0245		0.0481		0.0292		4.3820		2.4754	
	✓	1/10	0.0289	+0.68	0.0237	+3.26	0.0493	-2.49	0.0313	-7.19	3.5980	+17.89	2.0891	+15.60
	✓	full	0.0264	+9.27	0.0218	+11.02	0.0439	+8.73	0.0263	+9.93	3.6497	+16.71	2.0762	+16.12

Quantitative comparison on 7Scenes, NRGBD and DTU. Value & relative improvement (Δ) after using **Puzzles**.

Visual Comparison

Puzzles: Unbounded Video-Depth Augmentation for Scalable End-to-End 3D Reconstruction

Project page: https://jiahao-ma.github.io/puzzles/