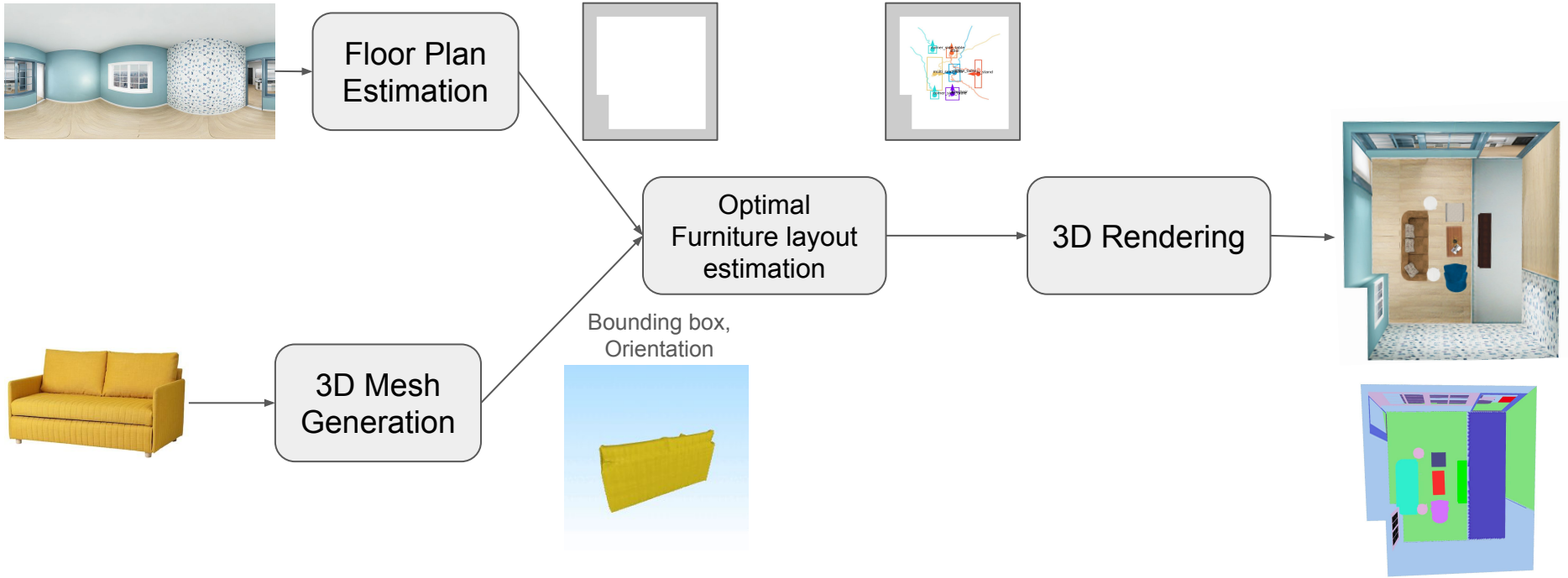
PanoFusion

Data Generation pipeline novel scene synthesis for contextually aware Furniture Placement

Taru Rustagi
Tanya Choudhary
Santhoshini Gongidi
Aman Goel

- **Increased training Data for indoor scene understanding tasks**
  - With annotations like semantic object labels, depth maps
- **Multi-view Rendering**: Given the 3D scenes, we can generate panoramas from various viewpoints

Our Pipeline



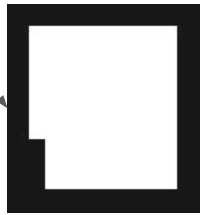
Floor Plan Estimation

- Estimate the room layout by identifying the junctions between the ceiling, walls, and floor.
- Render a 3D `.obj` model and apply texture mappings from the panorama image
- Utilize similar texture mappings to render new indoor scenes (Permutation)

Panorama Equirectangular image

Floor Plan Estimation

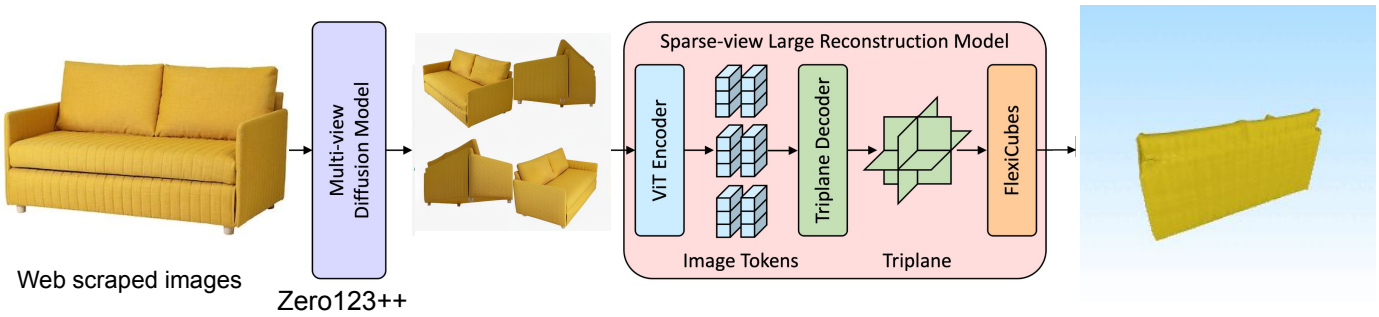
Estimated corners



Floor Plan Layout

3D Mesh Generation

- Use existing datasets such as 3D-FUTURES to get furniture meshes OR
- Create furniture meshes using **InstantMesh** from web-scraped real images
- Point cloud registration to adjust scale and orientation of meshes



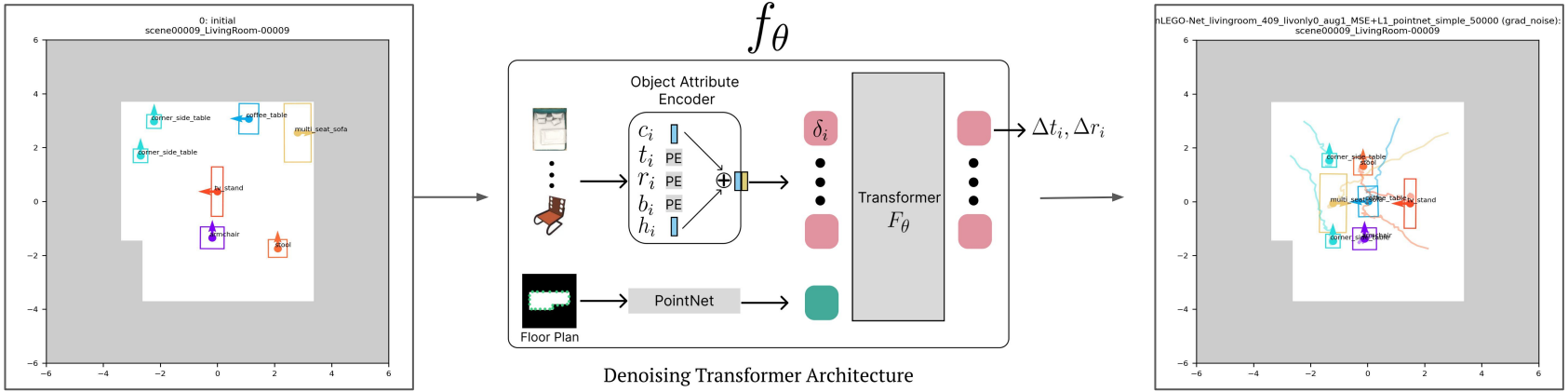
3D Furniture Mesh Generation

Point cloud registration

Jiale Xu et al. “InstantMesh: Efficient 3D Mesh Generation from a Single Image with Sparse-view Large Reconstruction Models”, arXiv 2024

Optimal Furniture Layout Estimation

- Use LEGO-Net to generate for our custom floor plans and furniture objects
- LEGO-Net follows a Transformer-based architecture to predict organized furniture positions and orientations



Random Furniture Initialization

Organized Furniture Placement

QA Wei et al. “LEGO-Net: Learning Regular Rearrangements of Objects in Rooms”, CVPR 2023

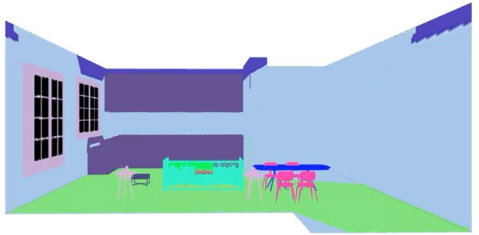
Physics Based Rendering

Reasons for using Open3D :

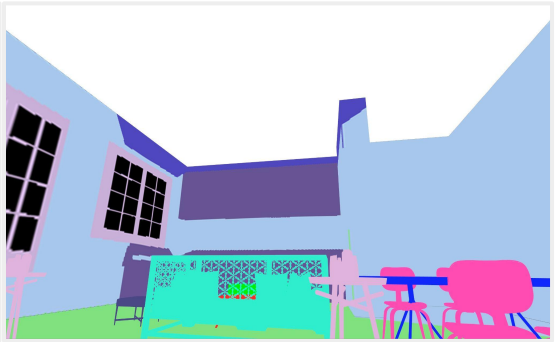
- Accurate Spatial Representation
- Flexibility in Customization
- Realistic Visualization

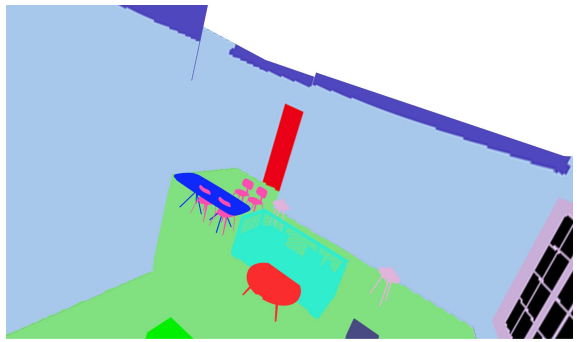
Pipeline Outcomes

- Render high fidelity details in RGB and Semantic map with 200 floor plans with permutation of 22 object classes
- Render multiple viewpoints of same indoor scene under various lighting conditions
- Account better for occlusions of objects from camera positions
- Output additional data such as depth maps, camera configurations



High Fidelity semantic labels

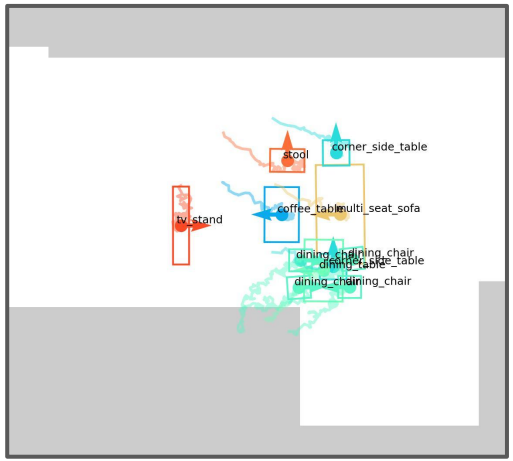




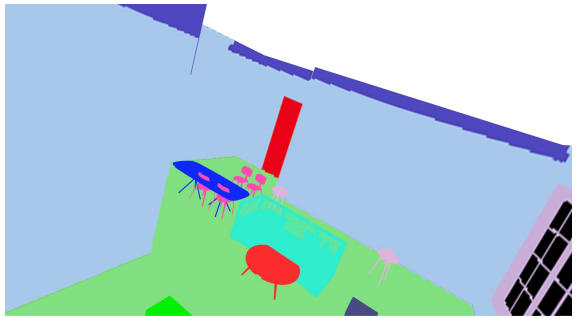
Semantic annotations from multiple viewpoints

Future work

- Using our pipeline for indoor scene understanding:
  - Export equirectangular images of the 3D renderings
  - Evaluate generated dataset on downstream tasks such as indoor scene segmentation
- Refining the data generation pipeline:
  - Fine-tune legonet with floor plans from Structured 3D dataset and random furniture initializations



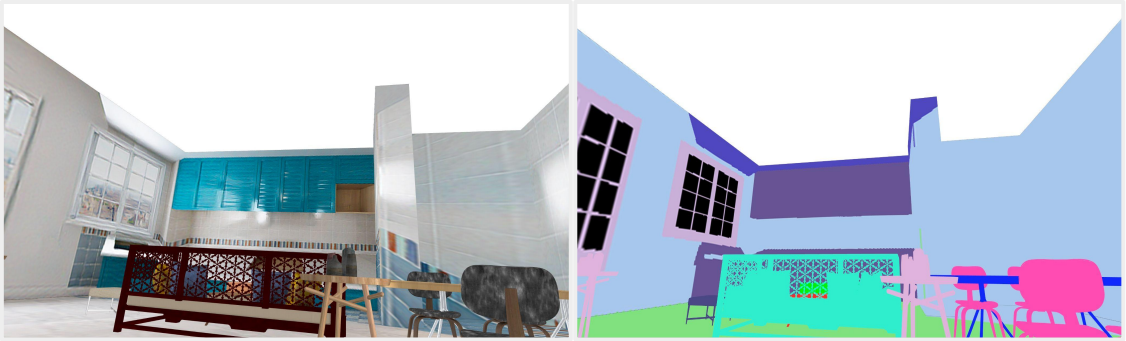
360 Panoramas & Equirectangular images

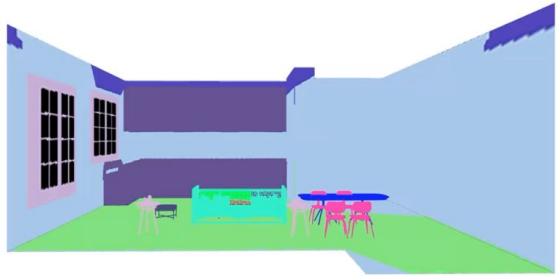


3D Rendering: Advantages

- Render high fidelity details in RGB and Semantic map
- Account for better occlusions of objects from camera positions

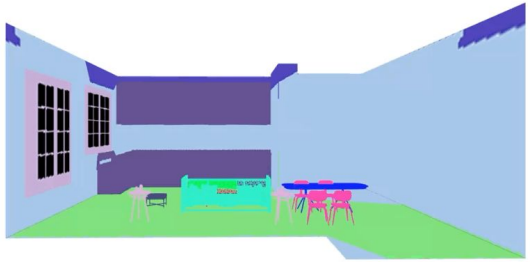
High Fidelity semantic labels



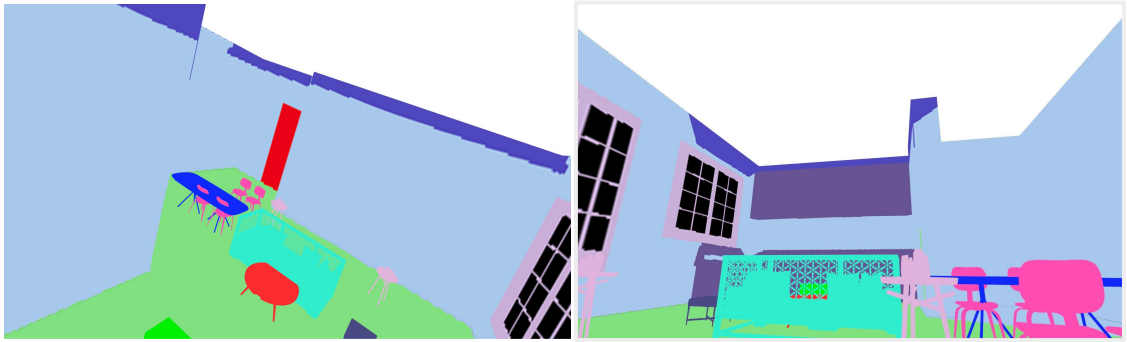


3D Rendering: Advantages

- Render multiple viewpoints of same indoor scene under various physics based simulations
- Output additional data such as depth maps, camera configurations



Semantic annotations from multiple view points

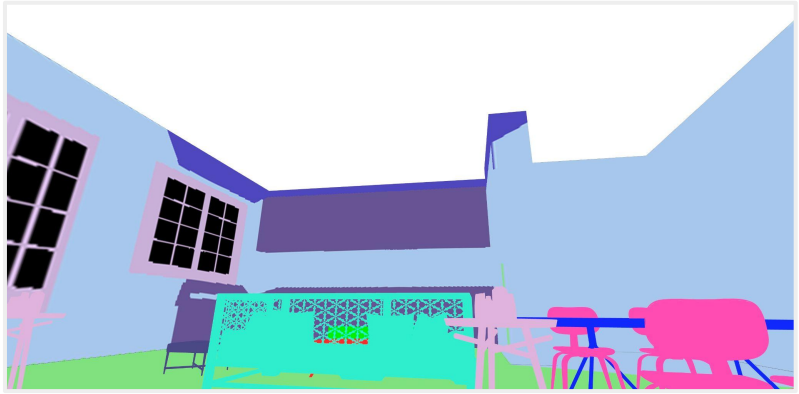


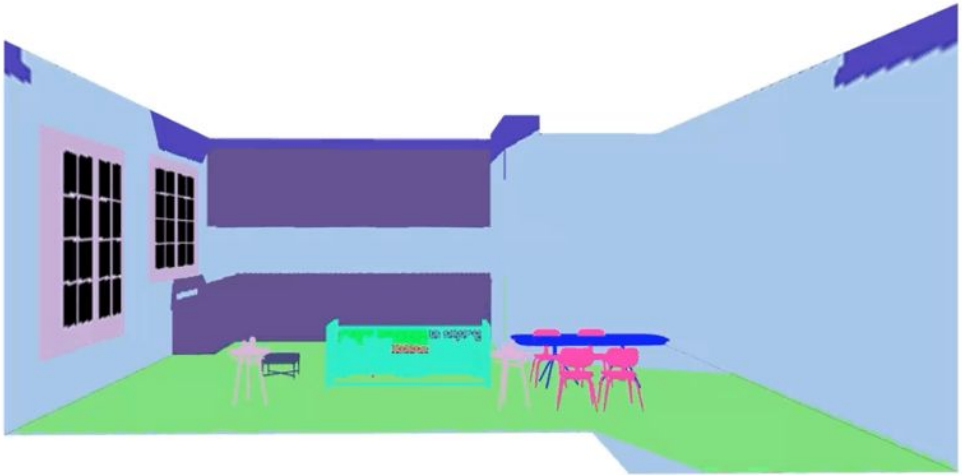
3D Rendering: Annotations

High fidelity textured images from multiple viewpoints

3D Rendering: Annotations

Semantic annotations from multiple viewpoints

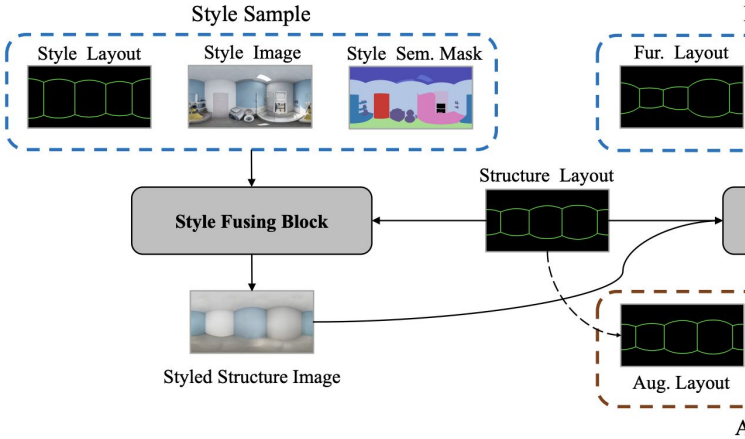




PanoMixSwap

Augmentation done by applying a stretching technique to each of the walls in the transformed layout.

This leads to distortion in the augmented images as shown below.



3D model texture outputs