PLAN LAB
Perception and LANguage Lab

# SILSA

## Sliding-Window Slice Latents for Topology-Preserving High-Resolution 3D Generation



**Tianjiao Yu**


Xinzhuo Li


Yifan Shen


Ying Shen


Kiet A. Nguyen


Adheesh Sunil Juvekar


Ismini Lourentzou

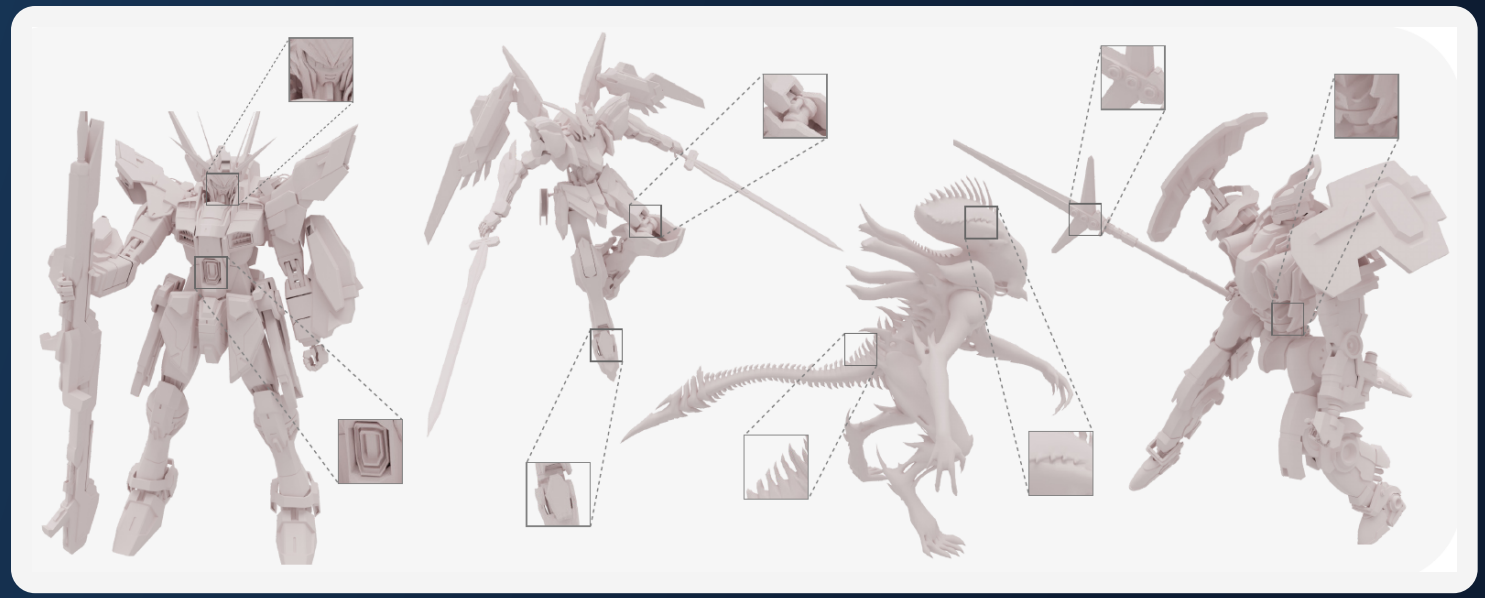
University of Illinois Urbana-Champaign




NeurIPS 2026 / Atlanta Poster Session 6

# FINE 3D STRUCTURE BREAKS

Shapes match the overall look, yet lose thin parts, holes, and connectivity



HARD CASES

- Thin supports
- Holes and openings
- Repeated components
- Long-range connectivity

**Dense voxels**
Memory and compute grow cubically with resolution

**Sparse / hierarchical tokens**
Tens of thousands of data-dependent tokens, multi-stage

**Compact set/triplane latents**
Cheap, but tokens lose their link to local geometry

# OVERLAPPING SLICE WINDOWS

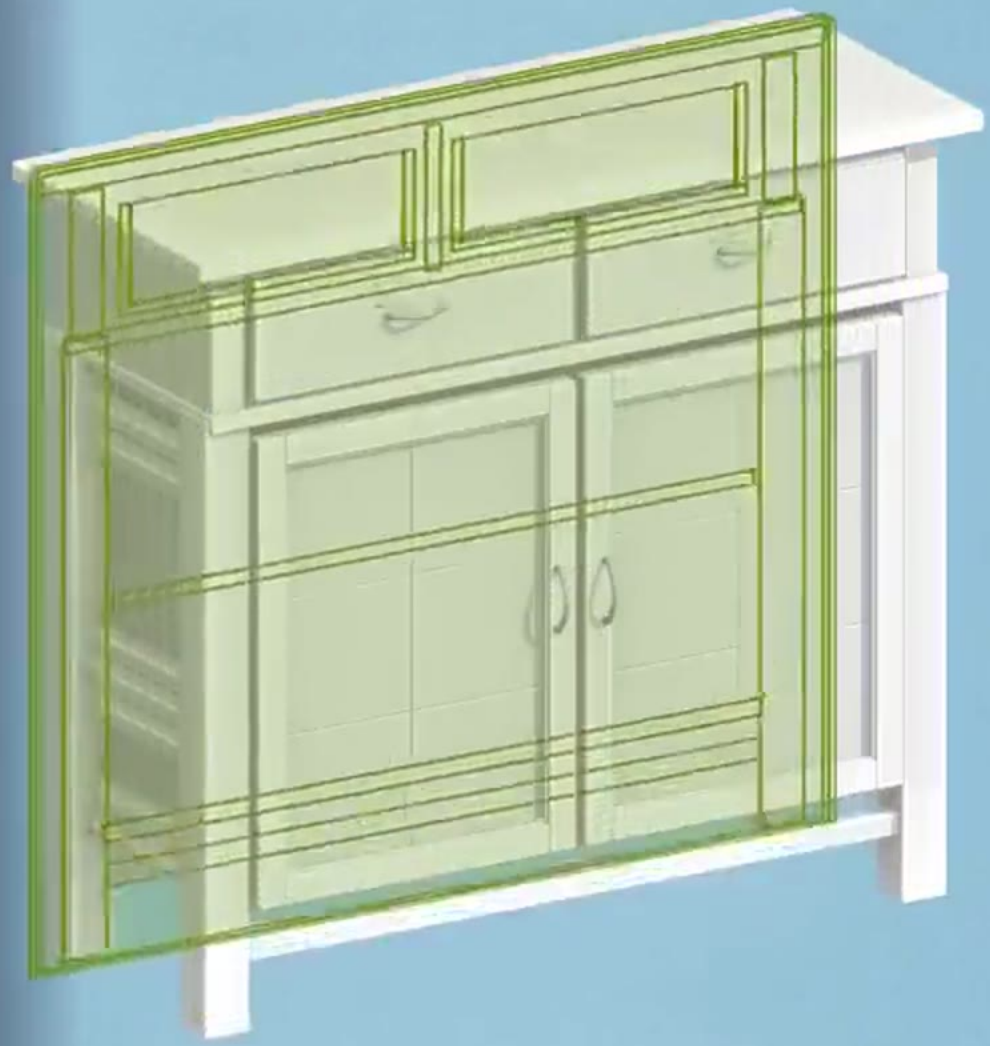
A thin slice captures the topology of a whole planar region in one token

X 128 | Y 128 | Z 128

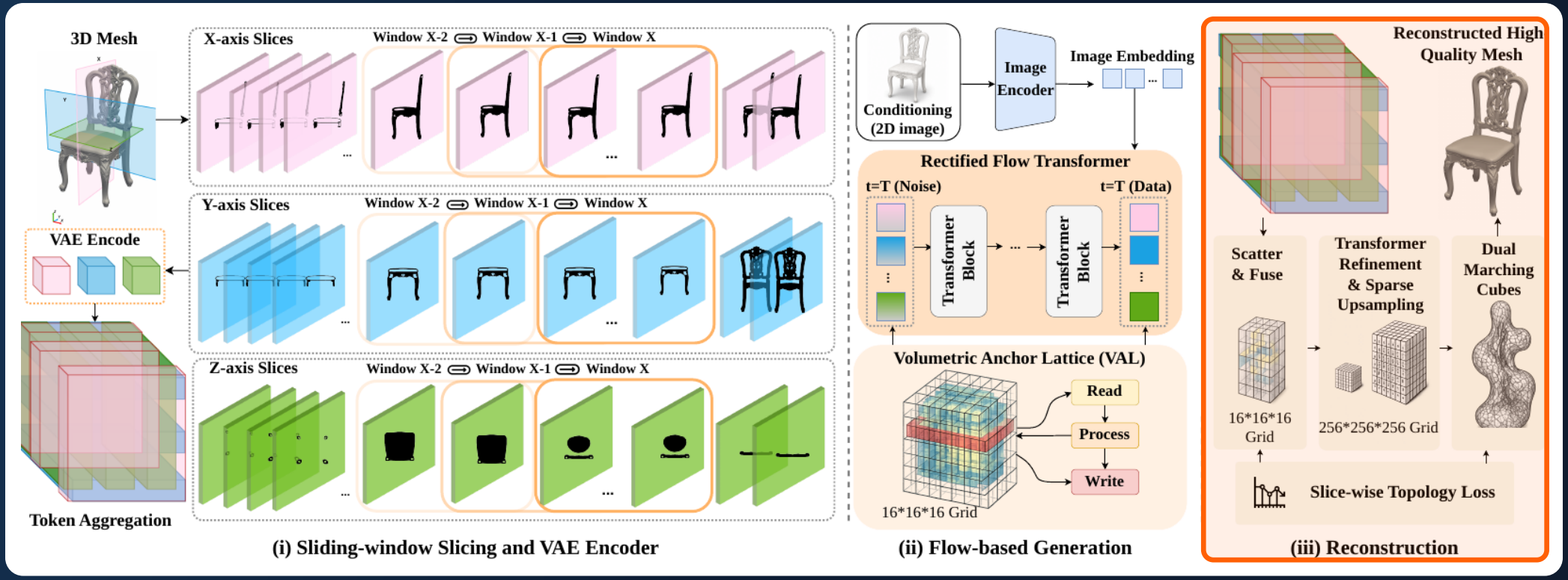
**384** slice tokens per shape

128 positions per axis · 8-bin windows · fixed for every shape





# SILSA PIPELINE




### i. Slice VAE encoder

Sliding windows of 8 bins pool surface points into Gaussian slice latents

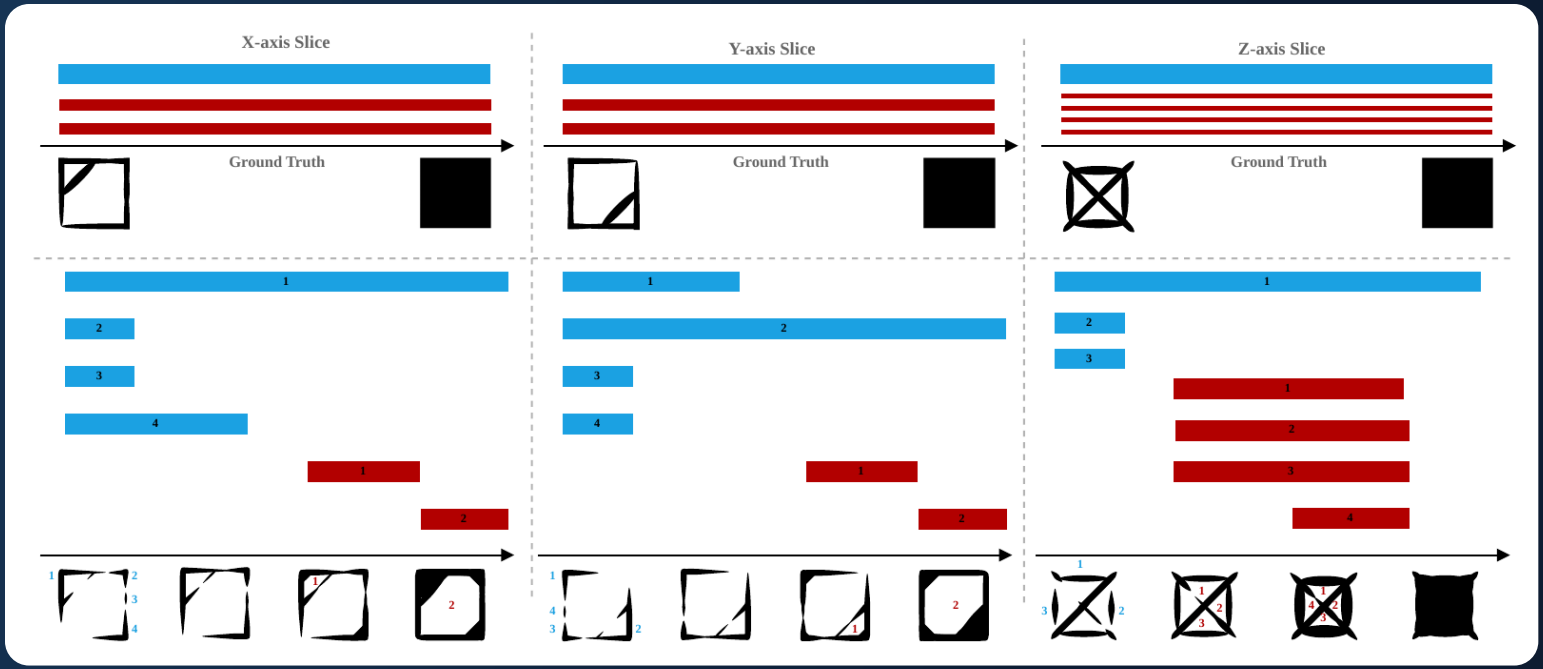
### ii. Single-stage rectified flow

Generates all 384 slice latents from one image; no active-voxel stage

### iii. Sparse volumetric decoder

Scatter to a shared grid, upsample $16^3 \rightarrow 256^3$, extract the mesh

# SLICE-WISE TOPOLOGY SUPERVISION




$\beta_0$ components

$\beta_1$ holes

**Per-slice persistence**

Match persistence diagrams within each cross-section

**Betti transitions**

Components and holes appear, merge, and split at the right depths

Targets errors rendering losses miss: holes closing early, supports disconnecting, parts merging.

# VOLUMETRIC ANCHOR LATTICE

Shared $16^3$ memory inside the flow transformer

X Y Z

**Read**
each token attends to its axis-aligned plane

**Write**
a gated update deposits its evidence

**One shape**
x, y, and z streams decode together



# IMAGE-TO-3D GENERATION

Toys4K and in-the-wild images · 10 baselines · best or tied-best on all 7 metrics

PSNR

32.74

dB vs 26.31 for SAR3D

+24.4%

COVERAGE

79.08%

vs 73.01% for XCube

+6.07 pts

CLIP

87.94

vs 85.03 for TRELLIS

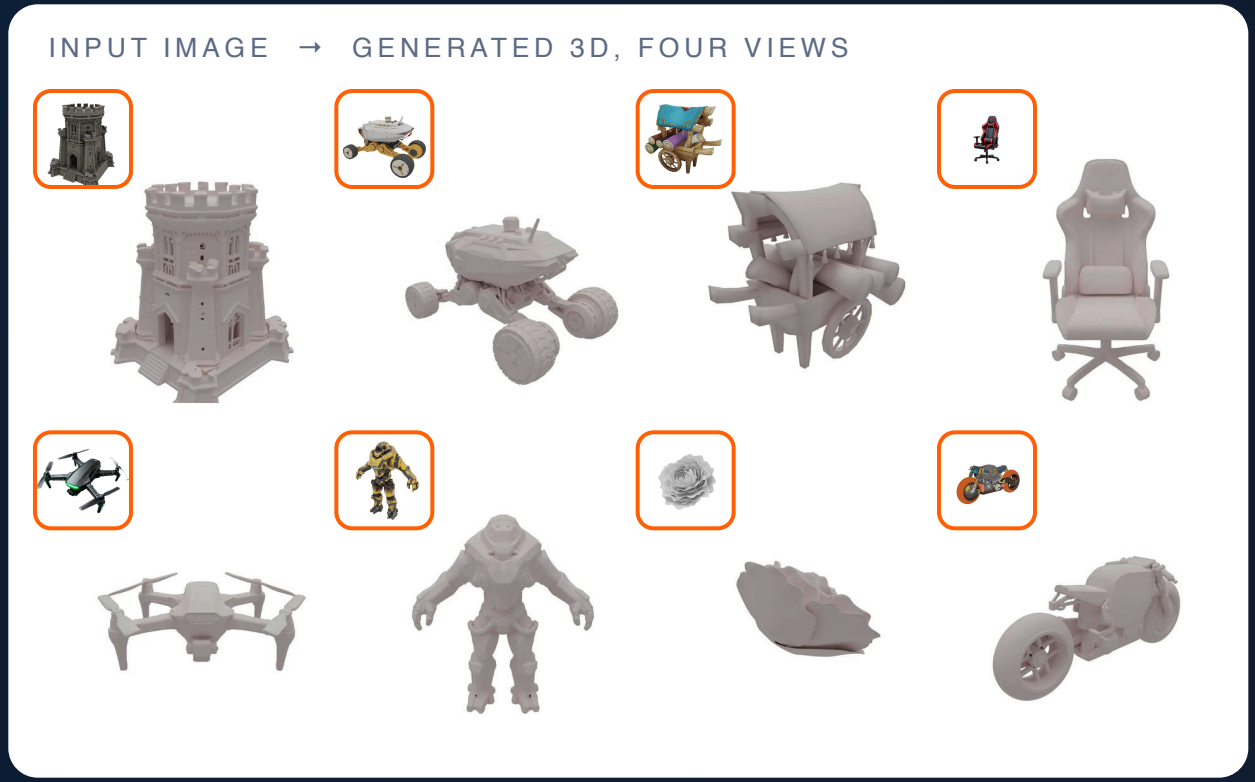
best

LPIPS

0.05

vs 0.10 for SAR3D

halved




In-the-wild inputs: plausible, view-consistent structure from a single image.

# GENERATION EFFICIENCY

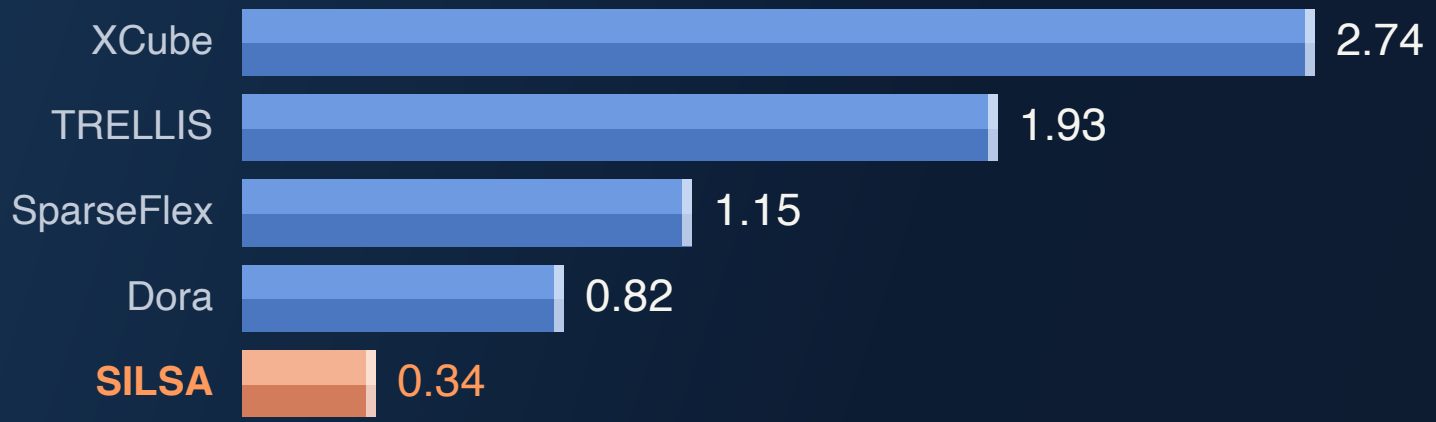
LATENT TOKENS PER SHAPE ↓




**70% fewer**

tokens than the most compact baseline; over 98% fewer than sparse voxel tokenizers

INFERENCE TIME (S / SHAPE) ↓



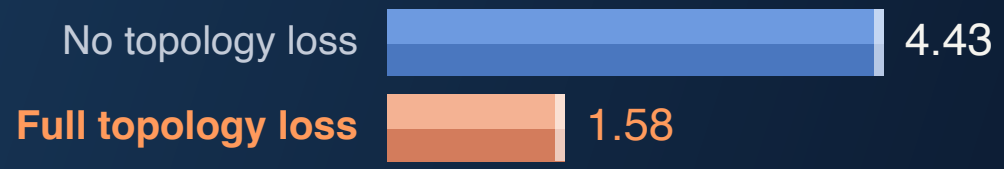

**0.34 s**

per shape, end to end, in a single stage

# ABLATIONS

Betti error ↓ (lower is better)

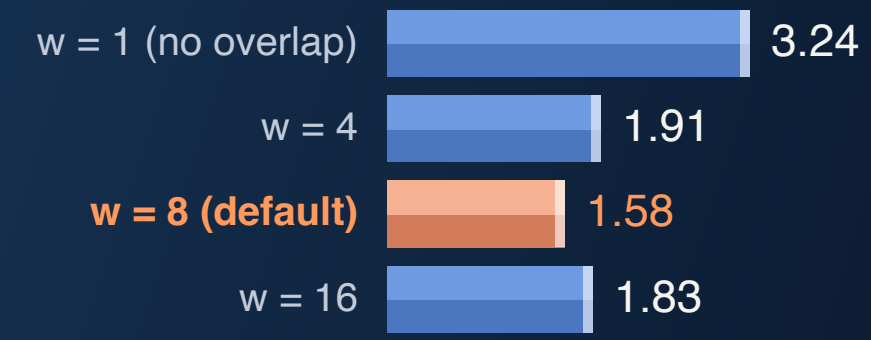
## TOPOLOGY SUPERVISION




## CROSS-AXIS COMMUNICATION




## WINDOW WIDTH




**Takeaway**

Topology losses, a shared 3D workspace, and overlapping windows each matter for structural correctness.

# SUMMARY

REPRESENTATION

## 384 tokens

Overlapping sliding-window slices along x, y, and z, fixed for every shape

STRUCTURE

## –9.2% Betti

Per-slice persistence and Betti-transition supervision

GENERATION

## +24.4% PSNR

Single-stage rectified flow with a Volumetric Anchor Lattice

**plan-lab.github.io/silsa**

# NEURIPS 2026 POSTER

Atlanta · Hall C1

PLAN LAB
Perception and LANguage Lab



THIS PAPER

## SILSA

Sliding-Window Slice Latents for Topology-Preserving High-Resolution 3D Generation

**FRIDAY, DECEMBER 11**
**4:30 PM ET**

Hall C1 · Poster Session 6

PLAN LAB POSTERS AT NEURIPS 2026

**WED, DEC 9 · 10:00 AM ET**
Cognitive Supersensing
Toward Cognitive Supersensing in Multimodal Large Language Models
**Poster Session 1**

**WED, DEC 9 · 10:00 AM ET**
Tiny but Trusted
Efficient Vision-Language Reasoning for Time-Series Anomaly Detection
**Poster Session 1**

**FRI, DEC 11 · 10:00 AM ET**
ELSA3D
Elastic Semantic Anchoring for Unified 3D Understanding and Generation
**Poster Session 5**

**FRI, DEC 11 · 4:30 PM ET**
SILSA
Sliding-Window Slice Latents for Topology-Preserving High-Resolution 3D Generation
**Poster Session 6**

PLAN LAB
Perception and LANguage Lab

# Thank you!




**Tianjiao Yu**

ty41@illinois.edu


plan-lab.github.io/silsa

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

SILSA · NeurIPS 2026

