ATISS: Autoregressive Transformers for Indoor Scene Synthesis

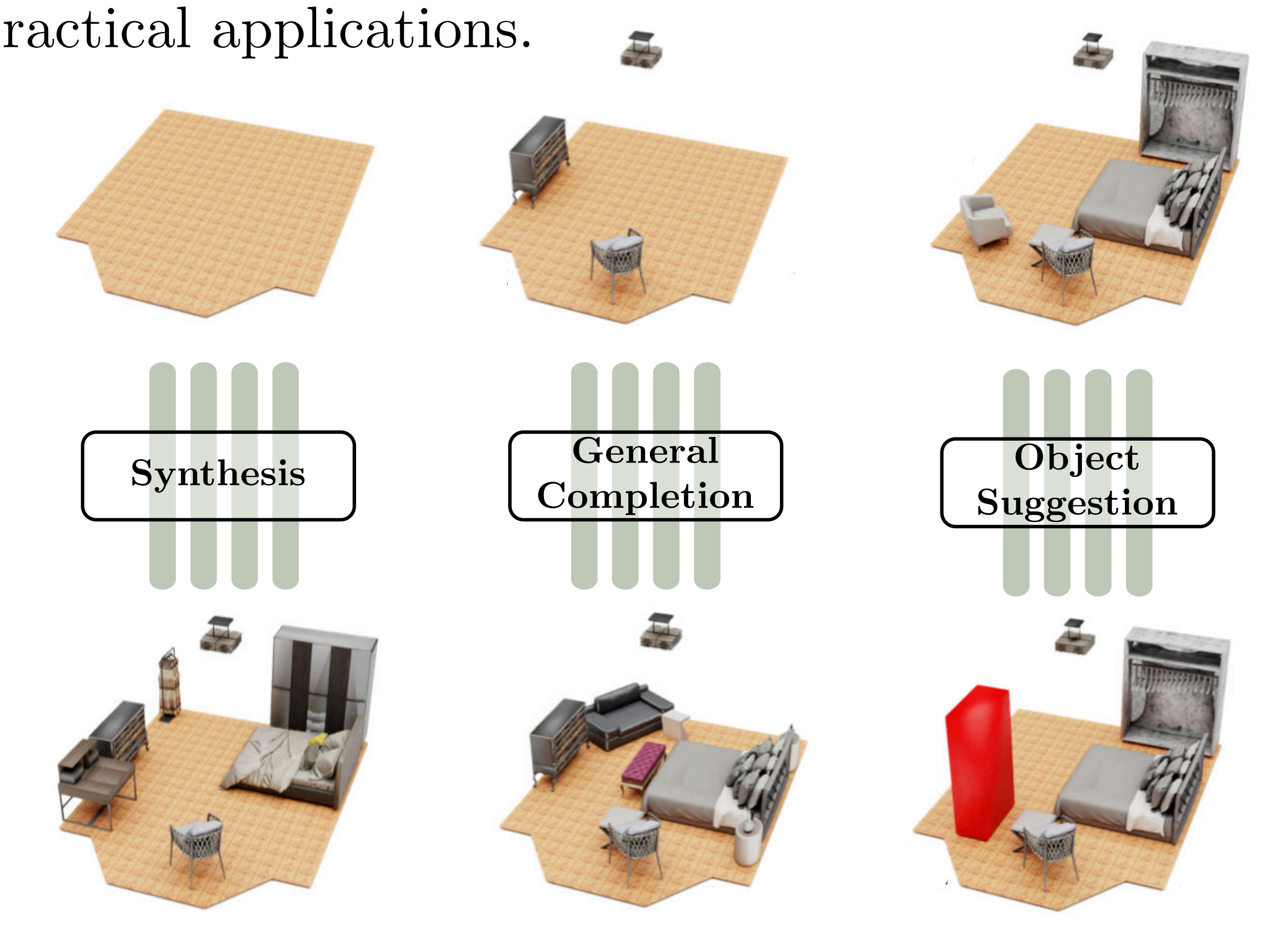
Despoina Paschalidou[1,3,4] Amlan Kar[4,5,6] Maria Shugrina[4]
Karsten Kreis[4] Andreas Geiger[1,2,3] Sanja Fidler[4,5,6]

[1]MPI for Intelligent Systems Tübingen [2]University of Tübingen [3]Max Planck ETH Center for Learning Systems [4]NVIDIA [5]University of Toronto [6]Vector Institute

https://nv-tlabs.github.io/ATISS/

Motivation

Existing scene synthesis pipelines represent scenes as **ordered sequences of objects**, thus inhibiting practical applications.

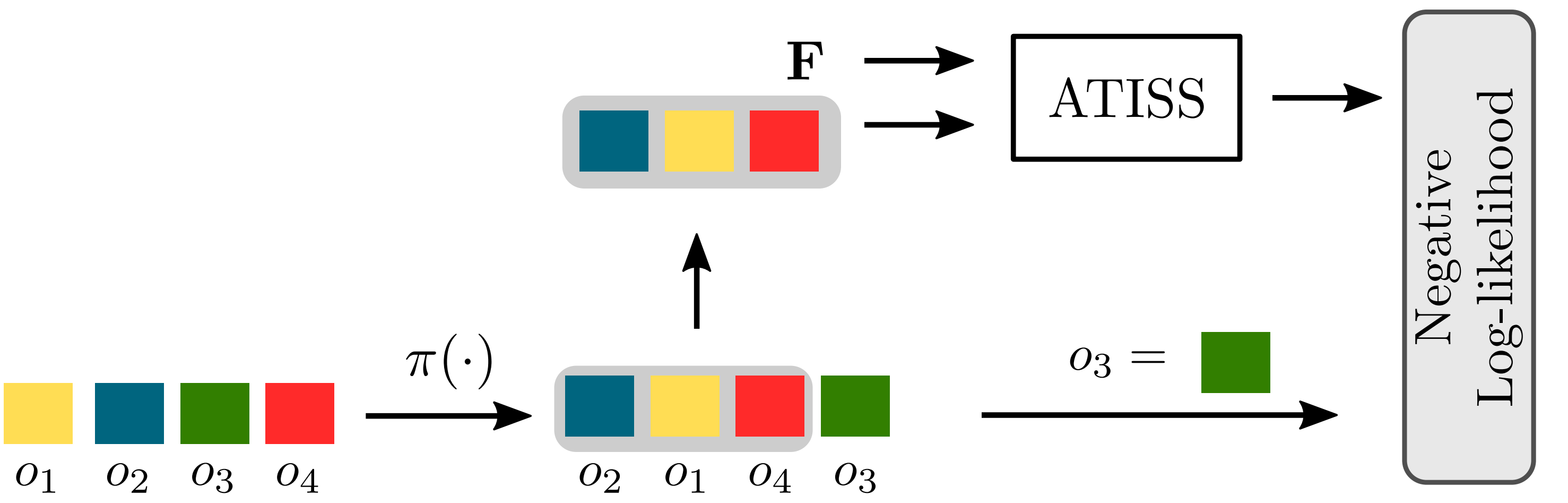


Contributions:

- Pose scene synthesis as autoregressive set generation.
- **State-of-the-art results on the scene synthesis.**
- Enables **new interactive applications**.
- **Renders a new scene up to 8 times faster.**

Training Overview

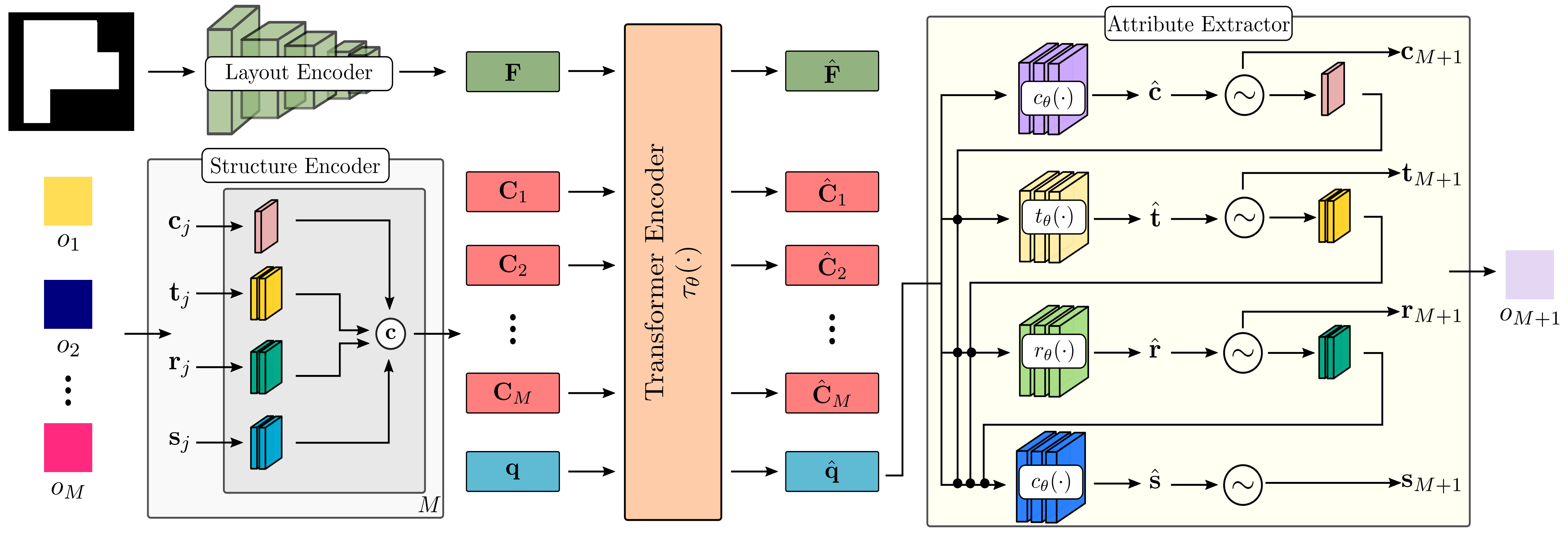
Each scene comprises an unordered set of objects in the scene $\mathcal{O} = \{o_j^i\}_{j=1}^M$ and its floor layout **F**.



Each object is represented as a **3D labelled bounding box**, modelled with four random variables that control its category **c**, size **s**, orientation **r** and location **t**.

Our Method

Our key idea is to **pose the scene synthesis task as an unordered set generation problem**. Given a room type and its shape, ATISS generates furniture arrangements by autoregressively placing objects in a **permutation-invariant fashion**.



The object attributes are **generated in an autoregressive manner:**

$$p_\theta(o_j \mid o_{<j}, \mathbf{F}) = p_\theta(\mathbf{c}_j | o_{<j}, \mathbf{F}) p_\theta(\mathbf{t}_j | \mathbf{c}_j, o_{<j}, \mathbf{F}) p_\theta(\mathbf{r}_j | \mathbf{c}_j, \mathbf{t}_j, o_{<j}, \mathbf{F}) p_\theta(\mathbf{s}_j | \mathbf{c}_j, \mathbf{t}_j, \mathbf{r}_j, o_{<j}, \mathbf{F}).$$

The **likelihood of generating a scene with any order** is:

$$p_\theta(\mathcal{O}|\mathbf{F}) = \sum_{\hat{\mathcal{O}} \in \pi(\mathcal{O})} \prod_{j \in \hat{\mathcal{O}}} p_\theta(o_j \mid o_{<j}, \mathbf{F})$$

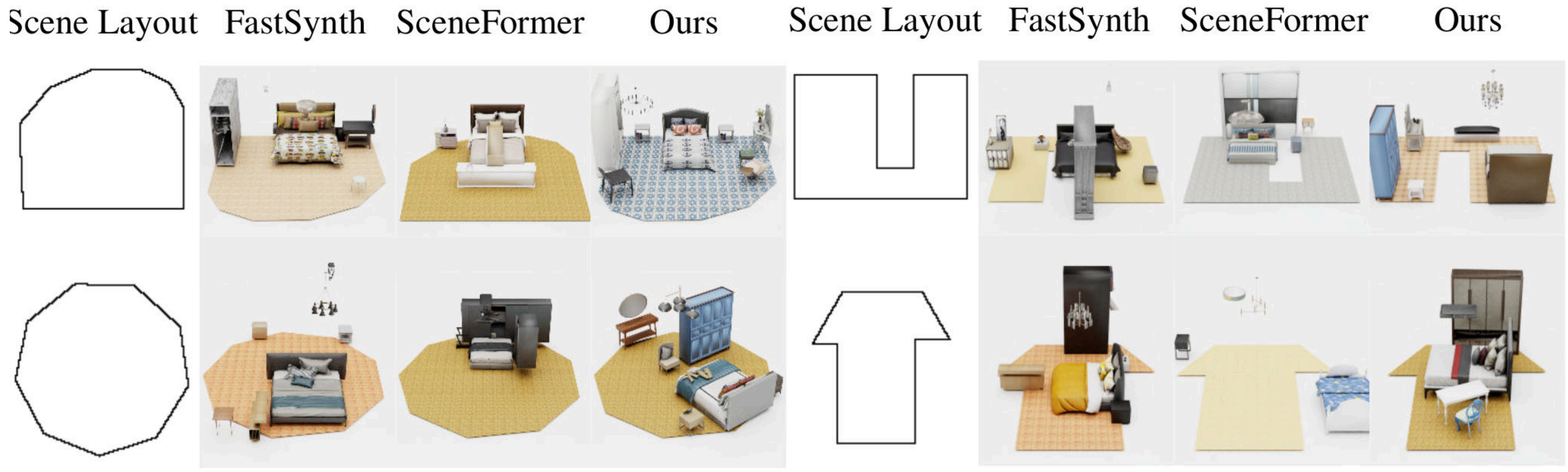
The **likelihood of generating a scene with all possible orders** is:

$$\hat{p}_\theta(\mathcal{O}|\mathbf{F}) = \prod_{\hat{\mathcal{O}} \in \pi(\mathcal{O})} \prod_{j \in \hat{\mathcal{O}}} p_\theta(o_j \mid o_{<j}, \mathbf{F})$$

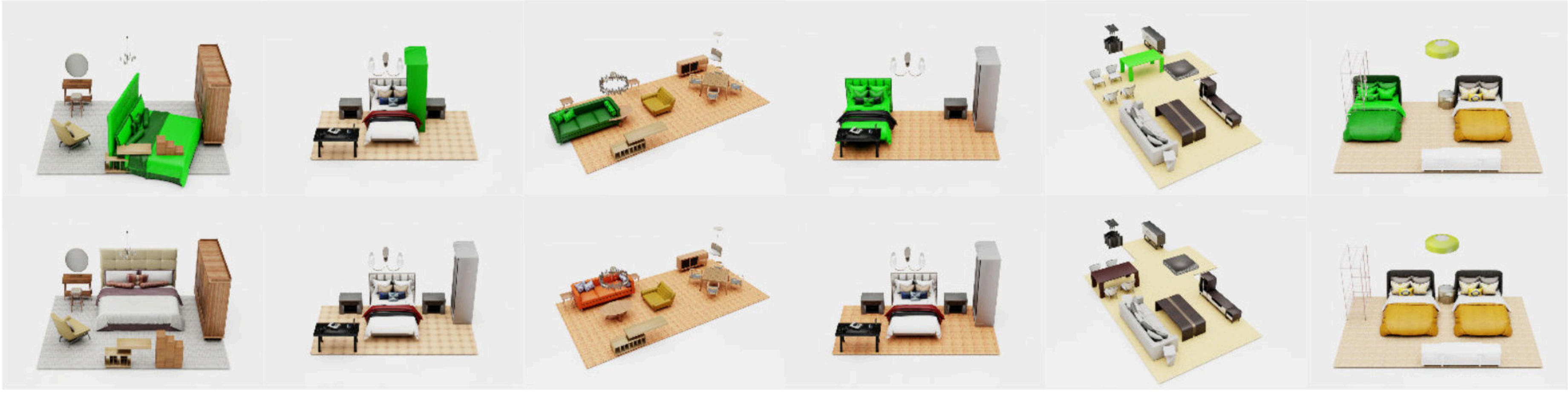
Scene Synthesis Results

Interactive Applications

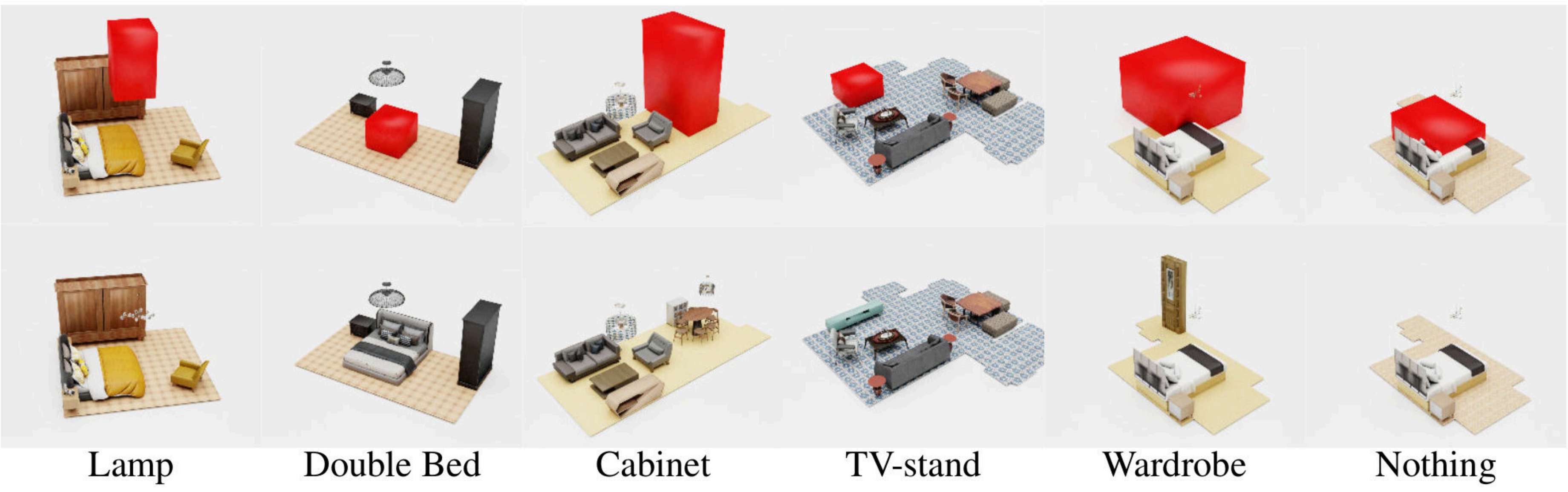
Generalization Beyond Training Data



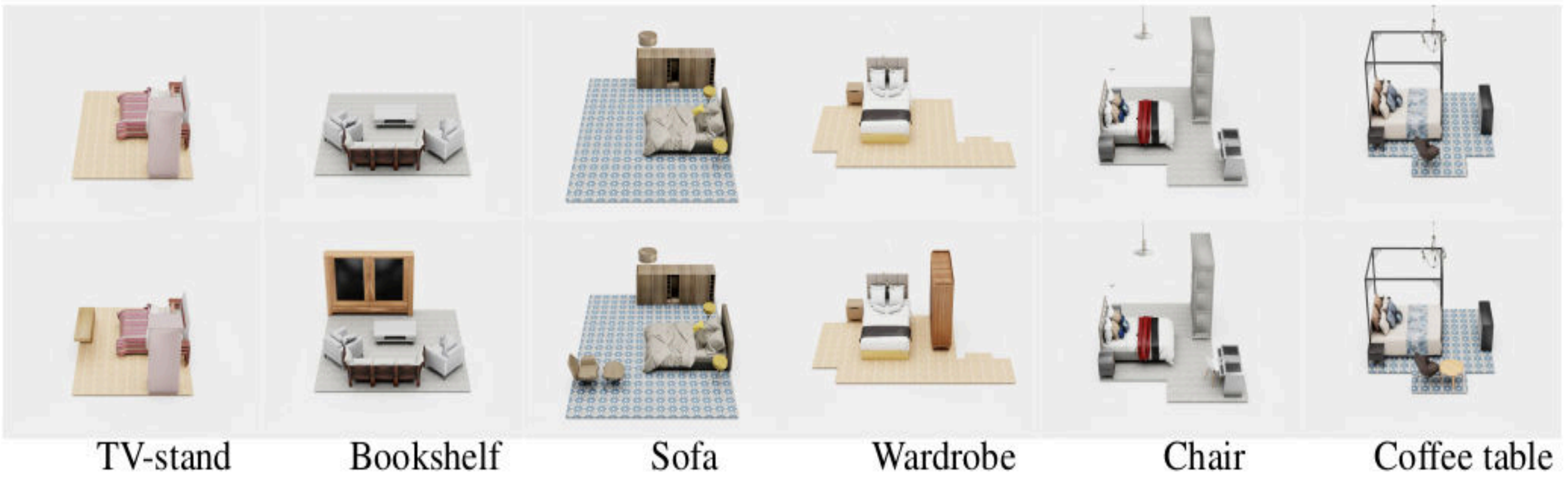
Failures Correction



Objects Suggestion



Object Placement


TV-stand
Bookshelf
Sofa
Wardrobe
Chair
Coffee table