PT43D: A Probabilistic Transformer for Generating 3D Shapes from Single Highly-Ambiguous RGB Images

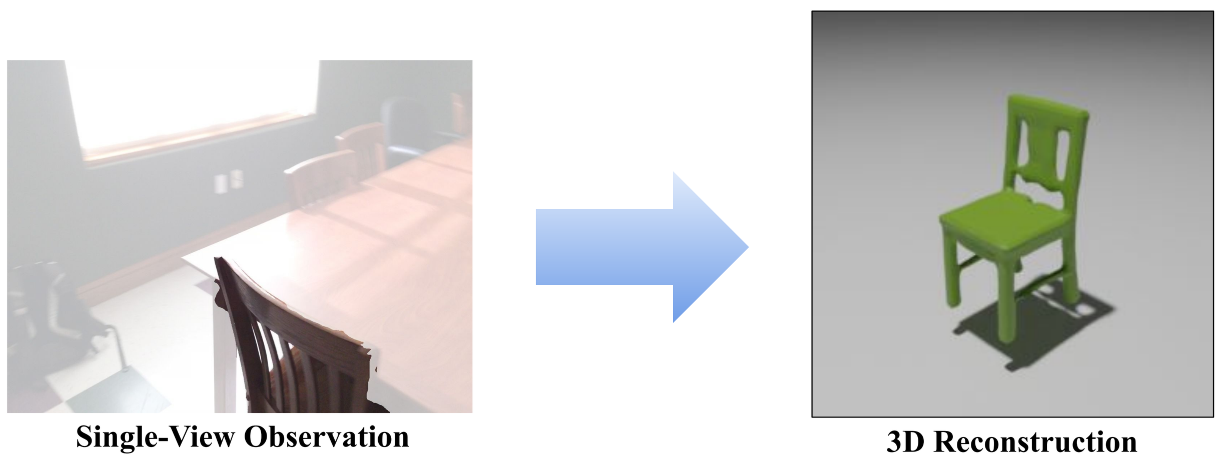
BMVC 2024

paper, code and data

Yiheng Xiong   Angela Dai
Technical University of Munich

Problem

Reconstructing 3D structure and geometry from **single-view**, **highly-ambiguous** object observations.

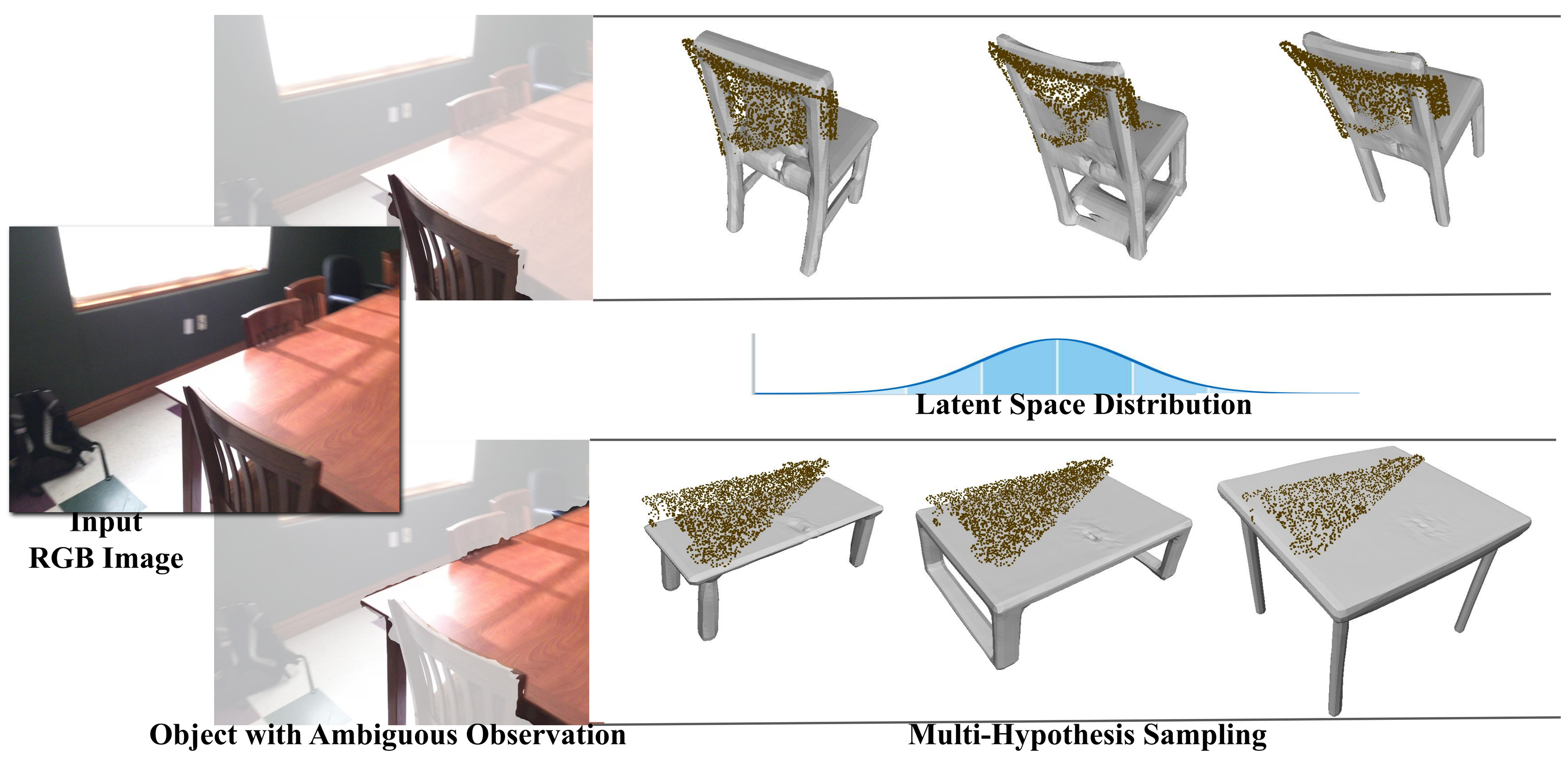


Challenges

Contributions

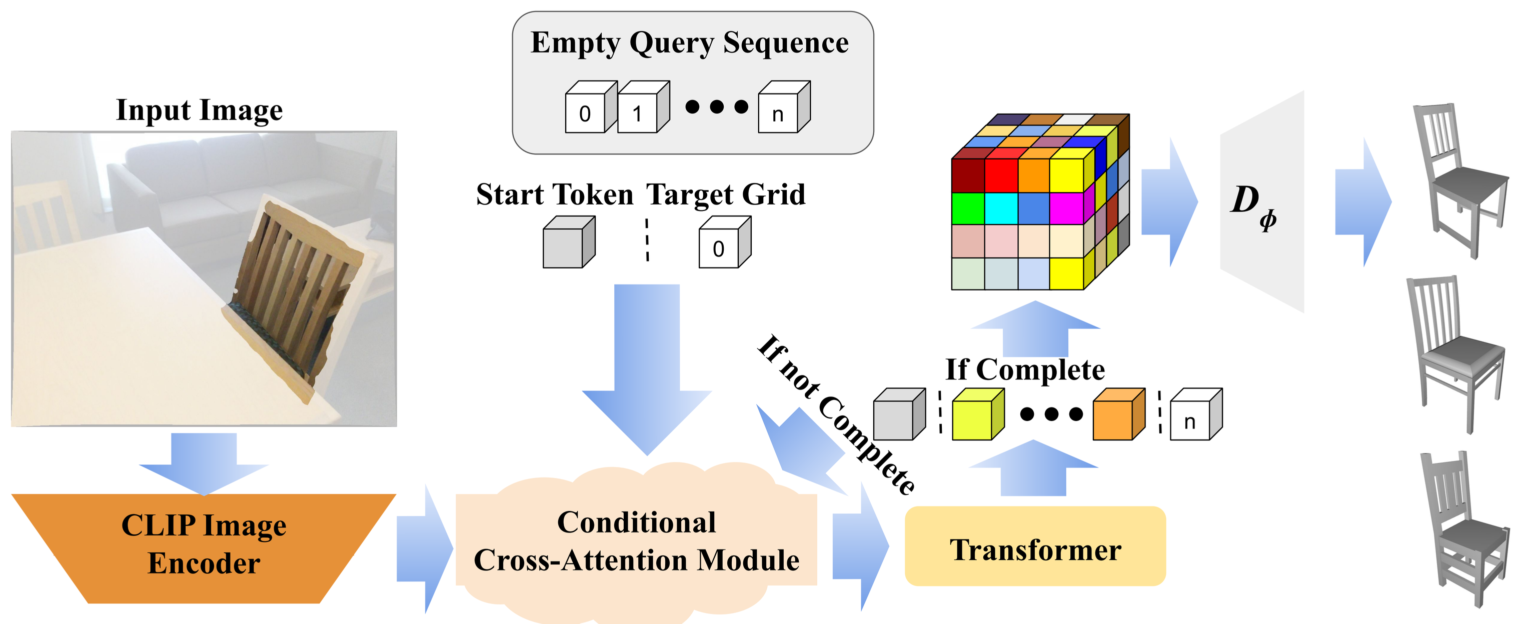
- **Probabilistic** approach for shape generation from **single highly-ambiguous** RGB images.
- **Synthetic** data augmentation for **improved real-world** fine-tuning.

Solution: Multi-Hypothesis Sampling



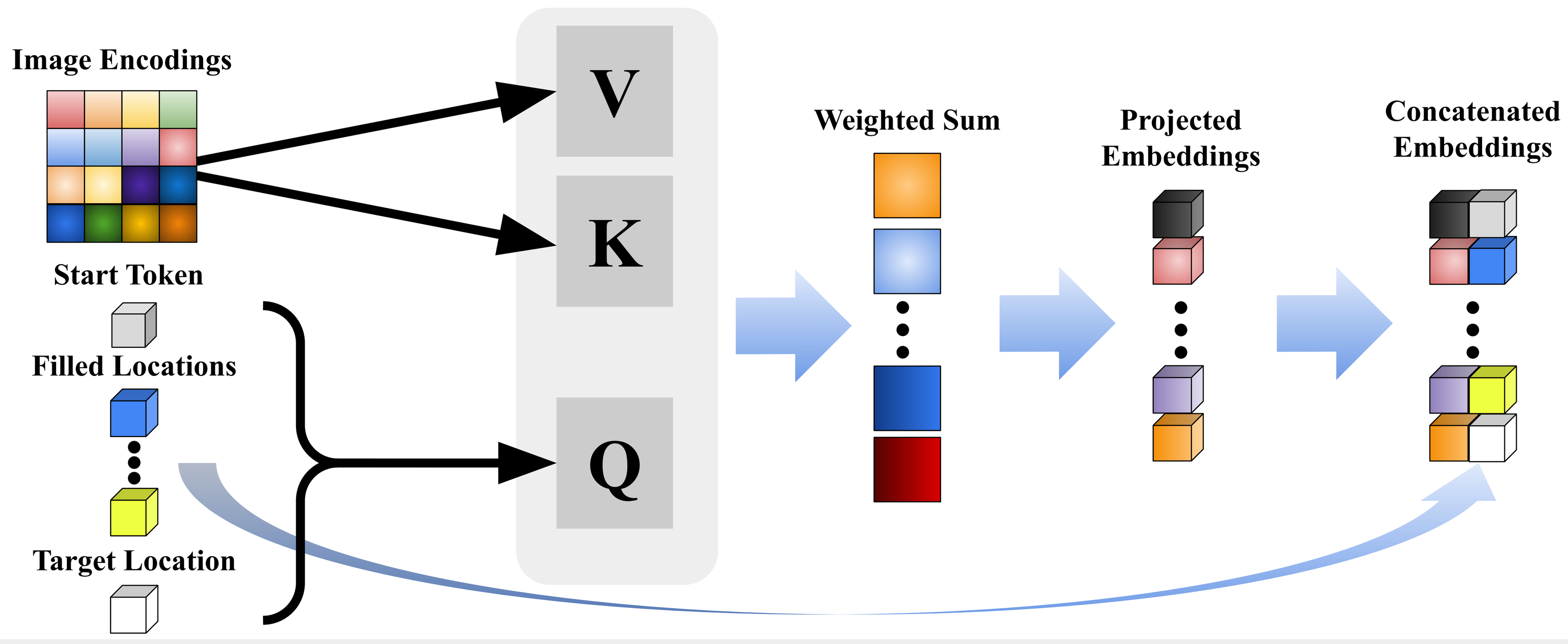
Our approach seeks to model the **distribution** of 3D shapes in **latent space** conditioned on a single highly ambiguous image, enabling the sampling of **multiple diverse hypotheses** during inference.

Probabilistic Shape Generation Approach



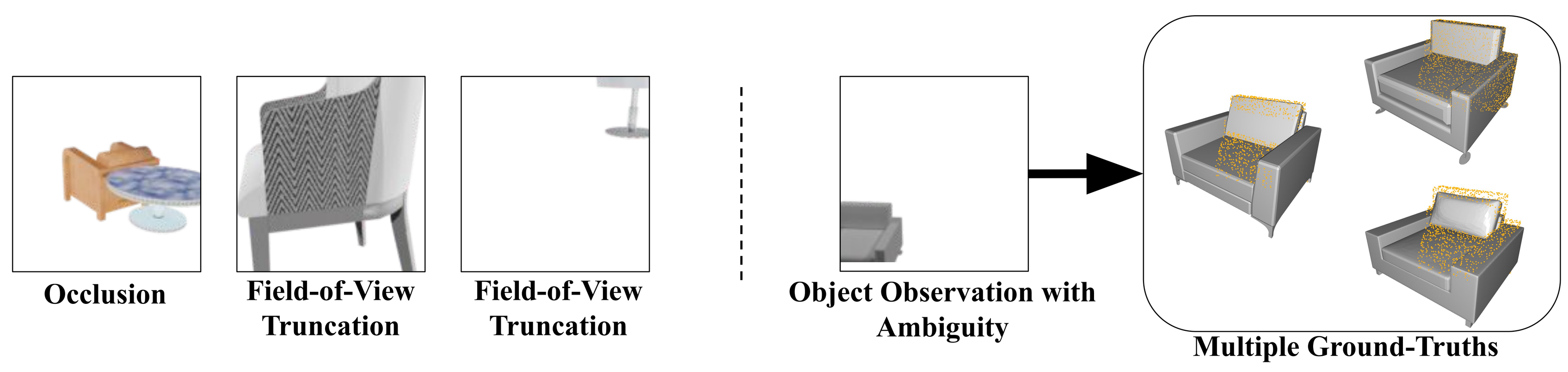
Our probabilistic approach uses transformer with conditional cross-attention to **autoregressively** predict a **distribution** in latent space for **each grid** from an input RGB image, enabling the generation of multiple plausible shape hypotheses.

Identifying the Most Relevant Regions in the Image via Conditional Cross-Attention



We perform **cross-attention** between image encodings and the input sequence to effectively **identify the most relevant region of interest** from the input image.

Synthetic Data Augmentation



Comparisons to State-of-the-Art

Our aim is to generate shapes that capture the **distribution** of possible reconstructions that explain an input image observation. Thus, we evaluate both **diversity** of reconstructed shapes as well as their reconstruction **quality**. Our method generates **higher** quality shapes compared with other baselines, with a **more reasonable diversity**, both quantitatively and qualitatively.

	Synthetic Data from ShapeNet				Real-World Data from ScanNet			
	TMD↑	bi-CD↓	F-score↑	visible CD↓	TMD↑	bi-CD↓	F-score↑	visible CD↓
AutoSDF	0.045	5.04	0.215	-	0.021	8.32	0.164	-
SDFusion	**0.091**	4.52	0.194	6.71	**0.132**	4.97	0.200	10.4
Ours	0.062	**4.41**	**0.224**	**5.41**	0.096	**3.61**	**0.228**	**5.89**