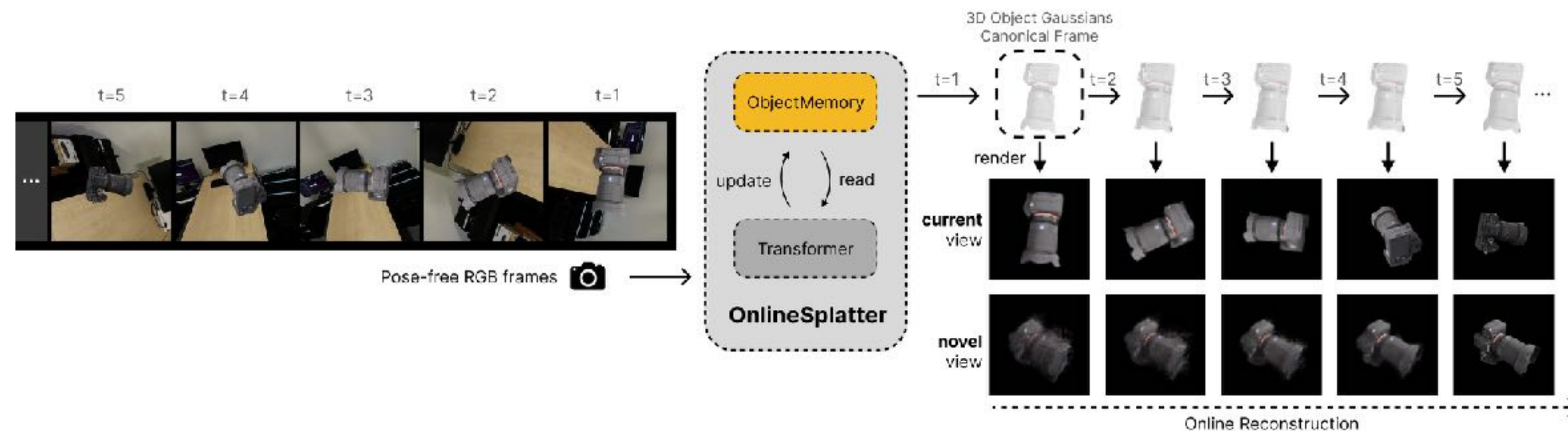


# OnlineSplatter: Pose-Free Online 3D Reconstruction for Free-Moving Objects

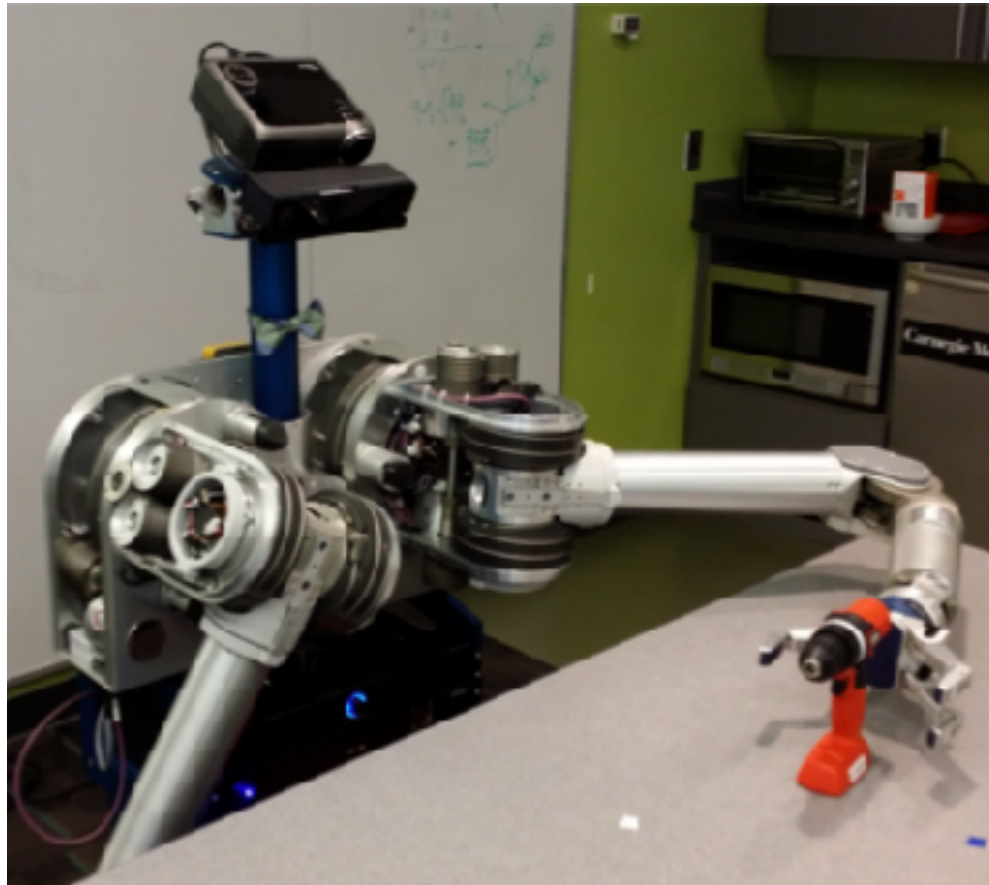
Mark He Huang   Lin Geng Foo   Christian Theobalt   Ying Sun   De Wen Soh

NeurIPS 2025 (Spotlight)



# Towards Unconstrained Assumption-Free Object Perception

Controlled Laboratory



**Static** Scene

**Static** Camera, **Known** Pose



Constrained Real-world

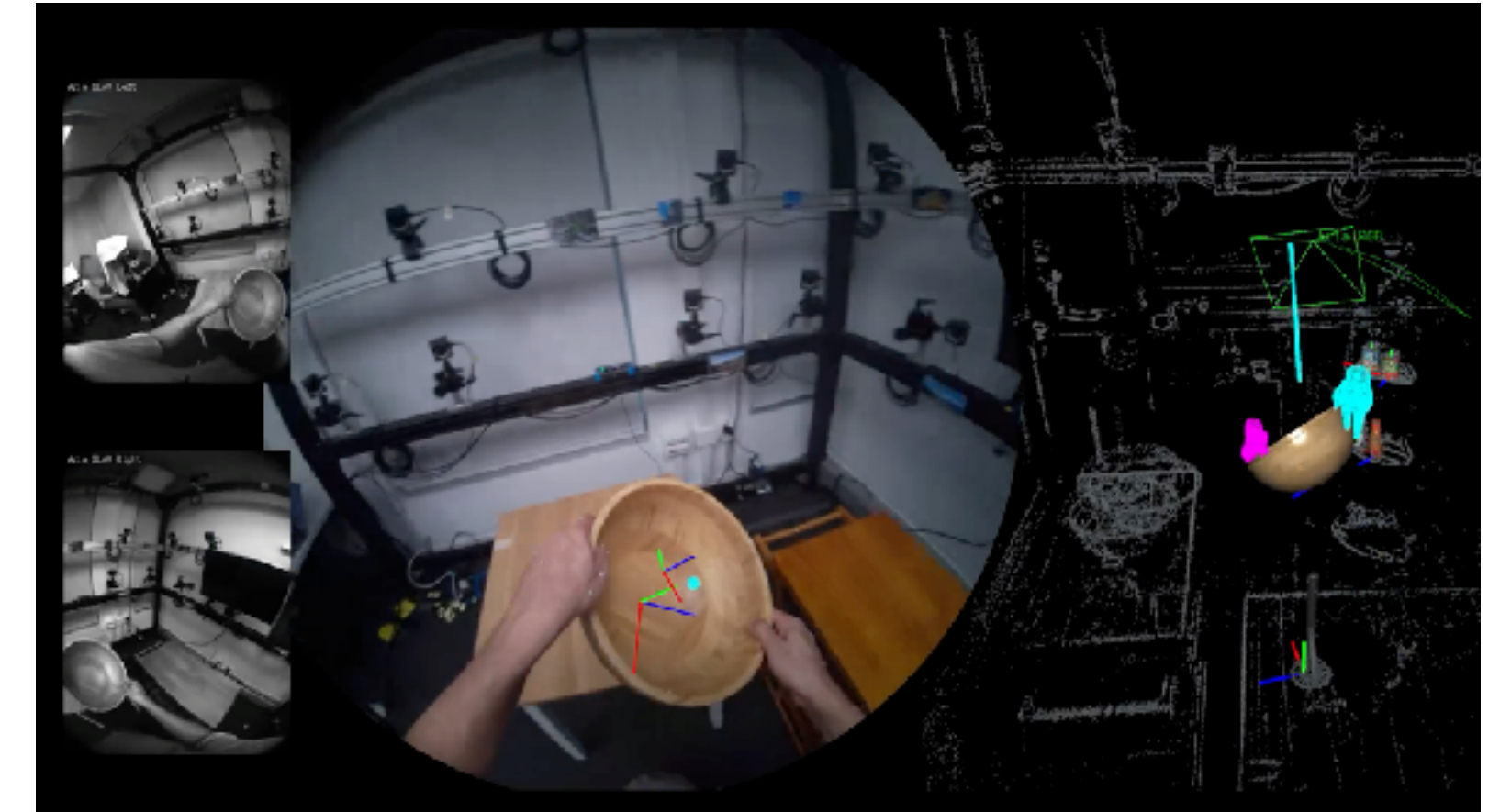


**Static** Scene

**Moving** Camera, **Known** SfM Pose



Unconstrained Real-world



**Dynamic** Scene

**Moving** Camera, **Unknown** Pose

Ego-centric Captures

Autonomous Robots Perception

*The YCB Object and Model Set: Towards Common Benchmarks for Manipulation Research (ICAR 2015)*

*OnePose: One-Shot Object Pose Estimation without CAD Models (CVPR 2022)*

*HOT3D: Hand and Object Tracking in 3D from Egocentric Multi-View Videos*



# Is Large Generative Model the Solution to Object Perception?



Single Image



Model

Feed-forward Model  
e.g. TRELIS, Hunyuan3D



3D Geometry & Texture

Enabled by large-scale object-level priors & use **Hallucination as a feature**

# Is Large Generative Model the Solution to Object Perception?

Enabled by Large-scale trained priors & use **Hallucination as a feature**

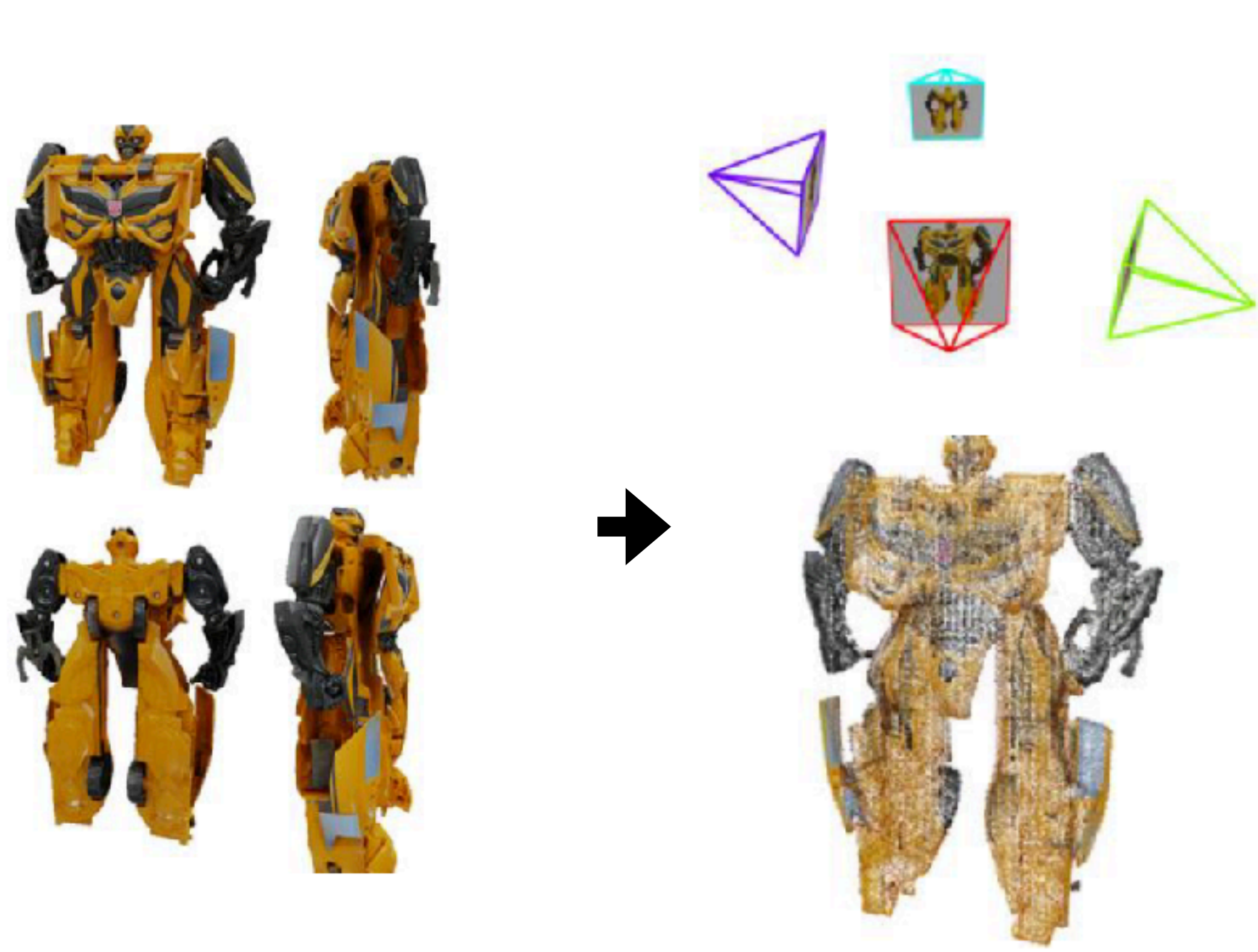
- Great for 3D asset creation, but not for real-time perception
- Performs poorly for out-of-domain objects
- Not applicable to real-time agents in dynamic environments

**How might we enable object perception for autonomous agents?**

**How might we perceive physical objects naturally and continuously?**

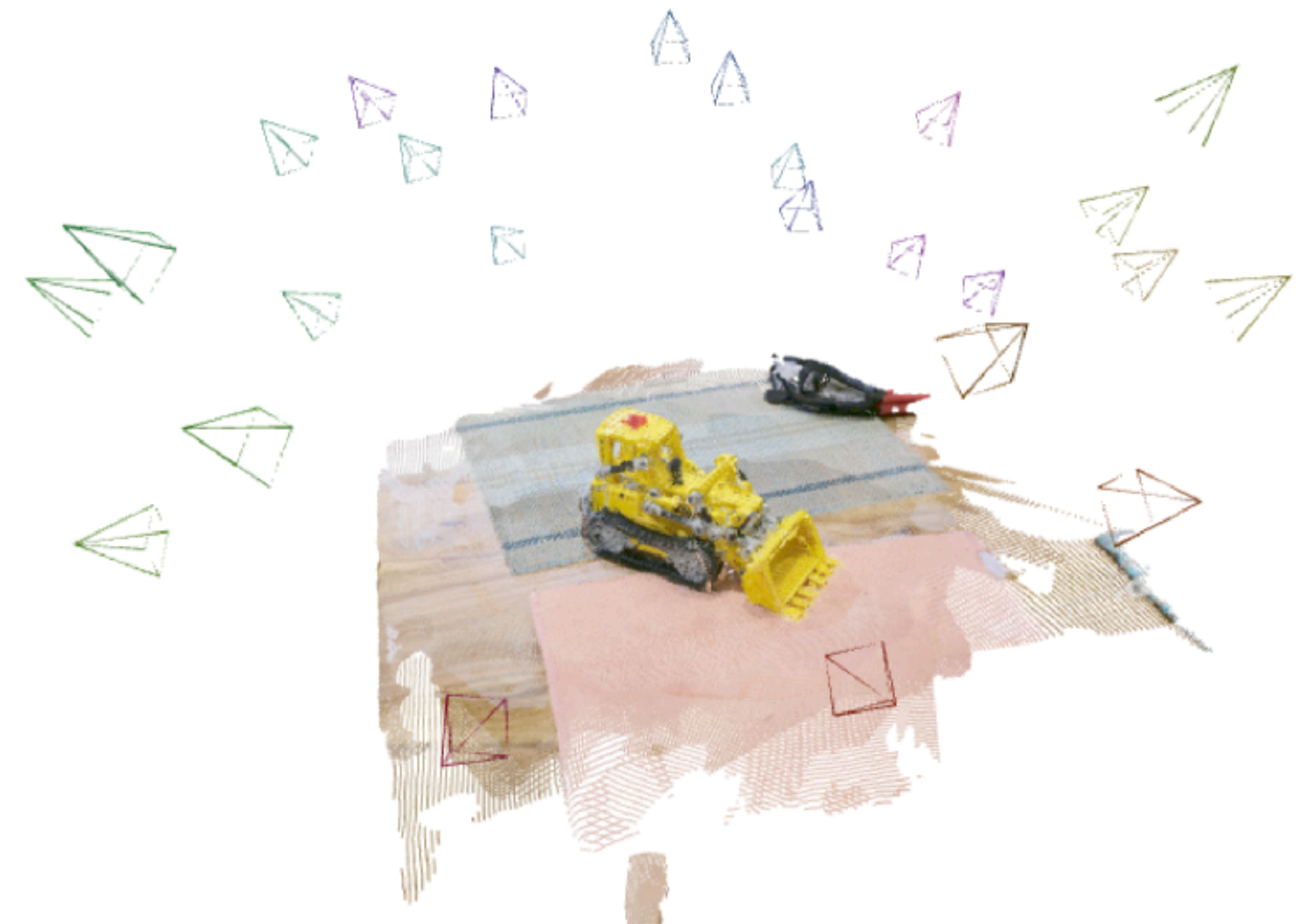


# Recent Advances in Feed-forward RGB-only Reconstruction



FreeSplatter

Pixel-align Gaussian Prediction



VGGT

Pixel-align Points Prediction

Enabled by large-scale learned capability to match and reason in pixel & 3D space

# Our Work

## **Pose-Free RGB Online 3D Reconstruction for Free-Moving Objects**

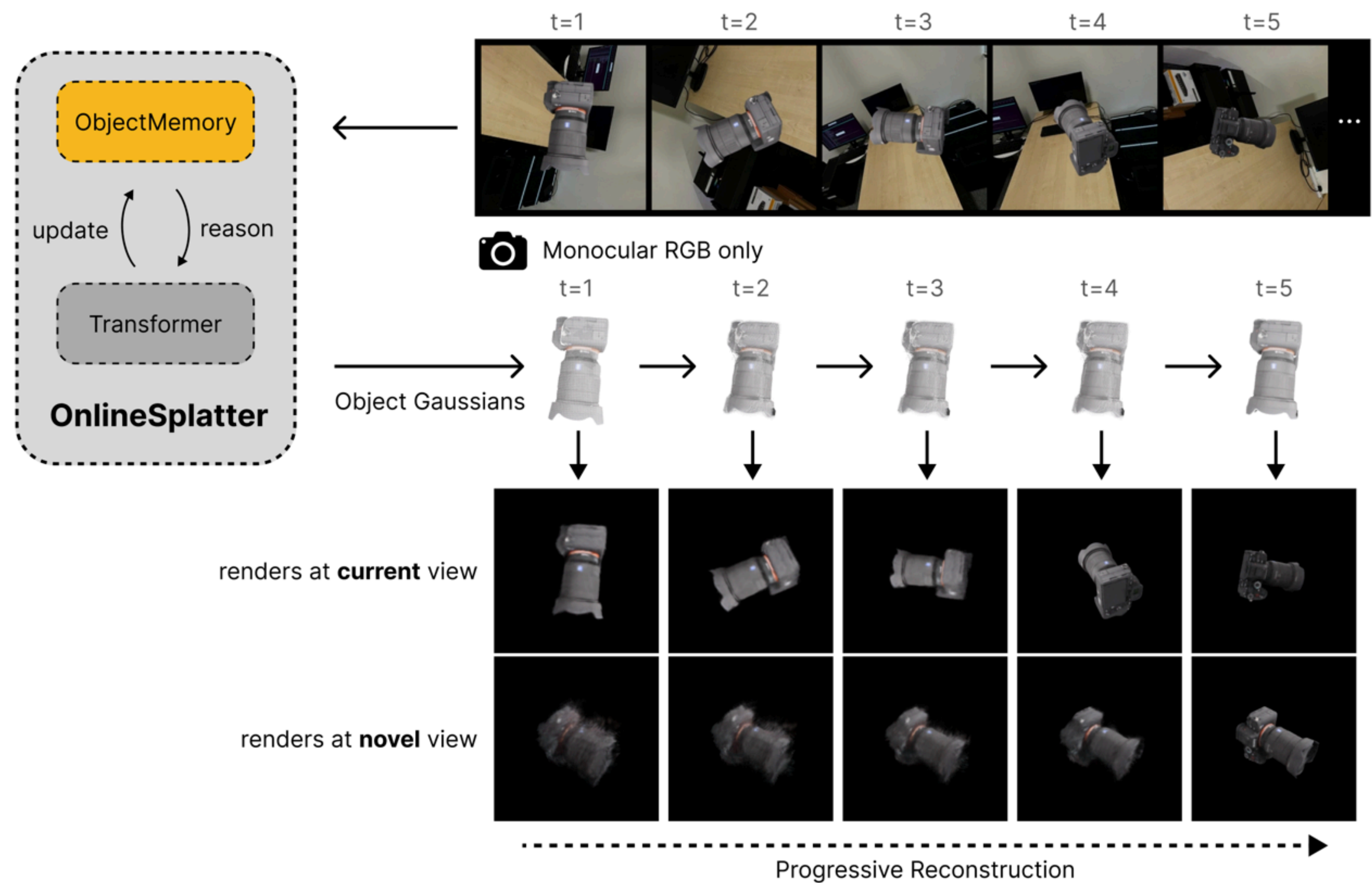
|                             |  |
|-----------------------------|--|
| <b>No Camera Pose</b>       | Removing the requirement for using camera poses as input     |
| <b>No Depth Sensor</b>      | Towards using only monocular RGB frames as inputs            |
| <b>No Pre-capturing</b>     | Towards using video stream to reconstruct objects on-the-fly |
| <b>No Static Assumption</b> | Alleviate the constraints of object placement & movements    |

Eliminating constraints & assumptions ➡ Unlock data in-the-wild

Bring offline reconstruction to online ➡ Unlock real-time experience

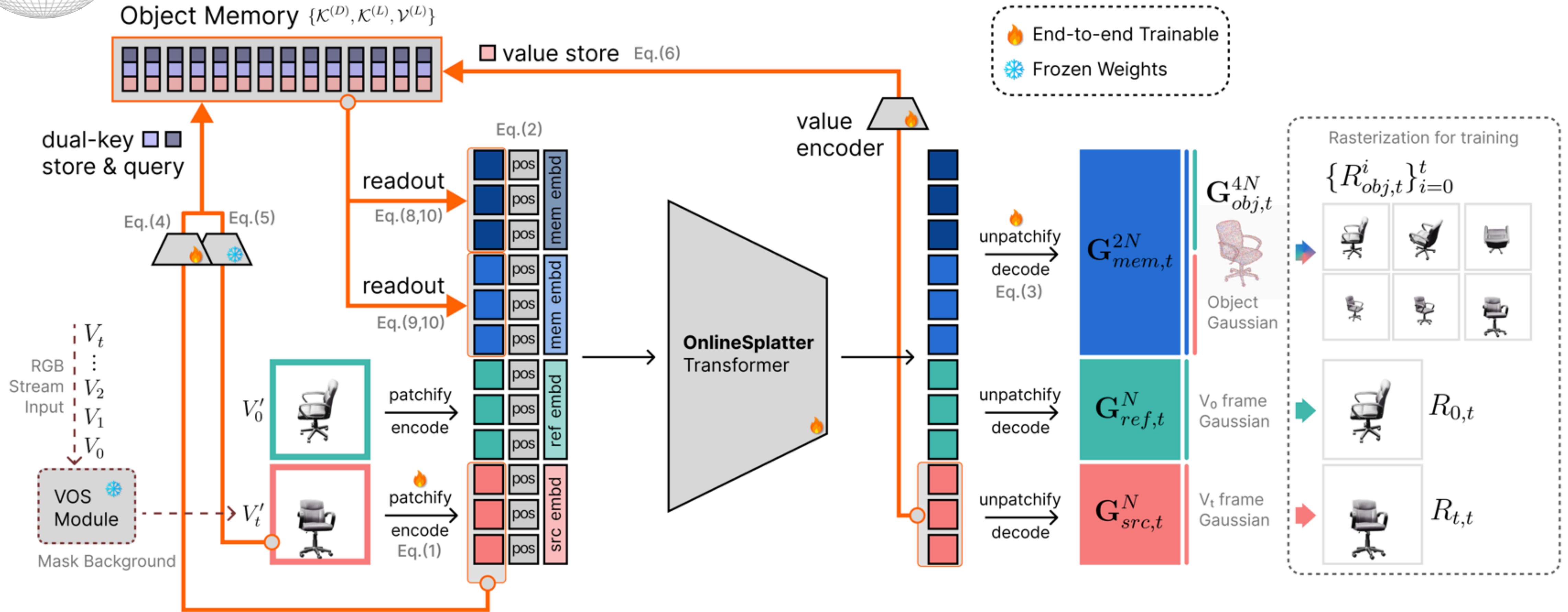
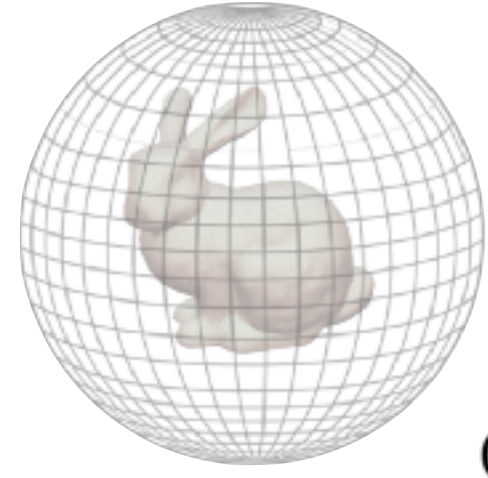


# OnlineSplatter: Pose-Free Online 3D Reconstruction for Free-Moving Objects



Overall Pipeline of OnlineSplatter

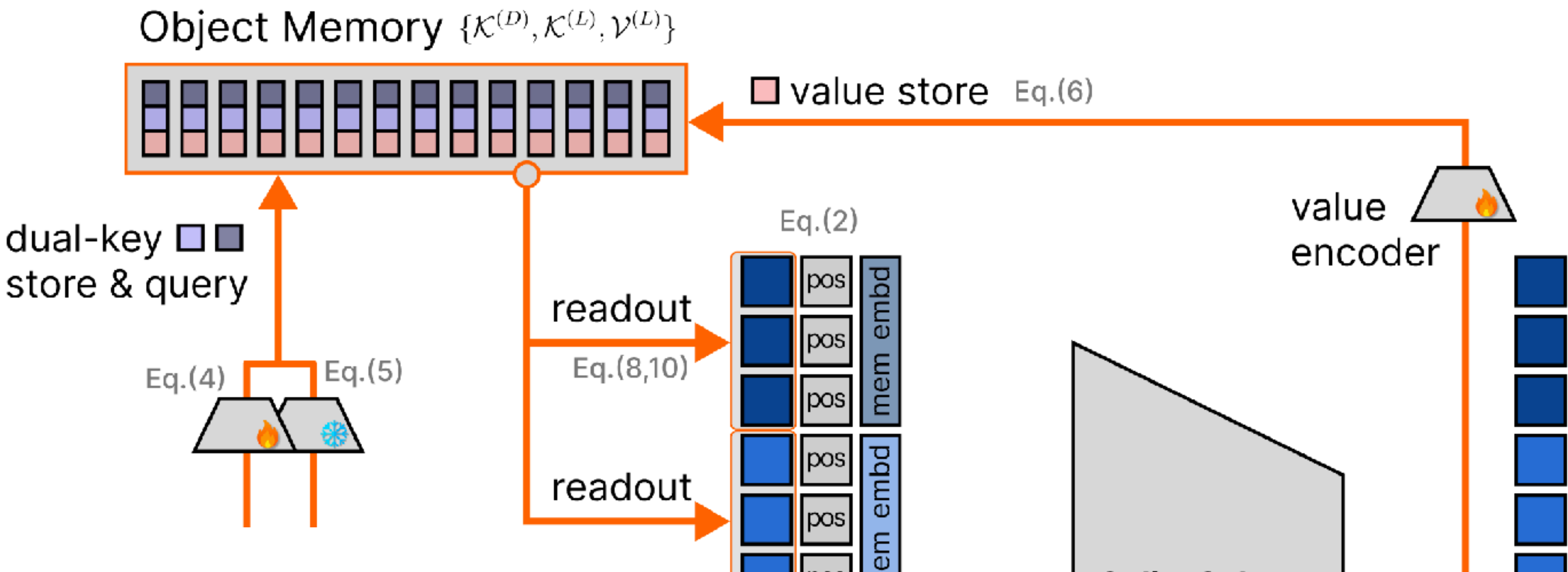
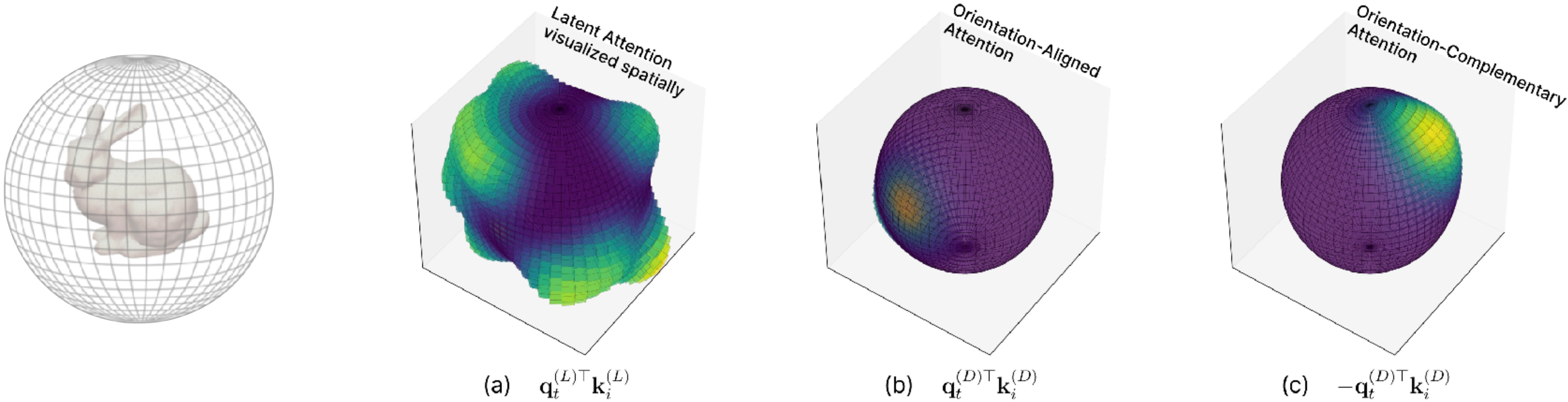
# OnlineSplatter: Pose-Free Online 3D Reconstruction for Free-Moving Objects



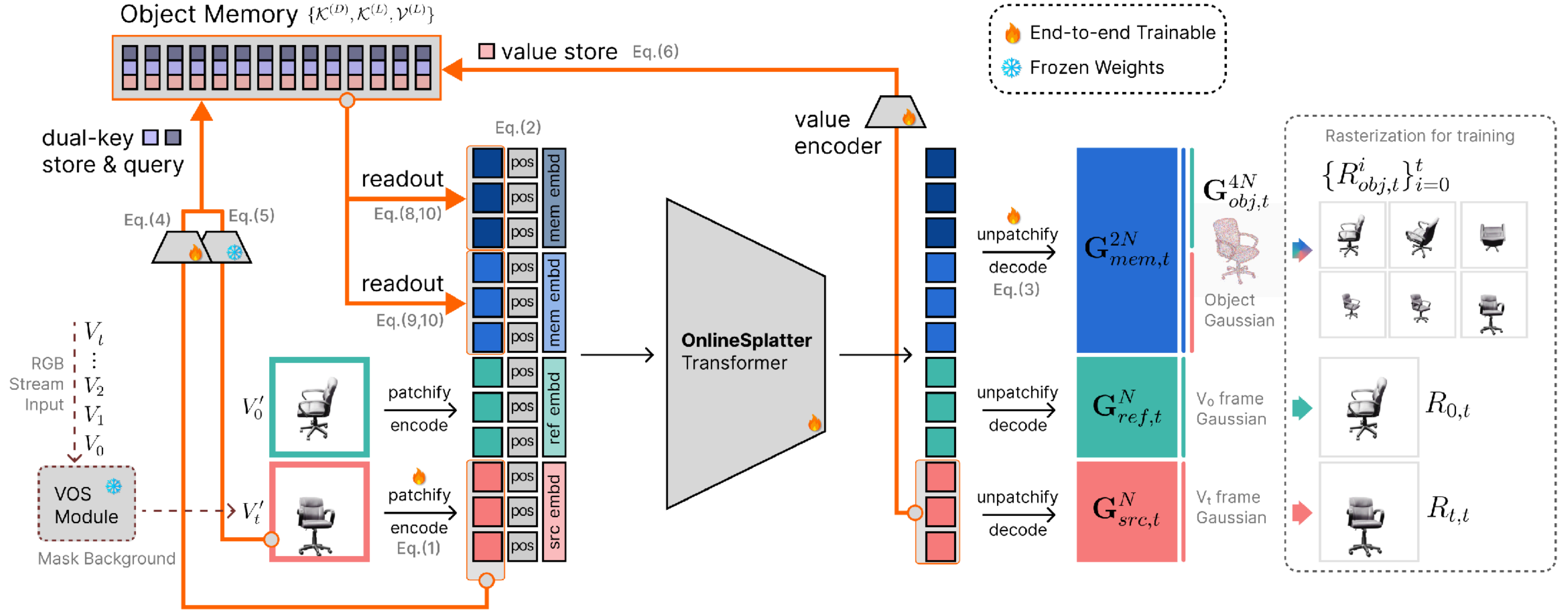
Overall Pipeline of OnlineSplatter



# Dual-Key Spatial Object Memory Insertion and Readout






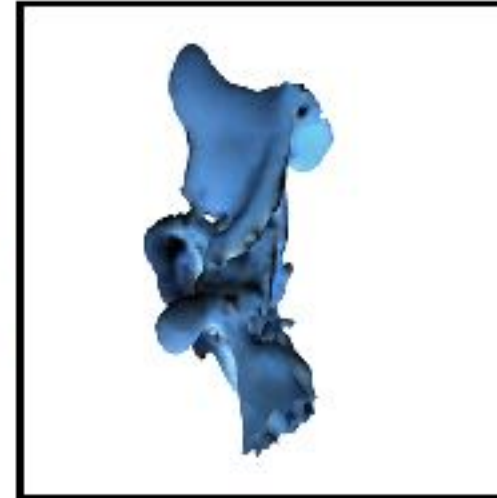
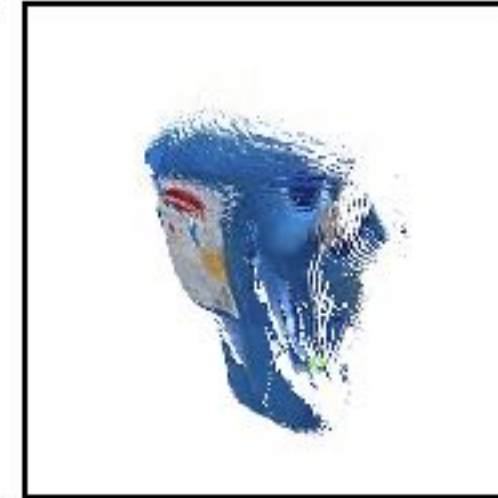


# Dual-Key Spatial Object Memory Insertion and Readout





# It is a Pretty Difficult Task

Many classical and latest methods fails

|   |   |  |   |   |   |   |
|---|---|--|---|---|---|---|
| Virtual-Cam<br>Optimization +<br>Global Refinement                                | Depth-based<br>Online<br>Optimization   | 3DGS-based<br>Offline Bundle<br>Adjustment   | Point-based<br>SfM + MVS<br>Offline BA  | Multi-view<br>Pointmap-based<br>Feed-Forward  | Multi-view<br>Learned LRM   | Multi-view<br>Pointmap-based<br>Feed-Forward  |
|  |  |  |  |  |  |  |
| Fmov  | BundleSDF   | 3DGS*  | COLMAP*   | Fast3R  | InstantMesh   | VGGT*   |

*Free-Moving Object Reconstruction and Pose Estimation with Virtual Camera (AAAI 2025)*

*BundleSDF: Neural 6-DoF Tracking and 3D Reconstruction of Unknown Objects (CVPR 2023)*

*3D Gaussian Splatting for Real-Time Radiance Field Rendering (SIGGRAPH 2023)*

*Structure-from-Motion Revisited (CVPR 2016)*

*Fast3R: Towards 3D Reconstruction of 1000+ Images in One Forward Pass (CVPR 2025)*

*InstantMesh: Efficient 3D Mesh Generation from a Single Image with Sparse-view Large Reconstruction Models (2024)*

*VGGT: Visual Geometry Grounded Transformer (CVPR 2025)*



# Quantitative and Qualitative Results

| GSO                  |               |              |              |               |              |              |               |              |              |
|----------------------|---------------|--------------|--------------|---------------|--------------|--------------|---------------|--------------|--------------|
| Method               | Early-Stage   |              |              | Mid-Stage     |              |              | Late-Stage    |              |              |
|                      | PSNR↑         | SSIM↑        | LPIPS↓       | PSNR↑         | SSIM↑        | LPIPS↓       | PSNR↑         | SSIM↑        | LPIPS↓       |
| FSO <sub>rand4</sub> | 21.358        | 0.861        | 0.177        | 21.919        | <u>0.877</u> | 0.181        | 21.737        | 0.855        | 0.181        |
| FSO <sub>dist4</sub> | 22.365        | <u>0.874</u> | <u>0.119</u> | <u>23.757</u> | 0.862        | <u>0.117</u> | 23.751        | 0.873        | <u>0.120</u> |
| NPS <sub>dist2</sub> | 22.986        | 0.859        | 0.155        | 23.050        | 0.863        | 0.162        | 22.949        | <u>0.878</u> | 0.156        |
| NPS <sub>dist3</sub> | <u>23.331</u> | 0.862        | 0.149        | 23.206        | 0.861        | 0.138        | <u>24.141</u> | 0.863        | 0.125        |
| <b>Ours</b>          | <b>26.329</b> | <b>0.921</b> | <b>0.084</b> | <b>27.553</b> | <b>0.933</b> | <b>0.066</b> | <b>31.737</b> | <b>0.969</b> | <b>0.075</b> |

| HO3D                 |               |              |              |               |              |              |               |              |              |
|----------------------|---------------|--------------|--------------|---------------|--------------|--------------|---------------|--------------|--------------|
| Method               | Early-Stage   |              |              | Mid-Stage     |              |              | Late-Stage    |              |              |
|                      | PSNR↑         | SSIM↑        | LPIPS↓       | PSNR↑         | SSIM↑        | LPIPS↓       | PSNR↑         | SSIM↑        | LPIPS↓       |
| FSO <sub>rand4</sub> | 18.488        | 0.820        | 0.187        | 18.552        | 0.817        | 0.191        | 17.683        | 0.810        | 0.199        |
| FSO <sub>dist4</sub> | 18.594        | 0.837        | 0.177        | 19.215        | 0.848        | 0.184        | 19.619        | 0.843        | 0.183        |
| NPS <sub>dist2</sub> | 21.063        | <u>0.855</u> | <u>0.160</u> | 22.803        | 0.841        | <u>0.158</u> | 22.134        | 0.846        | 0.164        |
| NPS <sub>dist3</sub> | <u>21.134</u> | 0.853        | 0.162        | <u>22.967</u> | <u>0.869</u> | 0.165        | <u>22.947</u> | <u>0.860</u> | <u>0.163</u> |
| <b>Ours</b>          | <b>23.627</b> | <b>0.910</b> | <b>0.152</b> | <b>25.803</b> | <b>0.912</b> | <b>0.122</b> | <b>27.928</b> | <b>0.952</b> | <b>0.099</b> |

Table 1: Comparison of different baselines on two datasets. Results are shown for early-stage mid-stage, and late-stage settings. Best results are **bolded** and second best results are underlined.

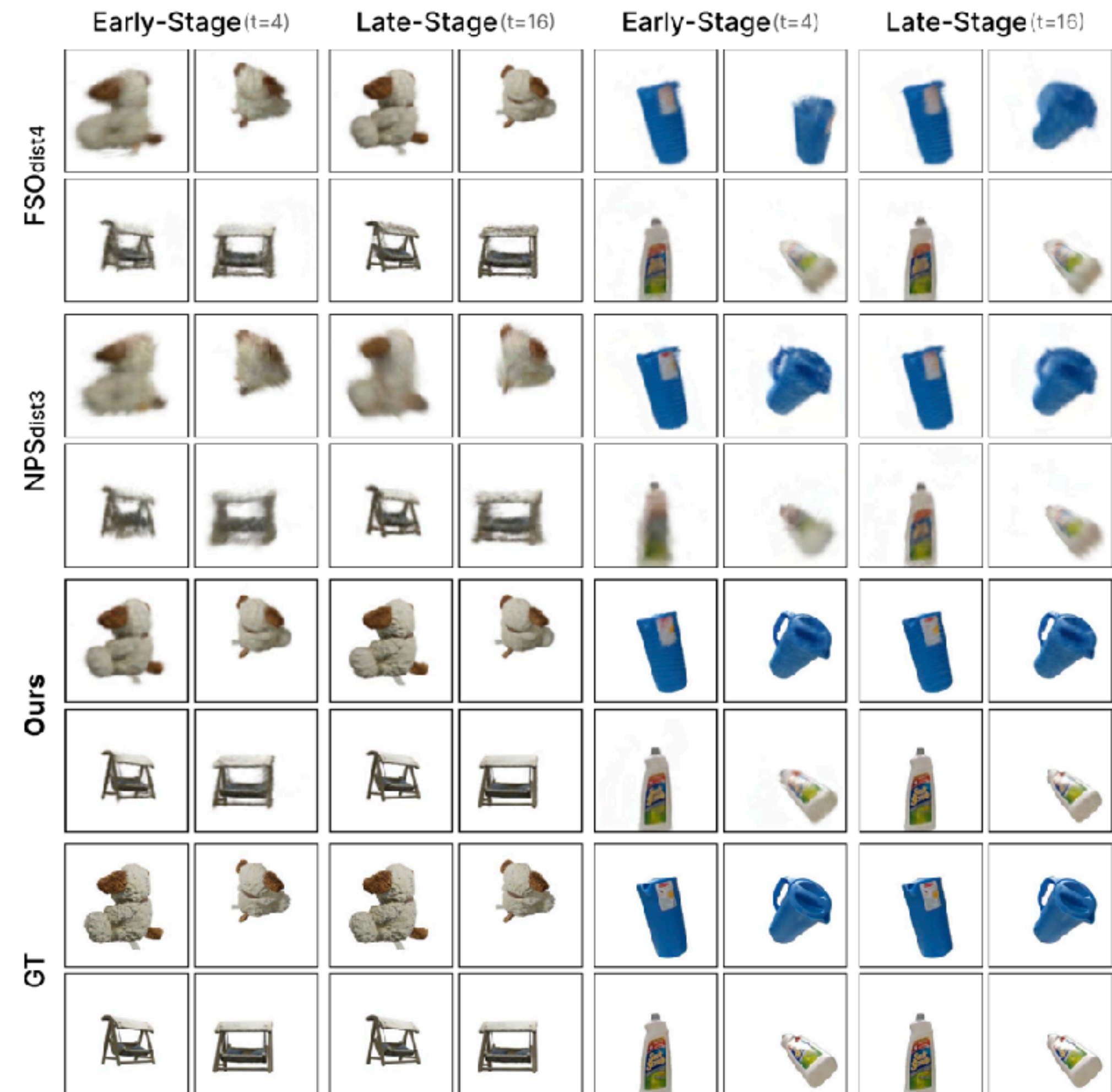
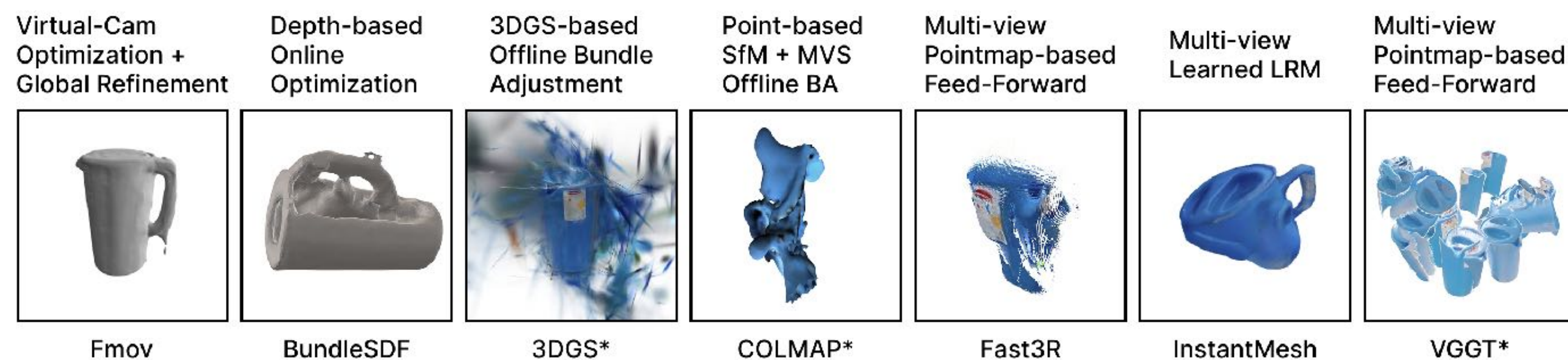


Figure 3: Qualitative results of different baselines and our method on the GSO (left) and HO3D (right) datasets. We visualize the results at inference timestep  $t = 4$  and  $t = 16$ , which corresponds to the early-stage and late-stage settings, respectively. Our reconstructed outputs shows significant better visual quality and geometric accuracy as more observations become available.



## Future Works

- Feedforward Model + Real-time Lightweight Optimization
- Pose-free Training
- Non-rigid Object On-the-fly Reconstruction
- Hybrid Object Representation

# Concurrent Works Exploring Stateful Continuous Reconstruction

- CUT3R: Continuous 3D Perception Model with Persistent State
- TTT3R: 3D Reconstruction as Test-Time Training



*[markhh.com/OnlineSplatter](http://markhh.com/OnlineSplatter)*