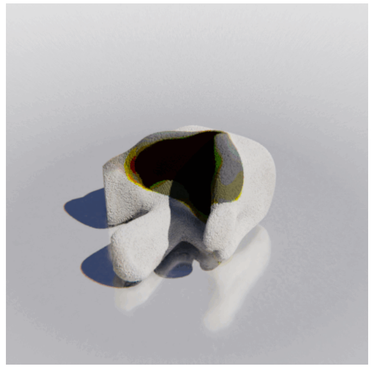
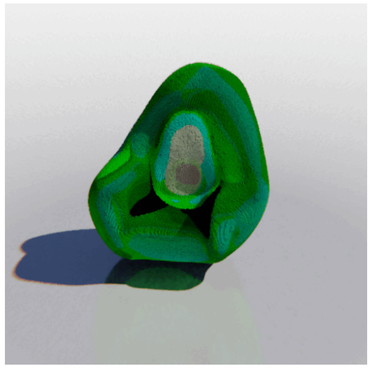
Multiview Compressive Coding (MCC) for 3D Reconstruction

@ WED-AM-080




Chao-Yuan Wu


Justin Johnson

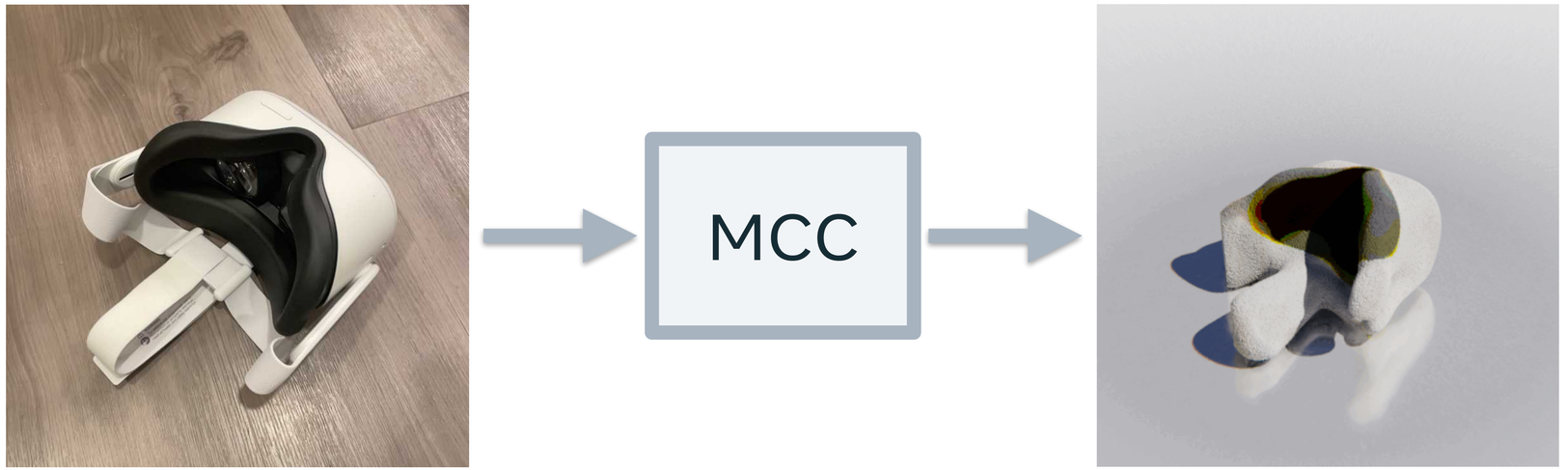

Jitendra Malik


Christoph Feichtenhofer


Georgia Gkioxari

Meta

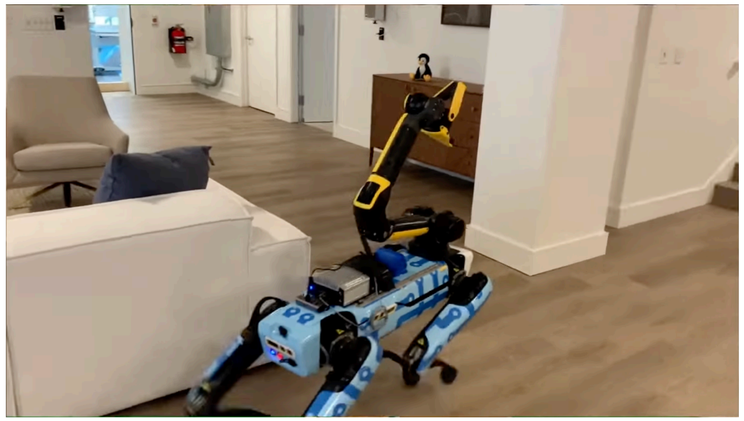
What is MCC?



RGB-D image

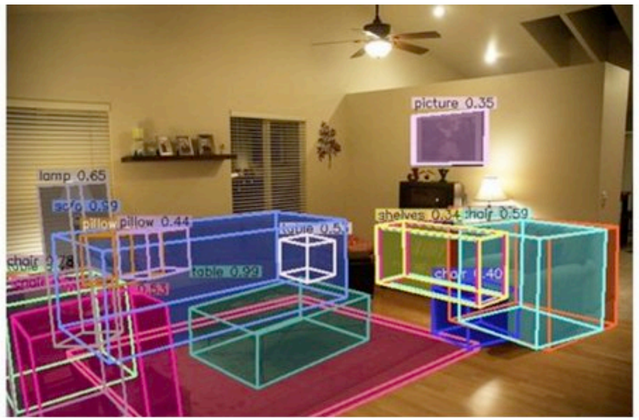
3D reconstruction

Why 3D Reconstruction?



Robotics

Mix Reality



Visual Reasoning

3D Asset Creation

Image sources: https://ai.facebook.com/blog/robots-learning-video-simulation-artificial-visual-cortex-vc-1/, Meta, Omni3D, sketchfab.com (Creative Commons)

Comparison to Prior Works

- Category-specific
- Rely on full 3D ground-truth

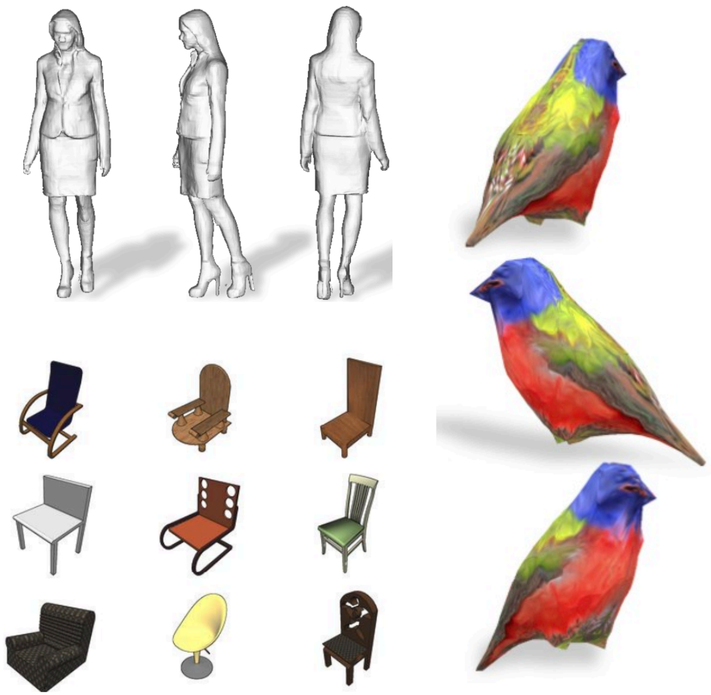
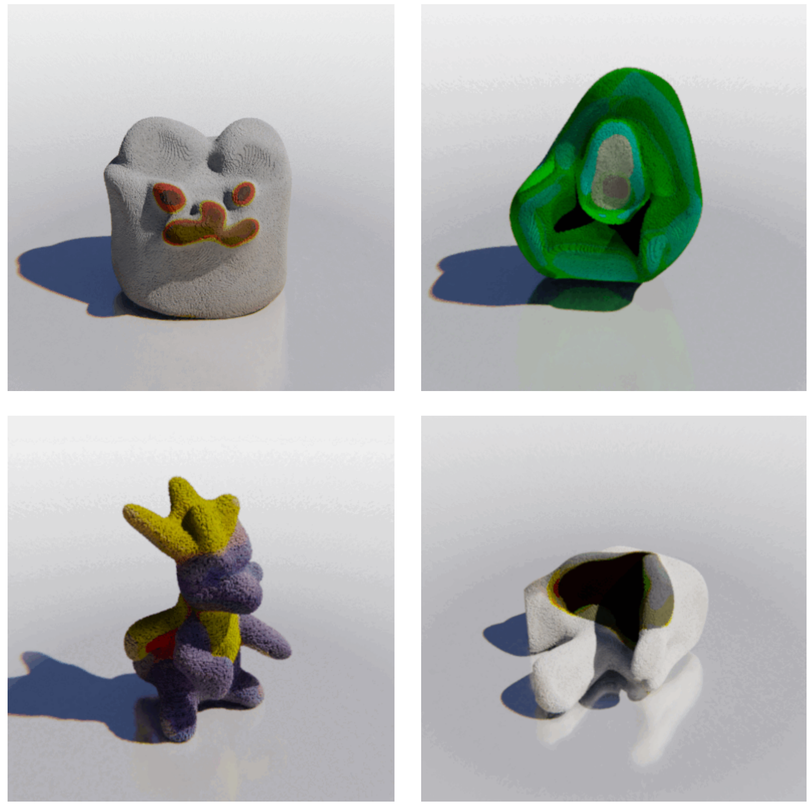
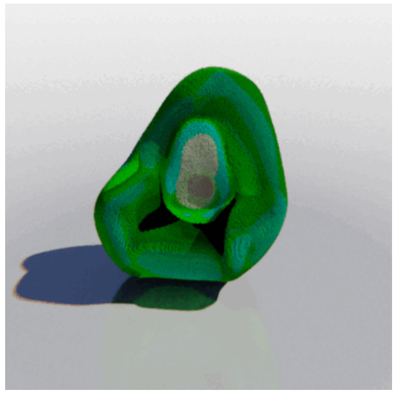
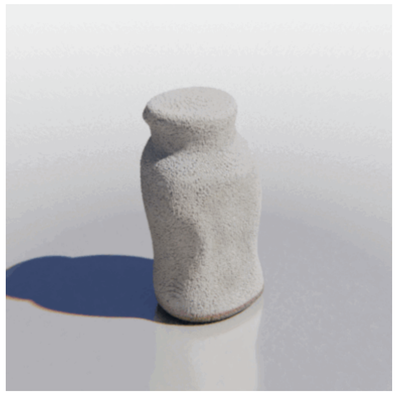
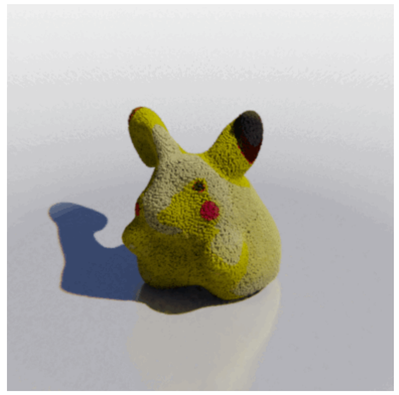
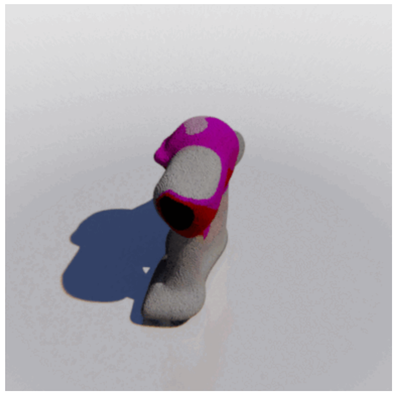


Image sources: “PIFu”, Saito et al. ICCV ’19, “CMR”, Kanazawa et al. ECCV ’18, “ShapeNet”, Chang et al., arXiv ’15.

- **Category-agnostic**
- **Trained on video frames**



Strong generalization to unseen objects

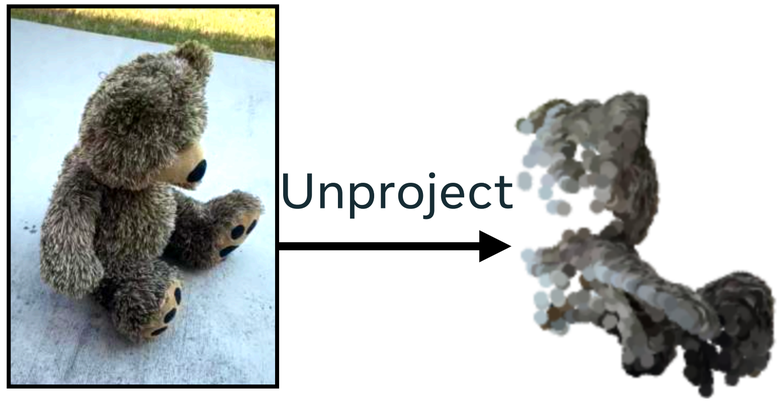
Input Image	**MCC output**	Input Image	**MCC output**	Input Image	**MCC output**
					
					

Method

Method

Input RGB-D image

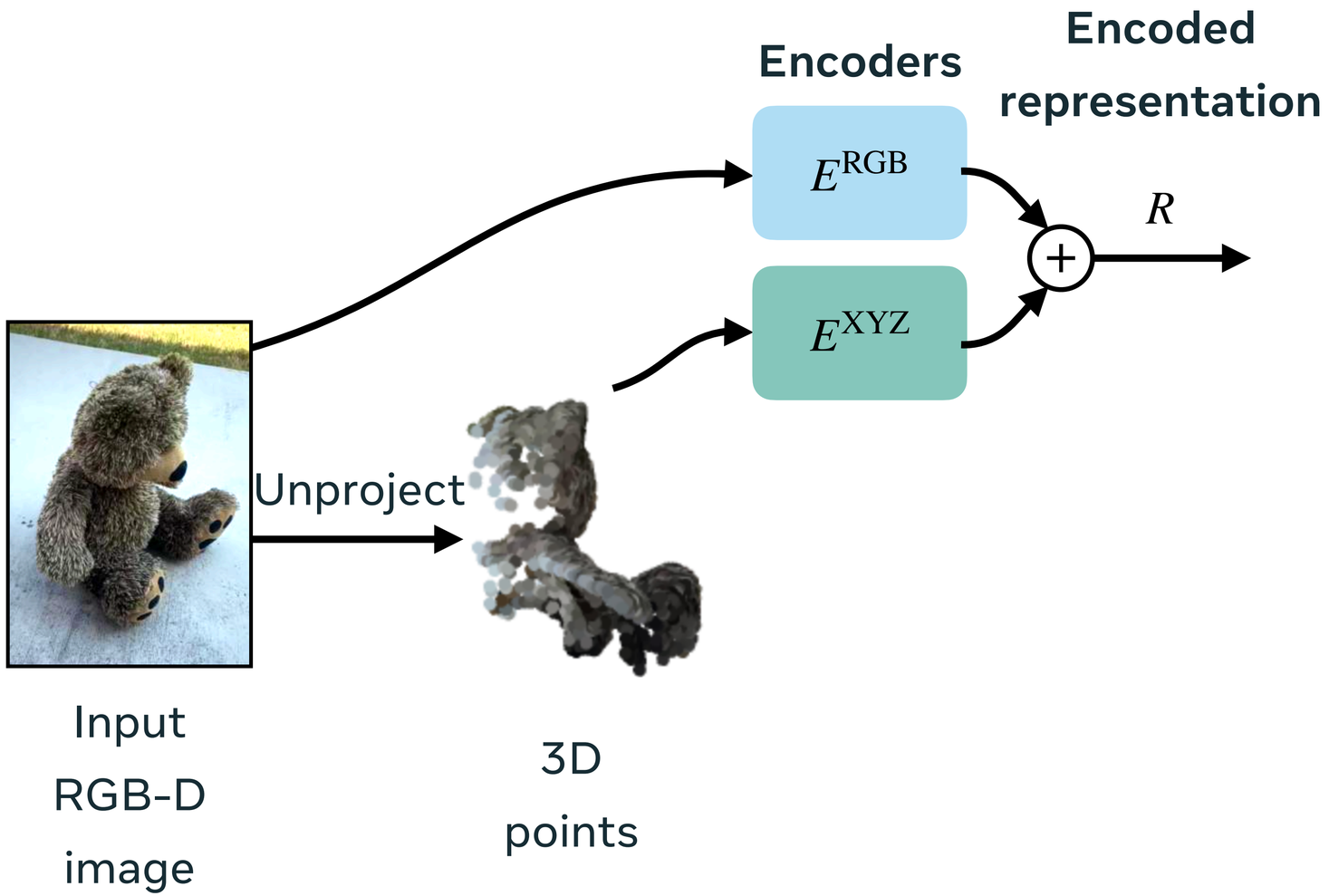
Method



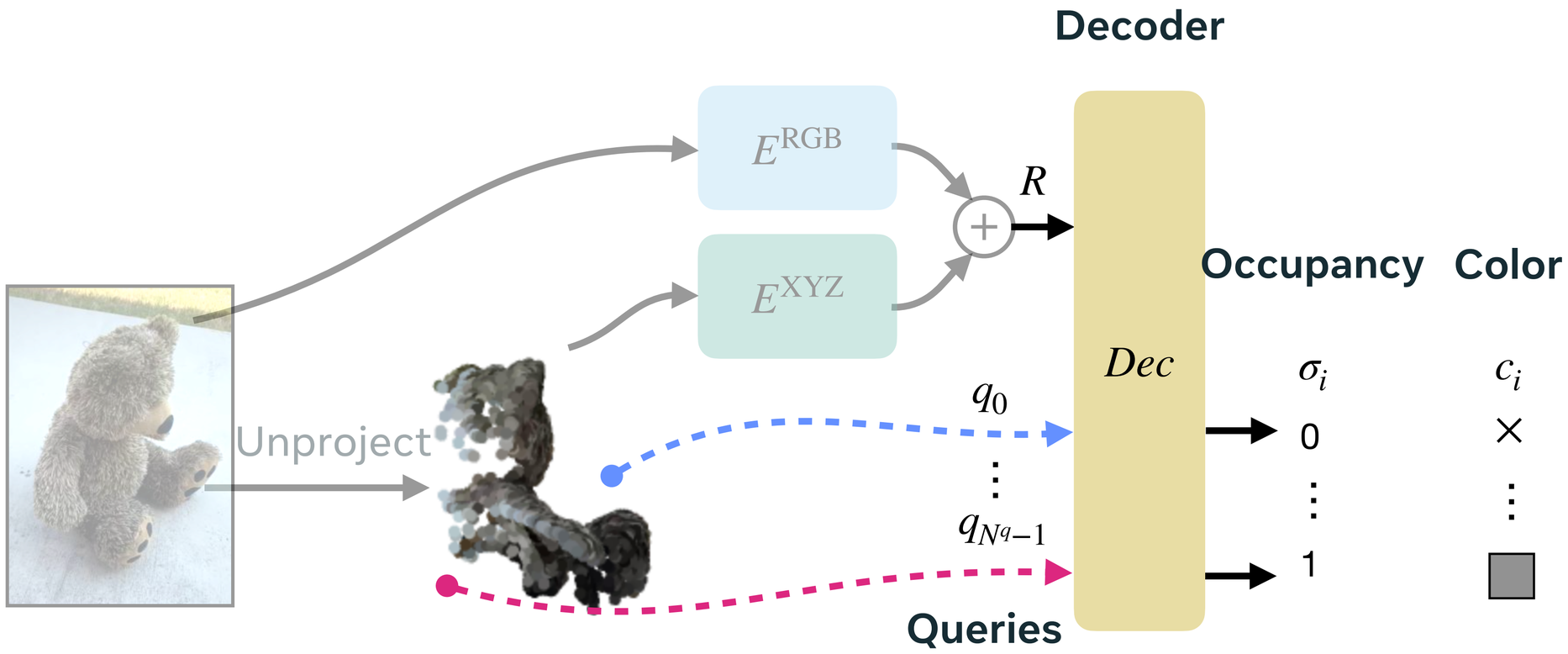
Input RGB-D image

3D points

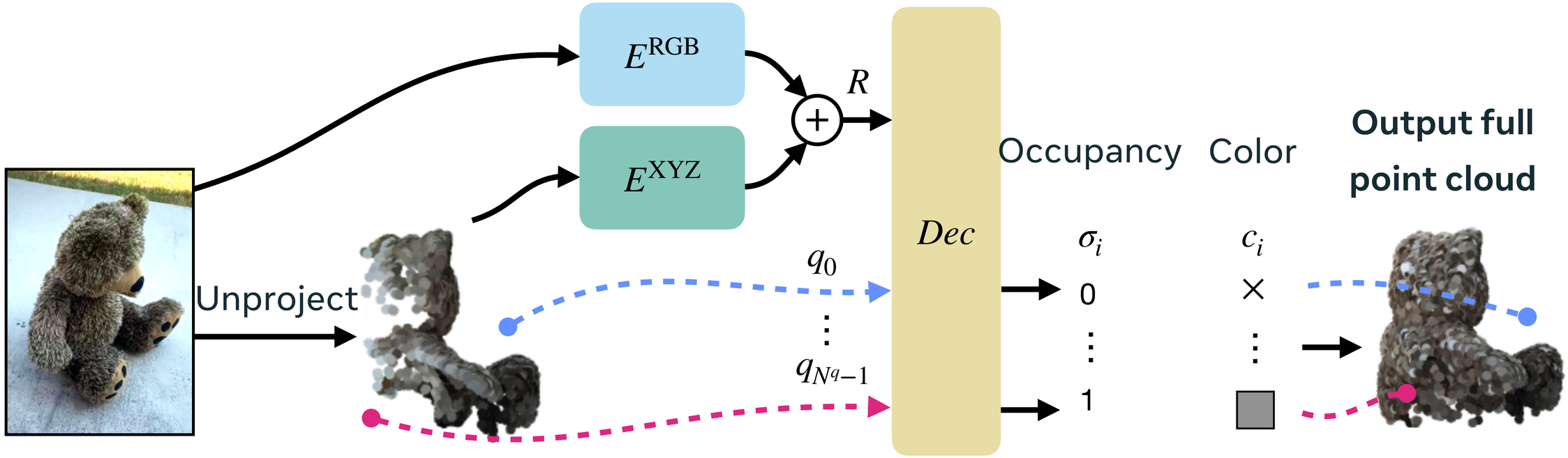
Method


Encoders
Encoded representation
E^{RGB}
E^{XYZ}
R
Unproject
Input RGB-D image
3D points

Method

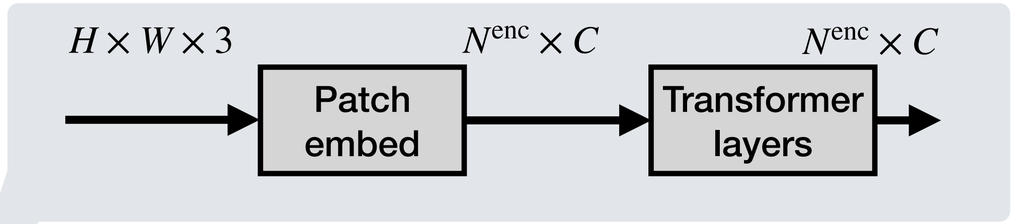

Decoder
E^{RGB}
E^{XYZ}
R
Unproject
q_0
q_{N^q-1}
Queries
Dec
Occupancy
Color
σ_i
c_i
0
1

Method

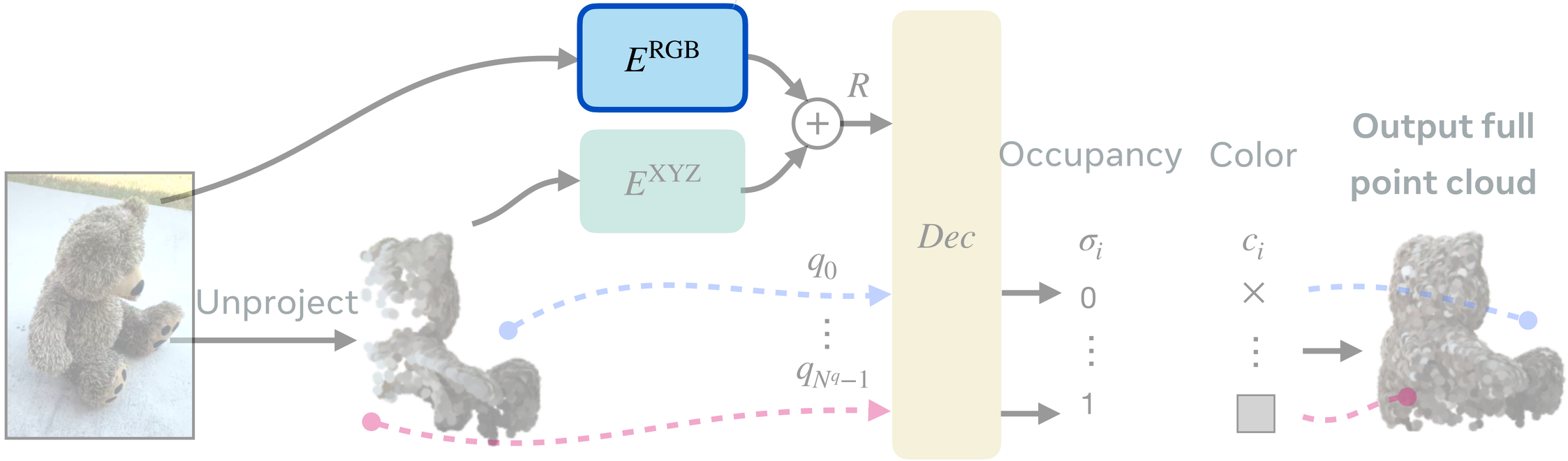

E^{RGB}
E^{XYZ}
R
Unproject
q_0
q_{N^q-1}
Dec
Occupancy
σ_i
0
1
Color
c_i
×
Output full point cloud

Architecture details

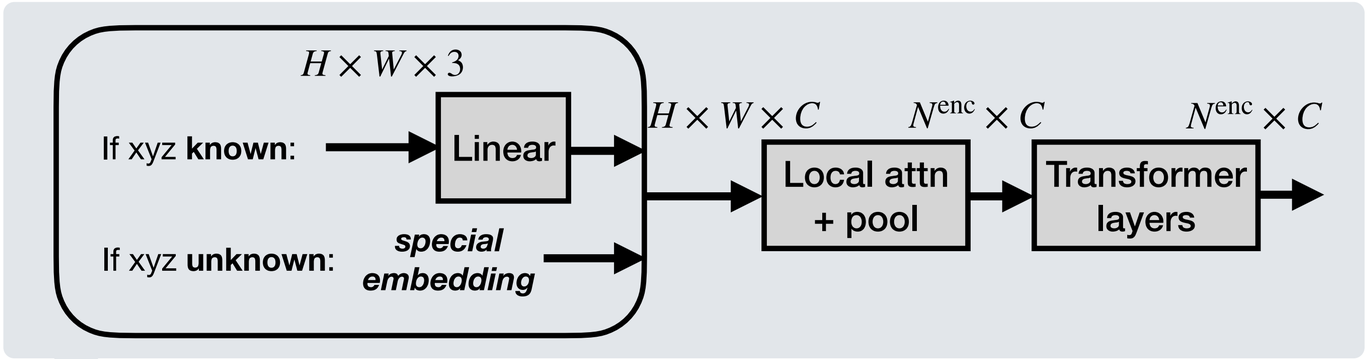
Architecture details

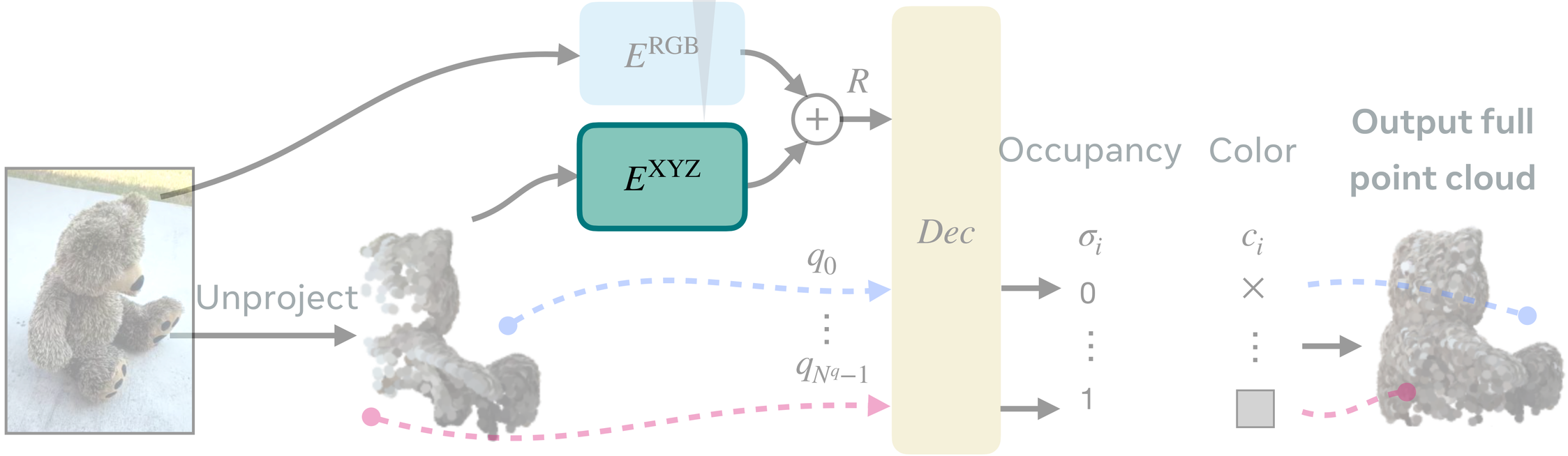


["ViT", Dosovitskiy et al., ICLR 2021]

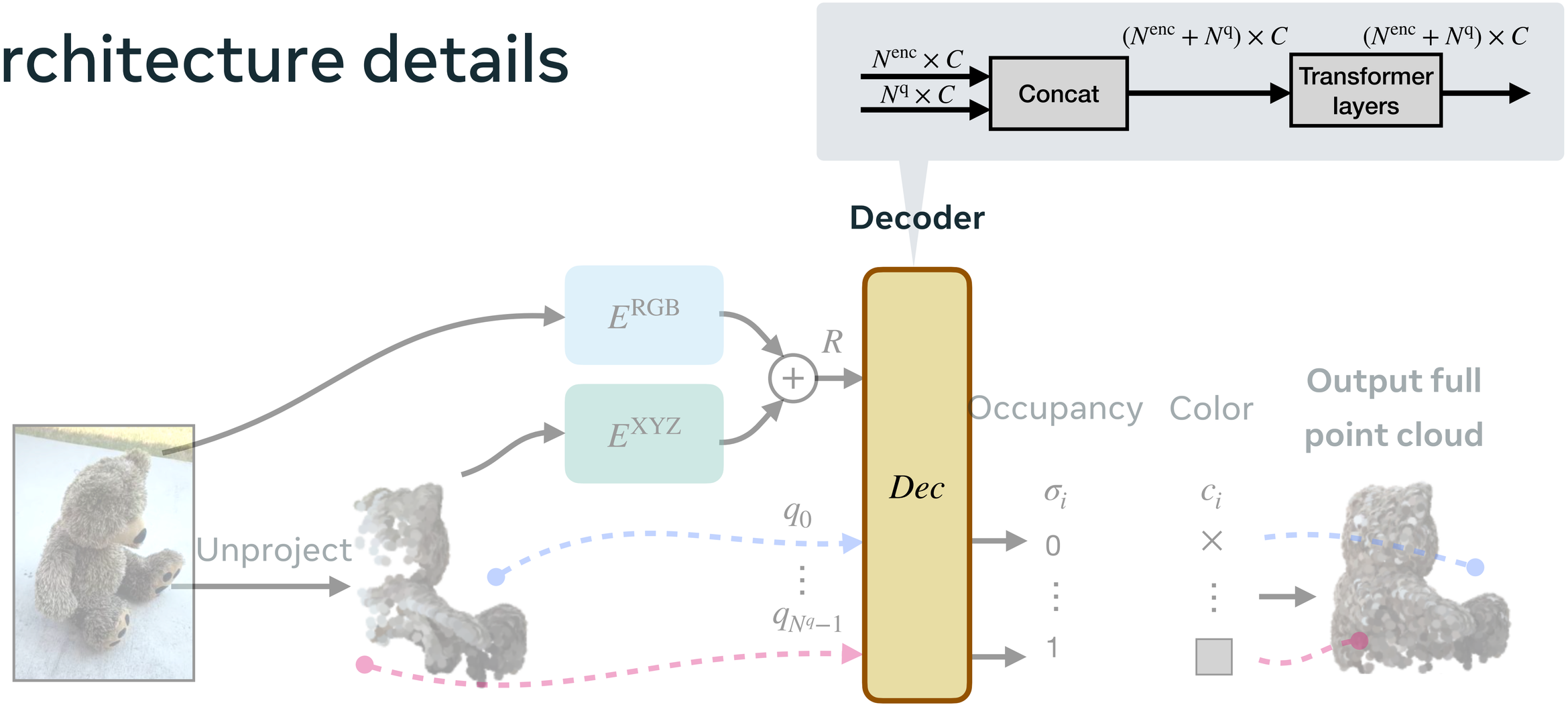


Architecture details





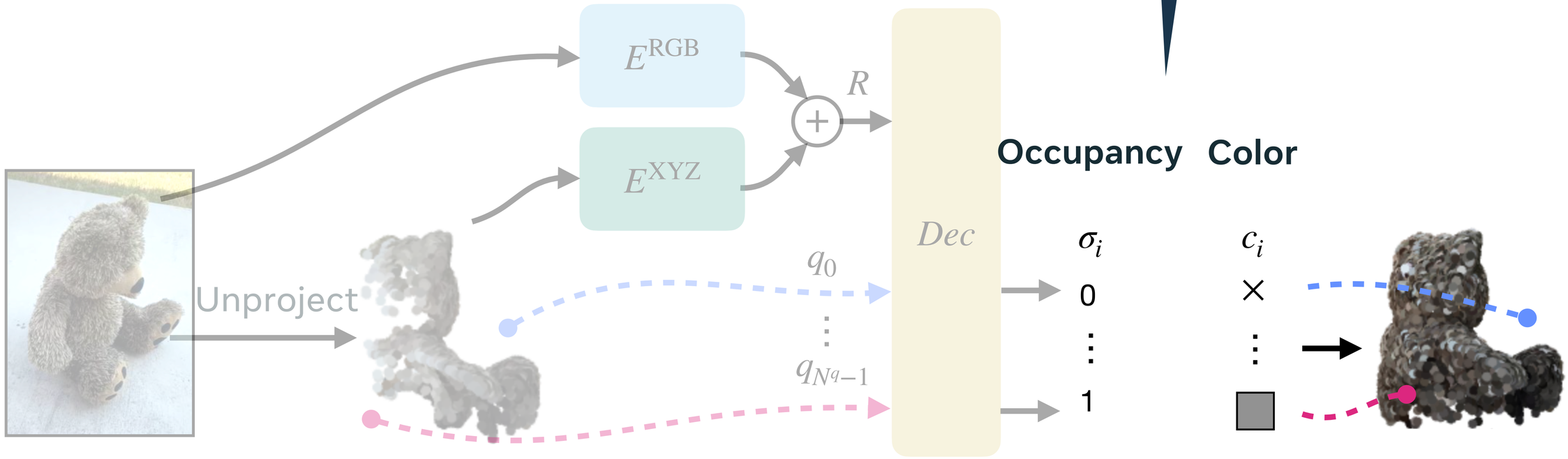
Architecture details



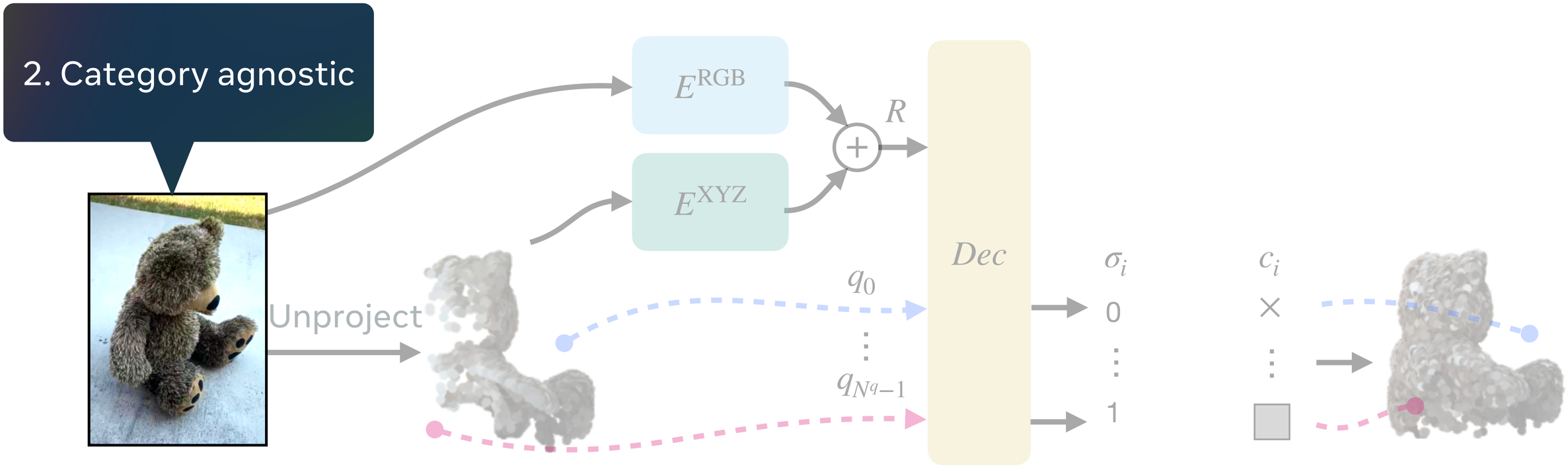
Key Properties

Key Properties

1. Needs only **points** as supervision

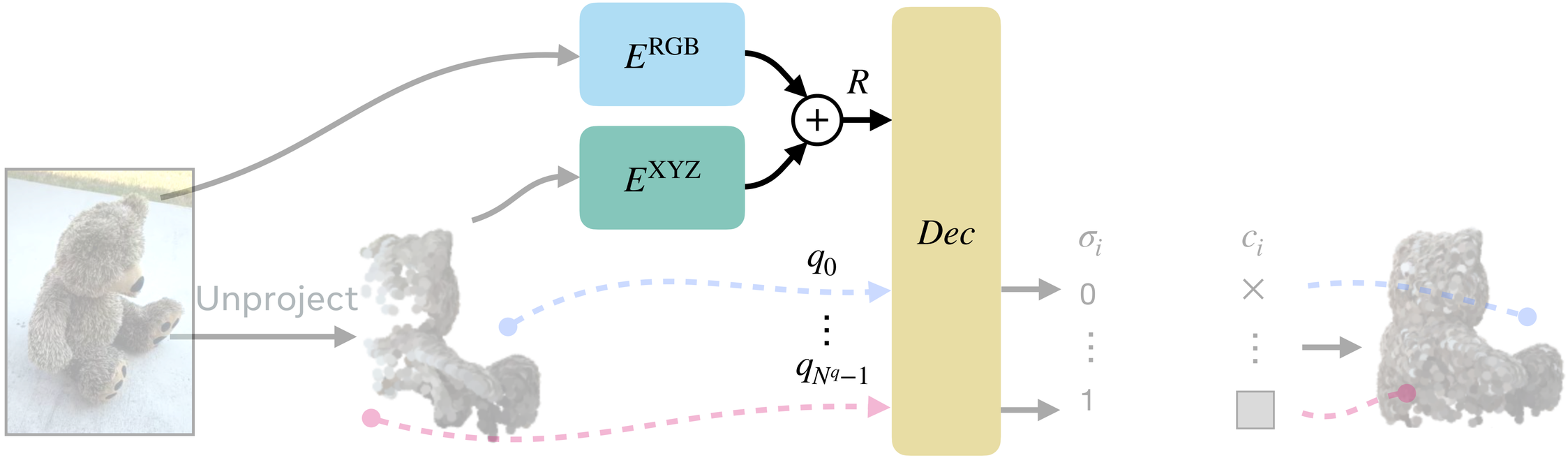


Key Properties

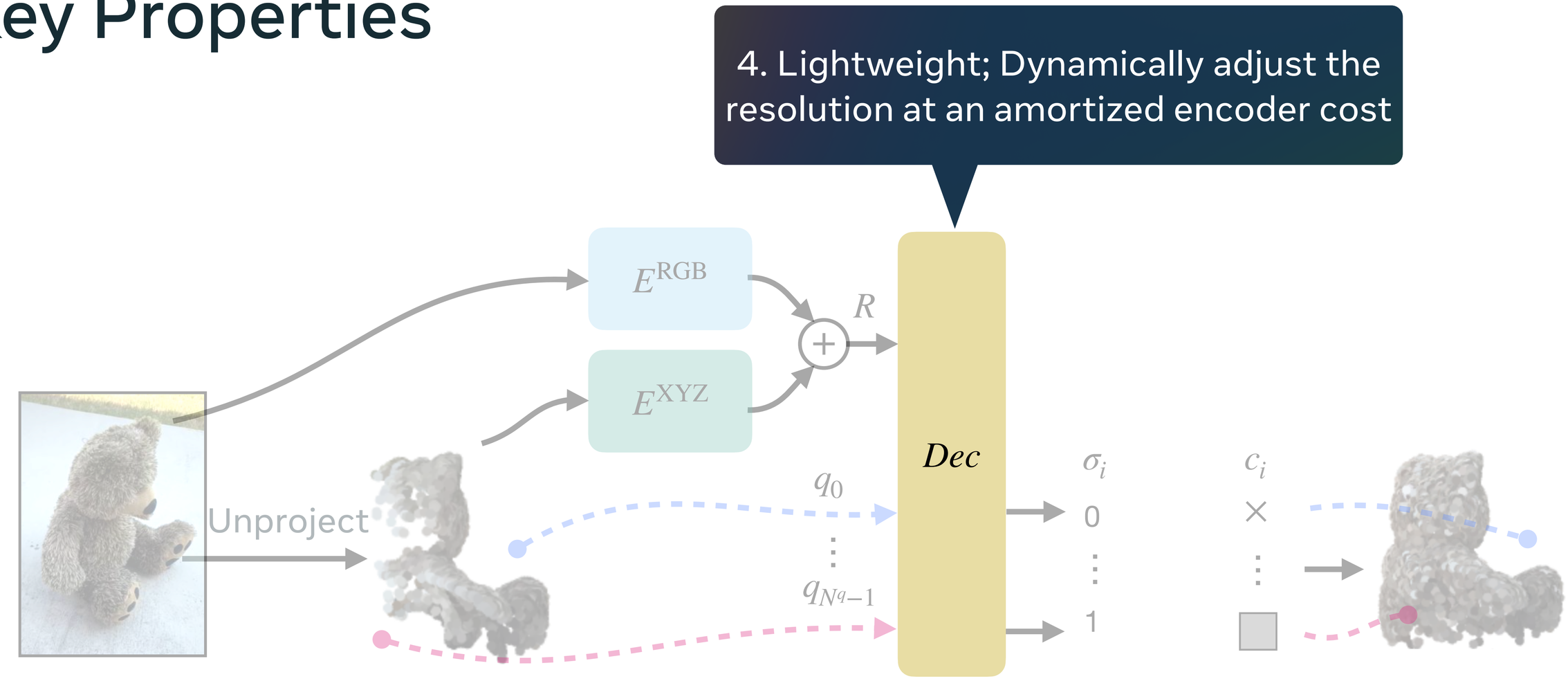


Key Properties

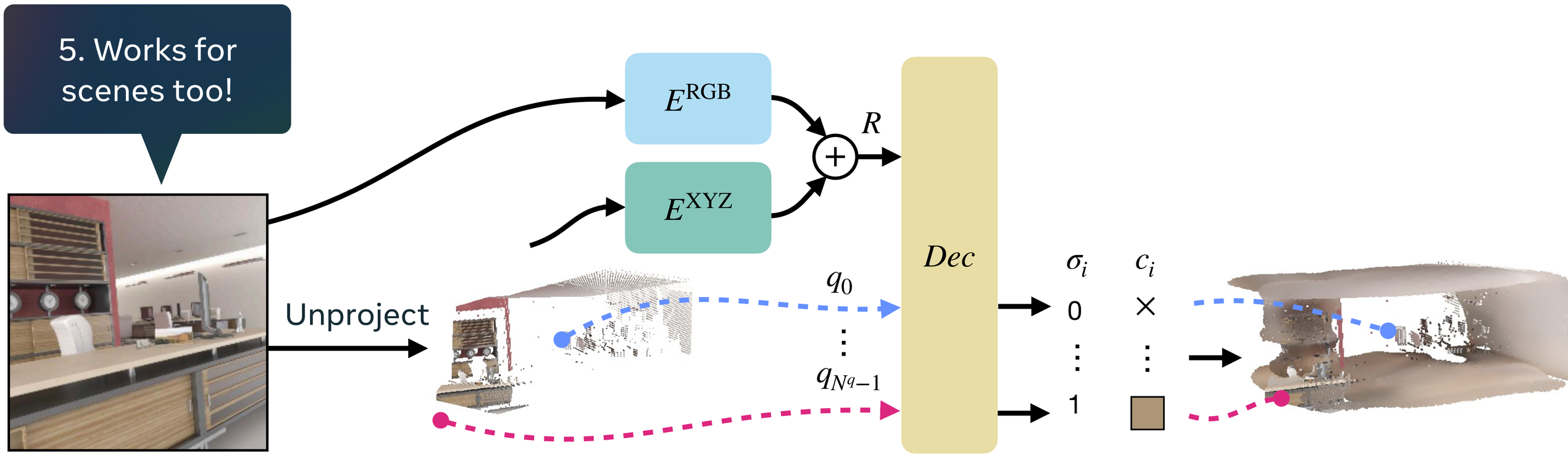
3. General Transformer-based design; No priors



Key Properties



Key Properties



Design philosophy

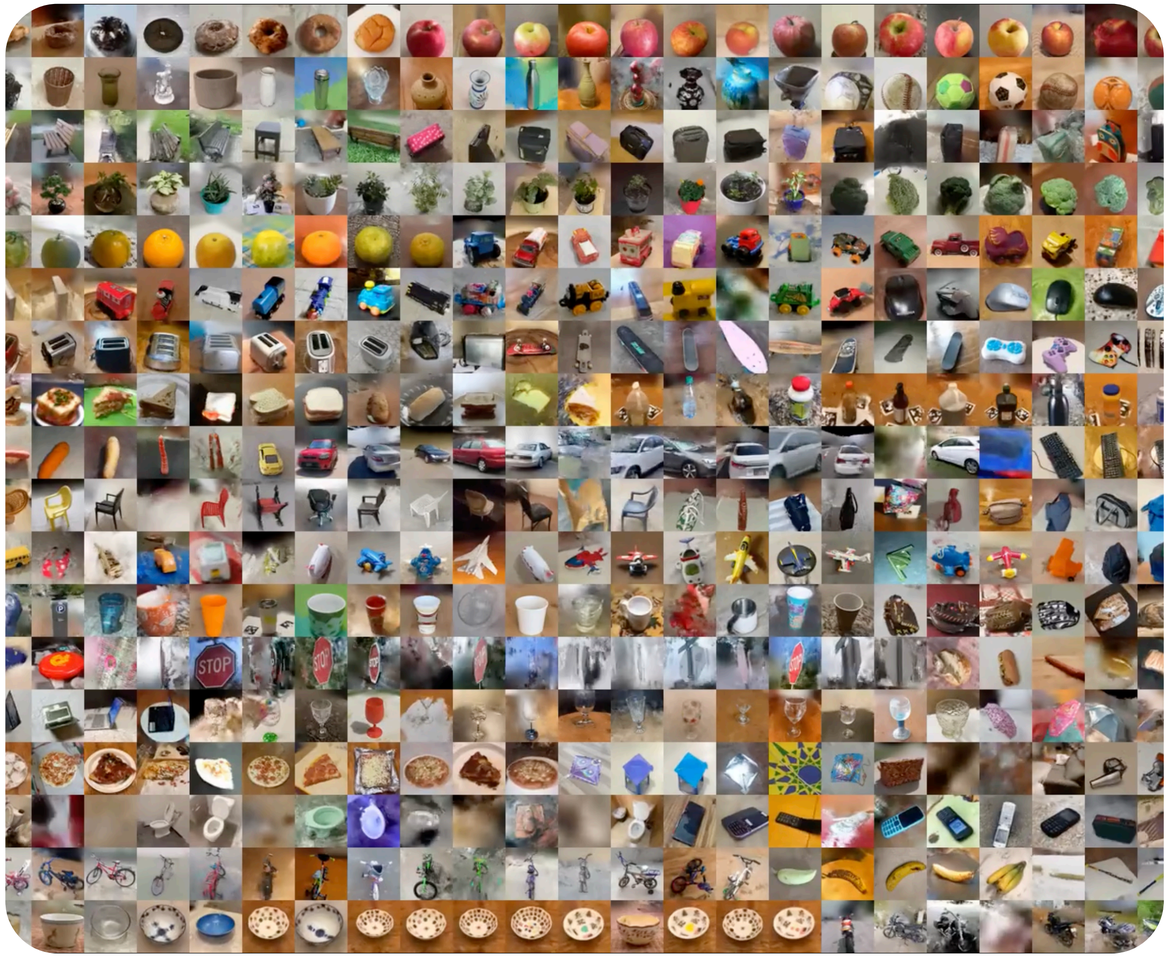
Simplicity

Scalability

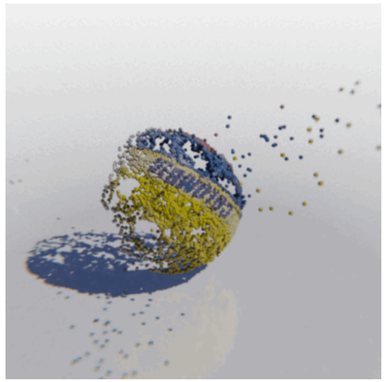
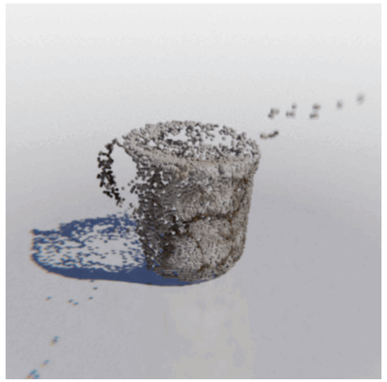
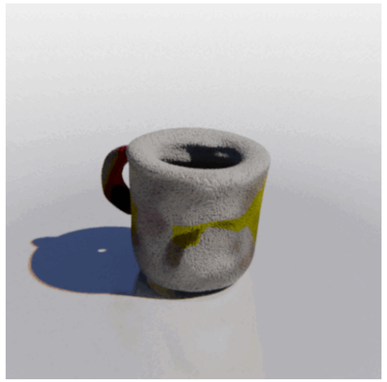
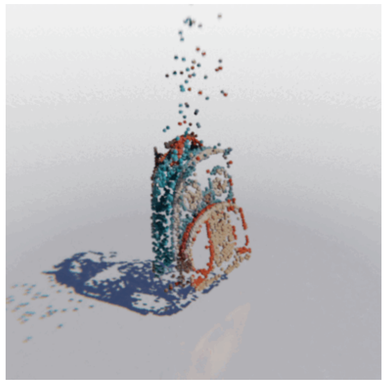
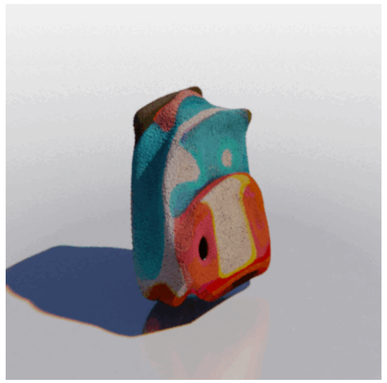
Experiments

Object-level data: **CO3D** dataset

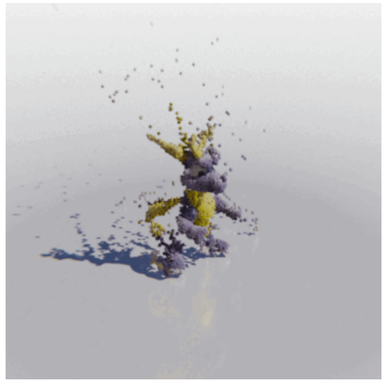
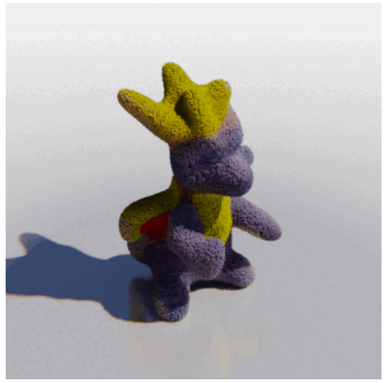
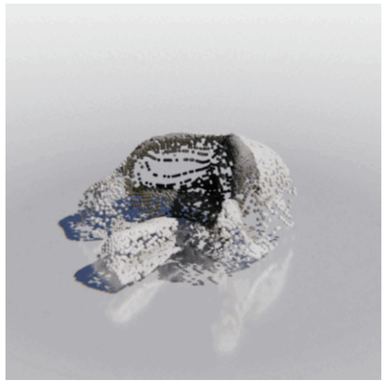
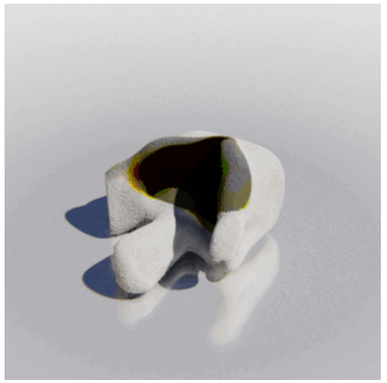
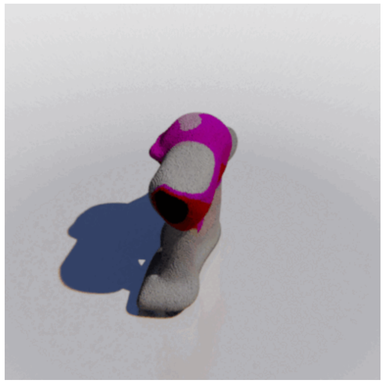
- 37K short videos
- 51 categories



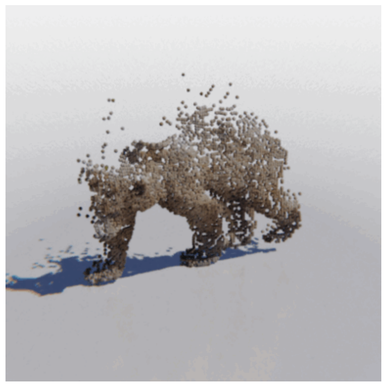
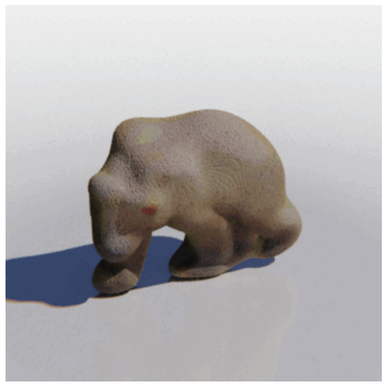
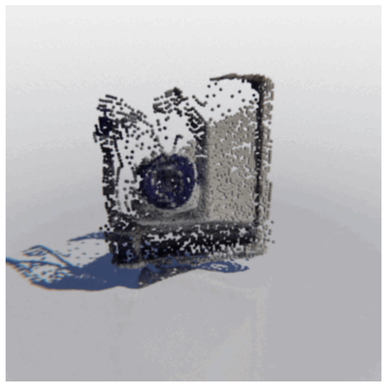
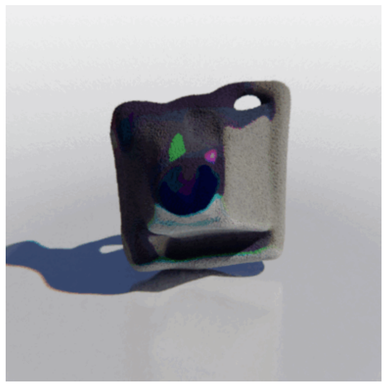
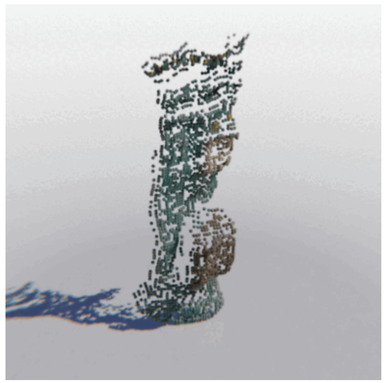
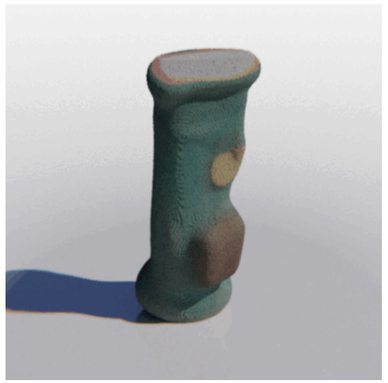
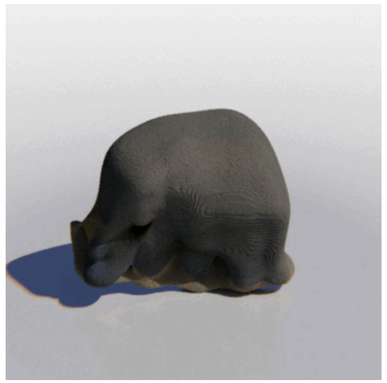
Generalization to **unseen CO3D categories**

Input	Seen	**output**	Input	Seen	**output**
					
					

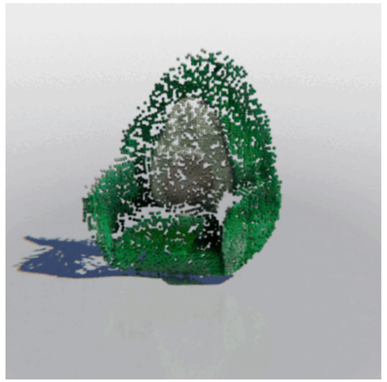
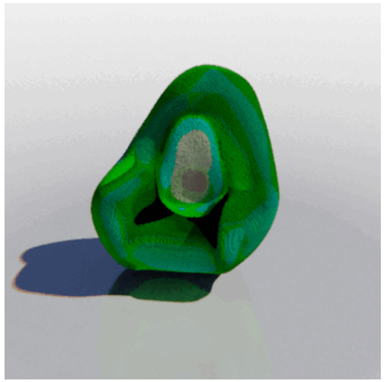
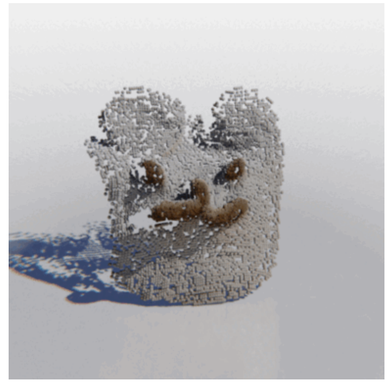
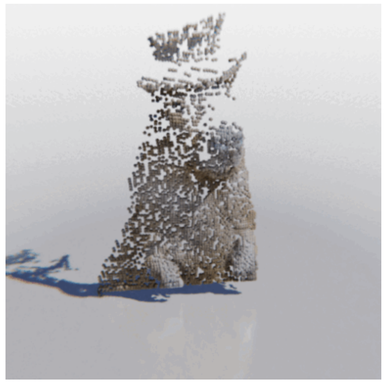
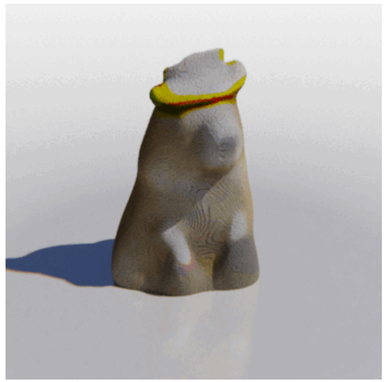
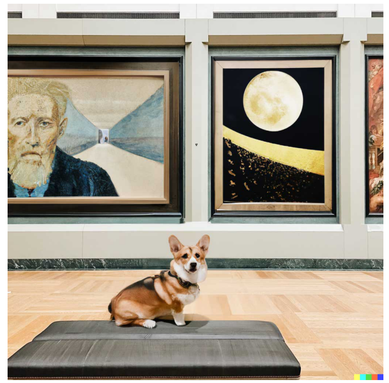
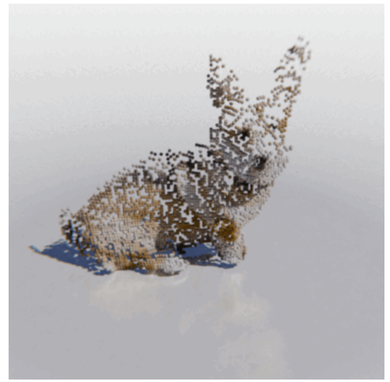
Generalization to **in-the-wild iPhone captures**

Input	Seen	**output**	Input	Seen	**output**
					
					

Generalization to **ImageNet** images

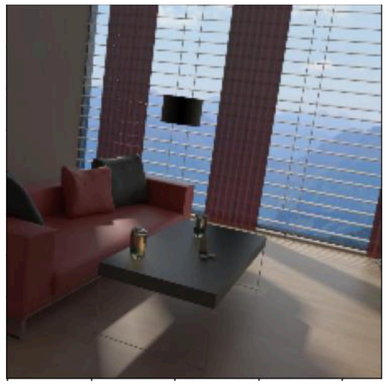
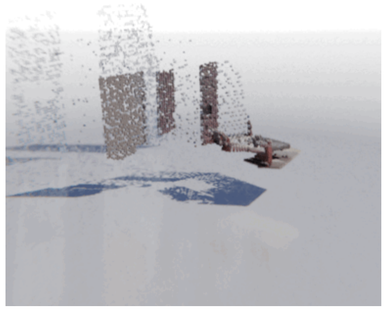
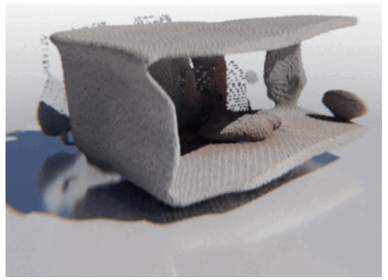
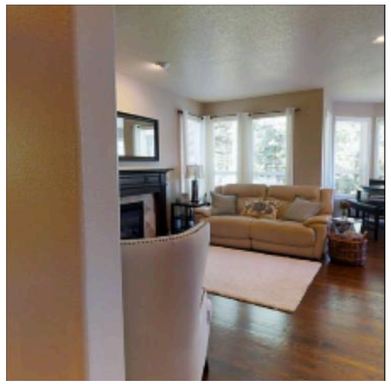
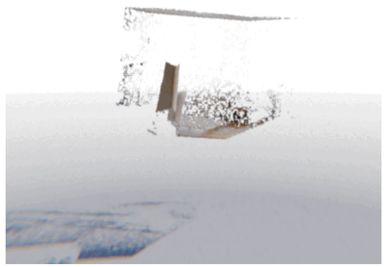
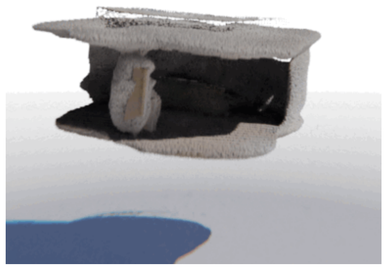
Input	Seen	**output**	Input	Seen	**output**
					
					

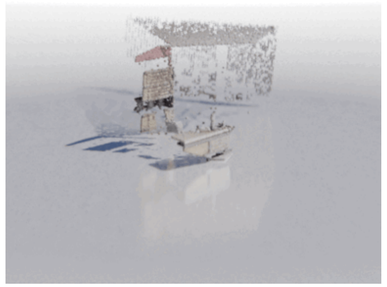
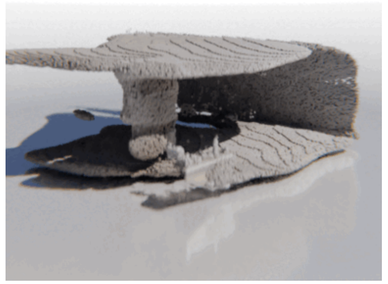
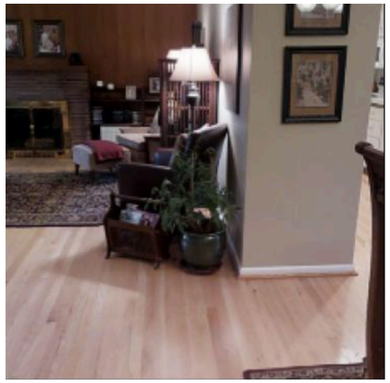
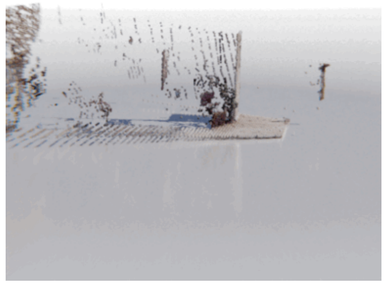
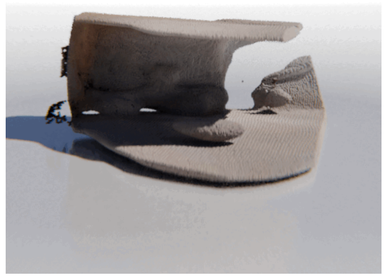
Generalization to DALL·E 2 images

Input	Seen	**output**	Input	Seen	**output**
					
					

Works on **Scenes**, too

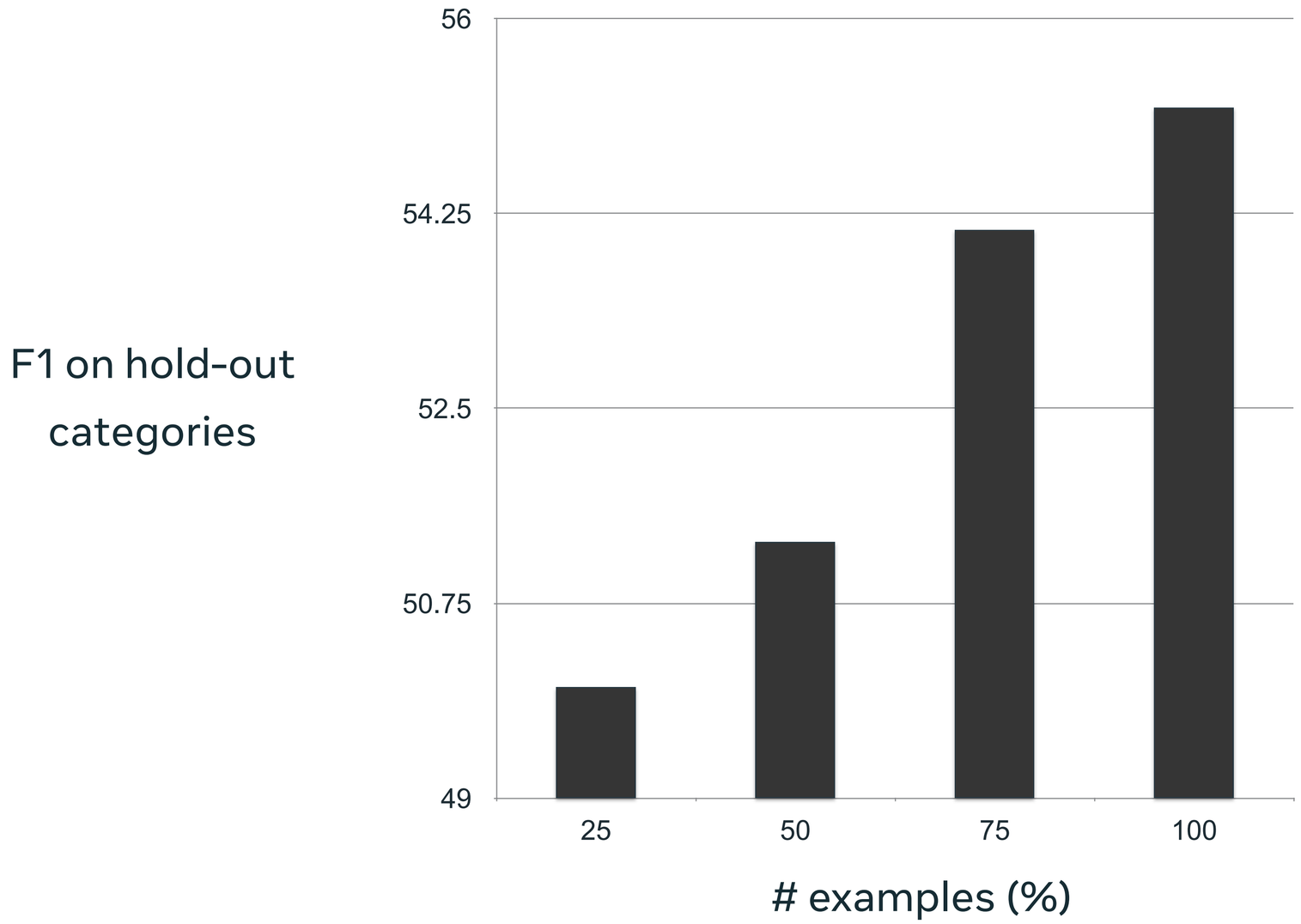
Input | Seen | **output** | Input | Seen | **output**

Model trained on **Hypersim**

Promising scaling behavior



Multiview Compressive Coding (MCC) for 3D Reconstruction

Code & models: https://github.com/facebookresearch/MCC
Project page: https://mcc3d.github.io/