Learning Unsupervised Hierarchical Part Decomposition of 3D Objects from a Single RGB Image

Despoina Paschalidou[1,2,5] Luc van Gool[2,3,5] Andreas Geiger[1,4,5]

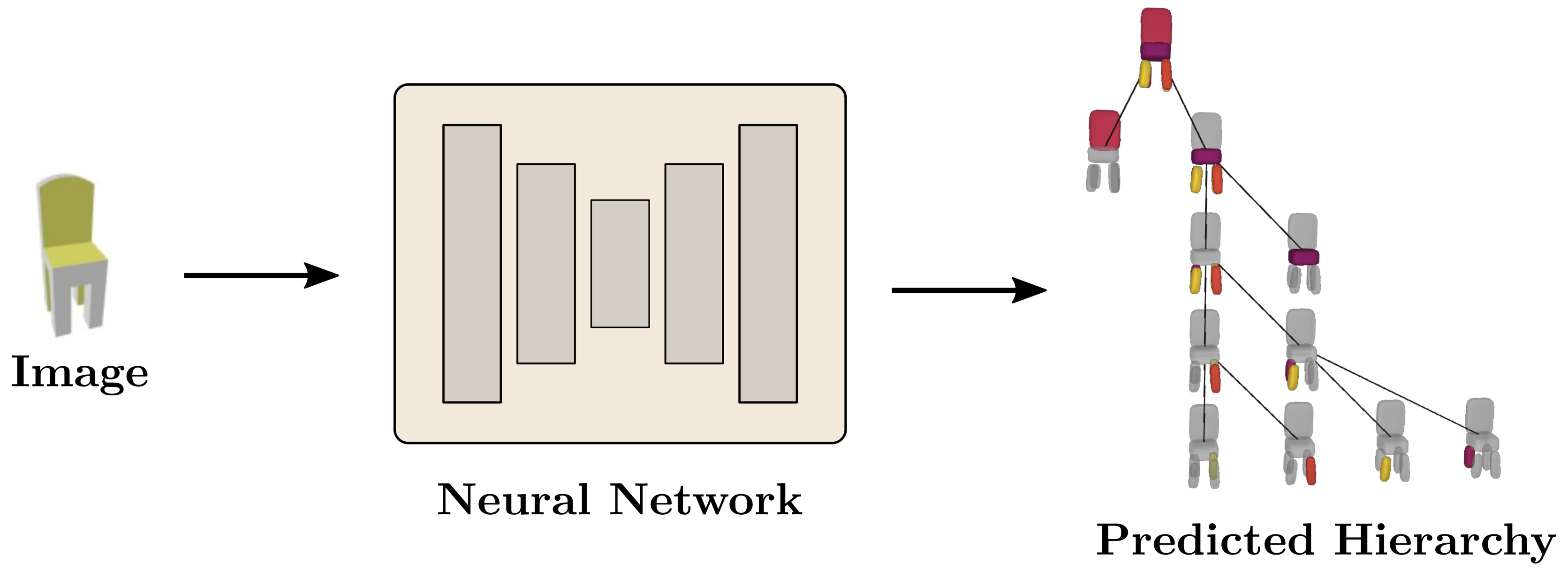
[1]Max Planck Institute for Intelligent Systems Tübingen
[2]Computer Vision Lab, ETH Zürich [3]KU Leuven [4]University of Tübingen
[5]Max Planck ETH Center for Learning Systems

CVPR SEATTLE WASHINGTON JUNE 16-18 2020

http://superquadrics.com/hierarchical_primitives

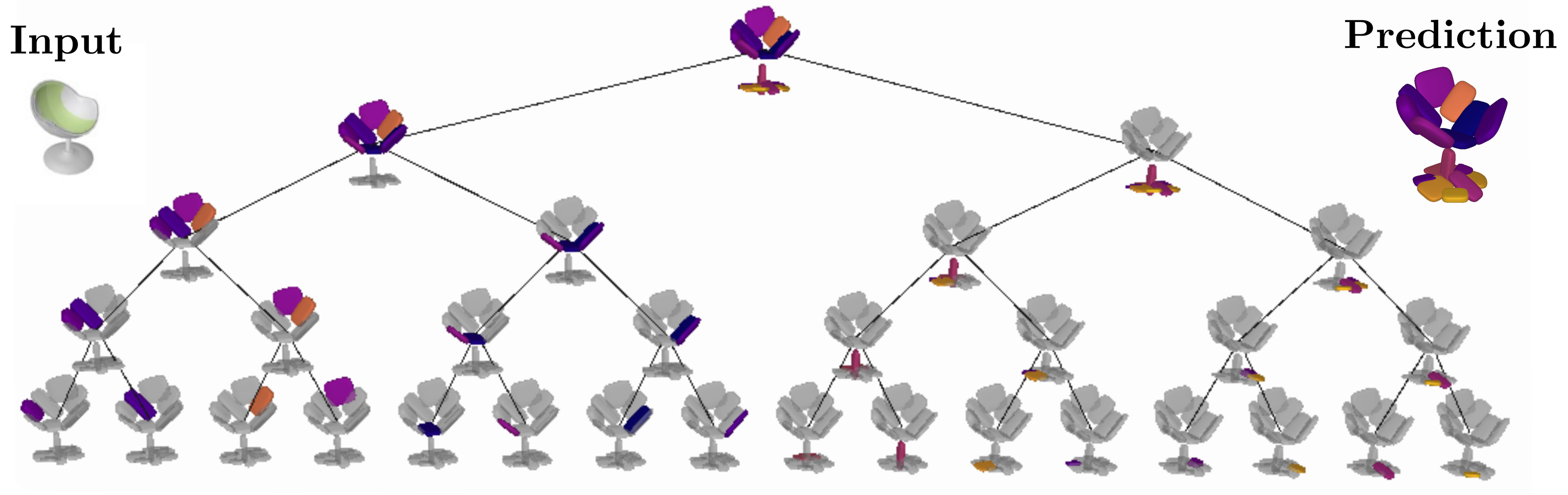
Motivation

- Existing shape representations **focus on recovering the local 3D geometry** of an object **without considering its part-based decomposition** or relations between its parts.
- **Goal: Jointly** recover **the geometry of an object as a set of primitives** and their **latent hierarchical layout without part-level supervision**.

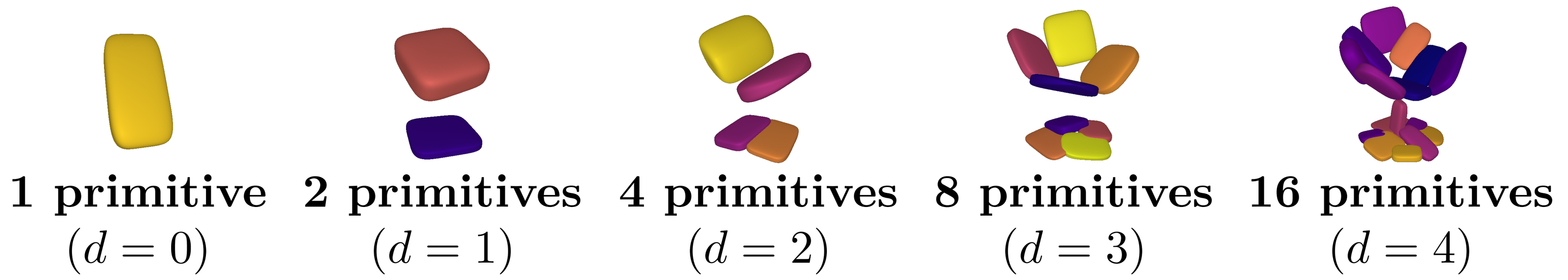


Our Representation

Idea: Model the **hierarchical decomposition** of various objects as **a binary tree of primitives** of depth D, where **simple parts are modelled with fewer primitives** and **more complex parts with more**.

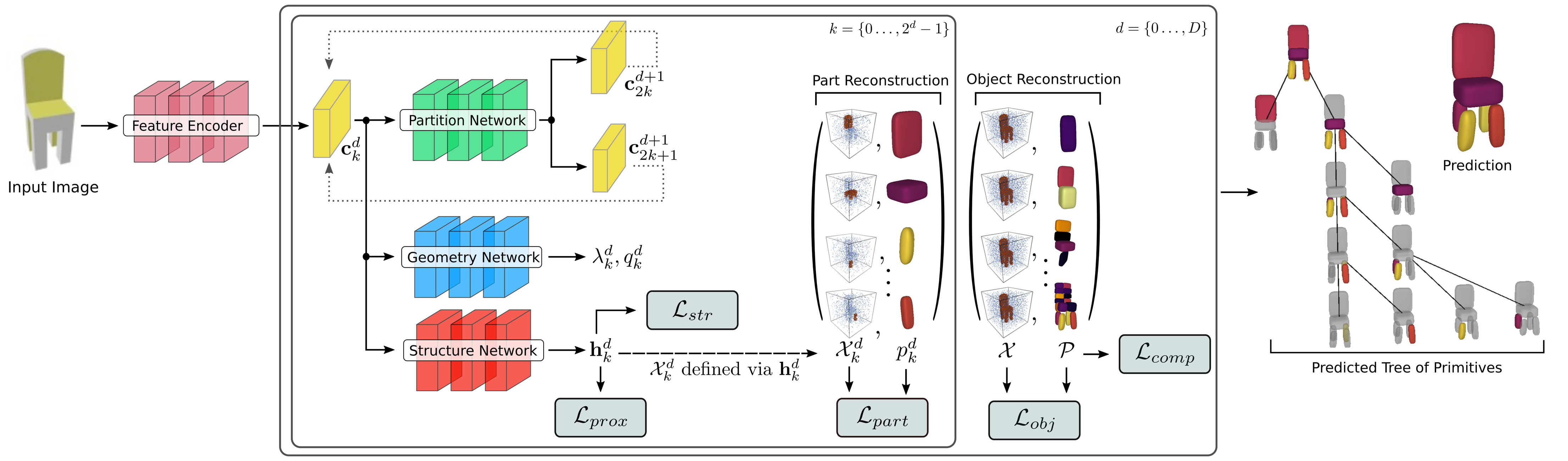


At every depth level, each of the $2^d \mid d = \{0, \dots, D\}$ **nodes is recursively split into two nodes** (its children) until reaching the maximum depth.



Network Architecture and Loss

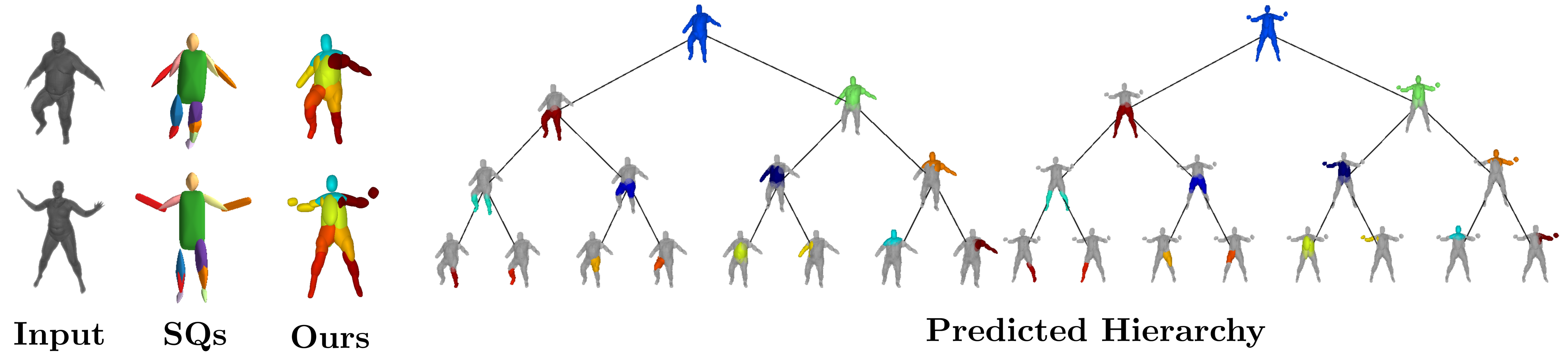
Given an input and a **target mesh** represented as a set of occupancy pairs $\mathcal{X} = \{(\mathbf{x}_i, o_i)\}_{i=1}^N$, our network predicts **a binary tree of primitives** $\mathcal{P} = \{\{p_k^d\}_{k=0}^{2^d-1} \mid d = \{0 \dots D\}\}$. For each primitive p_k^d the network regresses its **parameters** λ_k^d and its **reconstruction quality** q_k^d.



The **predicted shape** at depth d is $G^d(\mathbf{x})$ is the union of the per-primitive occupancy functions $g_k^d(\mathbf{x}, \lambda_k^d)$ at depth d.

- **Structure Loss:** $\mathcal{L}_{str}(\mathcal{H}; \mathcal{X}) = \sum_{h_k^d \in \mathcal{H}} \frac{1}{2^d - 1} \sum_{(\mathbf{x},o) \in \mathcal{X}_k^d} o \, \|\mathbf{x} - \mathbf{h}_k^d\|_2$
- **Reconstruction Loss:** $\mathcal{L}_{rec}(\mathcal{P}; \mathcal{X}) = \sum_{(\mathbf{x},o) \in \mathcal{X}} \sum_{d=0}^{D} L\left(G^d(\mathbf{x}), o\right) + \sum_{d=0}^{D} \sum_{k=0}^{2^d-1} \sum_{(\mathbf{x},o) \in \mathcal{X}_k^d} L\left(g_k^d\left(\mathbf{x}; \lambda_k^d\right), o\right)$
- **Compatibilty Loss:** $\mathcal{L}_{comp}(\mathcal{P}; \mathcal{X}) = \sum_{d=0}^{D} \sum_{k=0}^{2^d-1} \left(q_k^d - \mathbf{IoU}(p_k^d, \mathcal{X}_k^d)\right)^2$
- **Proximity Loss:** $\mathcal{L}_{prox}(\mathcal{P}) = \sum_{d=0}^{D} \sum_{k=0}^{2^d-1} \|\mathbf{t}(\lambda_k^d) - \mathbf{h}_k^d\|_2$

Single Image 3D Reconstruction on D-FAUST



Single Image 3D Reconstruction on ShapeNet

- Our representation **outperforms primitive-based methods and the OccNet w.r.t. IoU** and is the **second best w.r.t. Chamfer-$L1$**.
- Associate every primitive with a unique color, thus **primitives illustrated with the same color correspond to the same object part**