# Scal3R: Scalable Test-Time Training for Large-Scale 3D Reconstruction

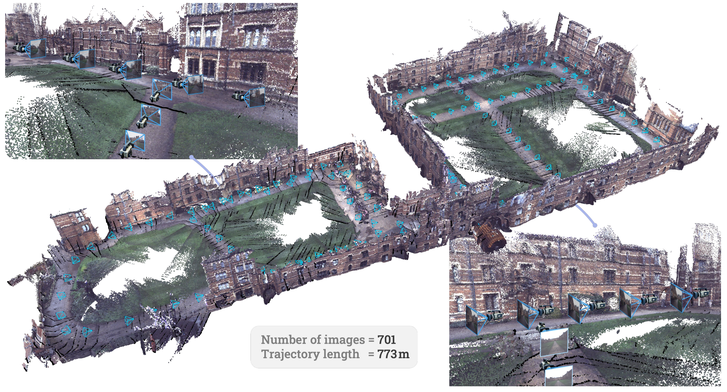
Tao Xie[1,2], Peishan Yang[1], Yudong Jin[1], Yingfeng Cai[2], Wei Yin[2], Weiqiang Ren[2], Qian Zhang[2], Wei Hua[1,3], Sida Peng[1], Xiaoyang Guo[2 †], Xiaowei Zhou[1†]
[1] Zhejiang University, [2] Horizon Robotics, [3] Zhejiang Lab

CVPR JUNE 3-7, 2026 DENVER COLORADO

https://zju3dv.github.io/scal3r

## Introduction

**Task:** large-scale 3D reconstruction from long RGB video.



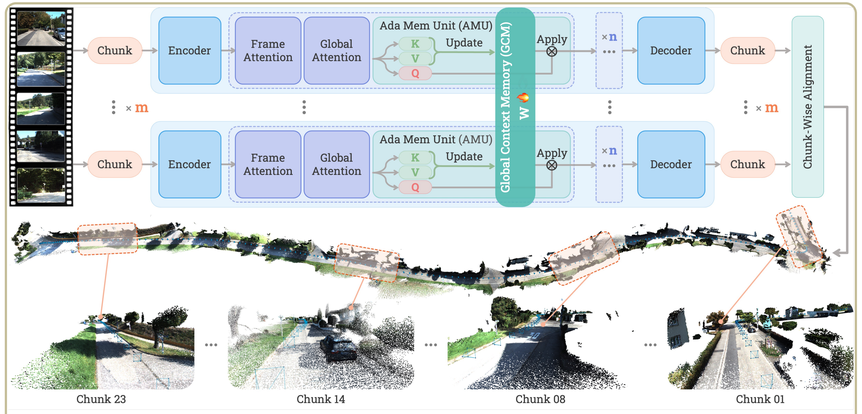

**Challenge 1:** quadratic computation cost of attention.
**Challenge 2:** behavior misalignment, train short but test long.
**Challenge 3:** chunk-wise methods lose cross-chunk context.

**Key idea:** chunk-wise processing, and put long-sequence training and inference and cross-chunk context updating and sharing into one unified pipeline.

## Contributions

- **A novel framework** capable of reconstructing high-quality kilometer-scale 3D scenes from long RGB-only sequences.
- **A global context representation** together with **a context aggregation mechanism** that jointly compresses, retains, and shares long-term information across sequences.
- Achieves **state-of-the art performance** with superior pose & 3D geometry accuracy and global consistency.

## Method

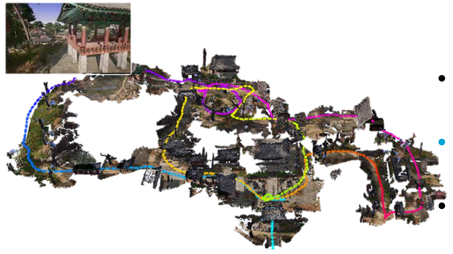



### 1. Global Context Memory (GCM)

**Adaptive Memory Units (AMU)** attached after 4 global-attention layers, which are updated **online** with chunk-level self-supervised key→value losses.

### 2. Global Context Synchronization (GCS)

Multi-GPU chunk partitioning as **context parallelism**. AMU gradients synchronized via PyTorch **all-reduce** every chunk.

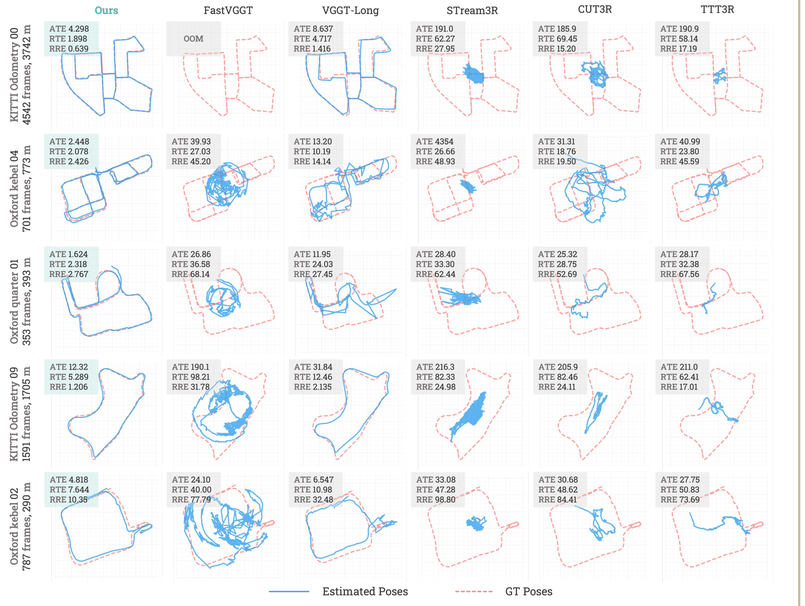


### 3. Why it works

- The **chunk-wise computation** tames long-video attention.
- **Cross-chunk synchronization**, not isolated chunk alignment.
- Same long-sequence flow in **training & inference**.

## Results

### 1. Camera Trajectory Comparison




### 2. Point Cloud Reconstruction Comparison