Exploring Superquadrics for 3D Shape Parsing

Thanasis Zoumpekas

Department of Mathematics and Computer Science, University of Barcelona

Topological Machine Learning Seminar

February 21, 2023

UNIVERSITAT DE BARCELONA

Marie Skłodowska-Curie Actions

Outline

- Introduction
- Superquadrics
- 3D Reconstruction
- Study of D. Paschalidou et al.[1]
  - Learning-based Scene Parsing
  - Experimentation and Results
- Conclusion
- Potential Research

Introduction

- Humans:
  - raw visual input into compact parsimonious representations
  - complex scenes from renderings of simple shape primitives [1]
- Early days of computer vision:
  - Images
  - 3D polyhedral shapes
  - Generalized cylinders
  - Superquadrics [1]
- Growing interest in 3D shape analysis
- Challenge: efficient and accurate parsing of complex shapes
- 3D data parsing into compact representations

Geometric Primitives

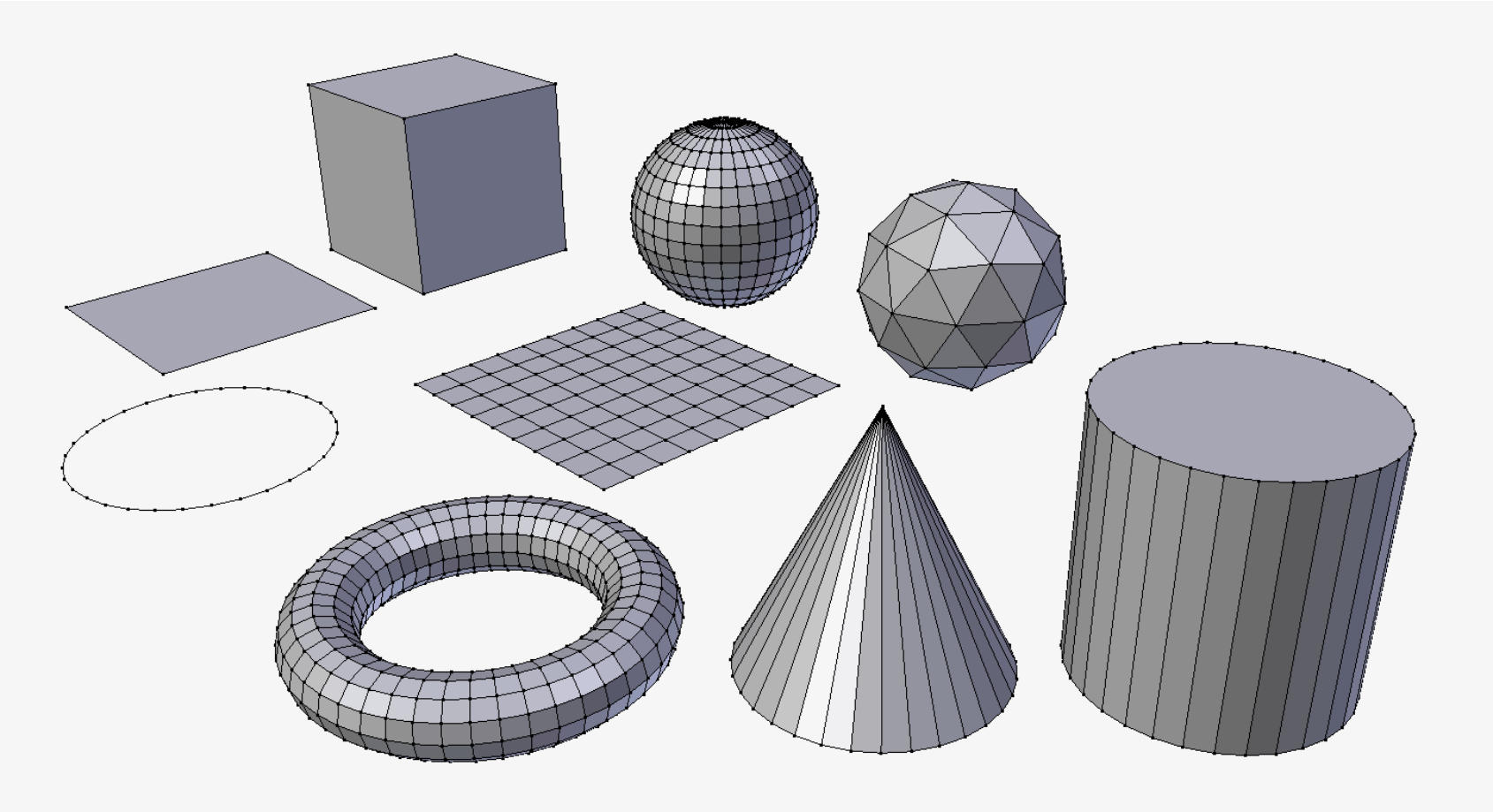


Image: https://www.openworldlearning.org/unlocking-the-potential-of-3d-modeling-with-primitives

Superquadrics

Barr [2] generalized superellipsoids to superquadrics (1981)

Parameters:

- Scale (control of overall size, e.g. larger/smaller)
- Shape (control of overall shape, e.g. rounded/angular)
- Orientation (specify orientation, e.g. Euler angles)
- Center point (coordinates of a point in 3D space)

Explicit equation [2]

$$\mathbf{r}(\eta,\omega) = \begin{bmatrix} \alpha_1 \cos^{\epsilon_1}\eta \cos^{\epsilon_2}\omega \\ \alpha_2 \cos^{\epsilon_1}\eta \sin^{\epsilon_2}\omega \\ \alpha_3 \sin^{\epsilon_1}\eta \end{bmatrix} \quad \begin{aligned} -\pi/2 \leq \eta \leq \pi/2 \\ -\pi \leq \omega \leq \pi \end{aligned}$$

Implicit equation

$$\left(\left(\frac{x}{\alpha_1} \right)^{\frac{2}{\epsilon_2}} + \left(\frac{y}{\alpha_2} \right)^{\frac{2}{\epsilon_2}} \right)^{\frac{\epsilon_2}{\epsilon_1}} + \left(\frac{z}{\alpha_3} \right)^{\frac{2}{\epsilon_1}} = 1$$

- $\alpha = [\alpha_1, \alpha_2, \alpha_3]$ control the size
- $\epsilon = [\epsilon_1, \epsilon_2]$ control of global shape

Rigid body transformation

- Translation vector: $\mathbf{t} = [t_x, t_y, t_z]$
- Rotation:
  - Euler angles: $[\phi, \theta, \psi]$
  - Quaternions: $[q_0, q_1, q_2, q_3]$
  - Quaternions help with gimbal lock problem, i.e. two of the three rotational axes happen to align
- Coordinate system transformation $\mathcal{T}(\mathbf{x}) = \mathbf{R}(\lambda)\,\mathbf{x} + \mathbf{t}(\lambda)$ from world coordinates to local primitive-centric coordinates (p_x, p_y, p_z)

Formally

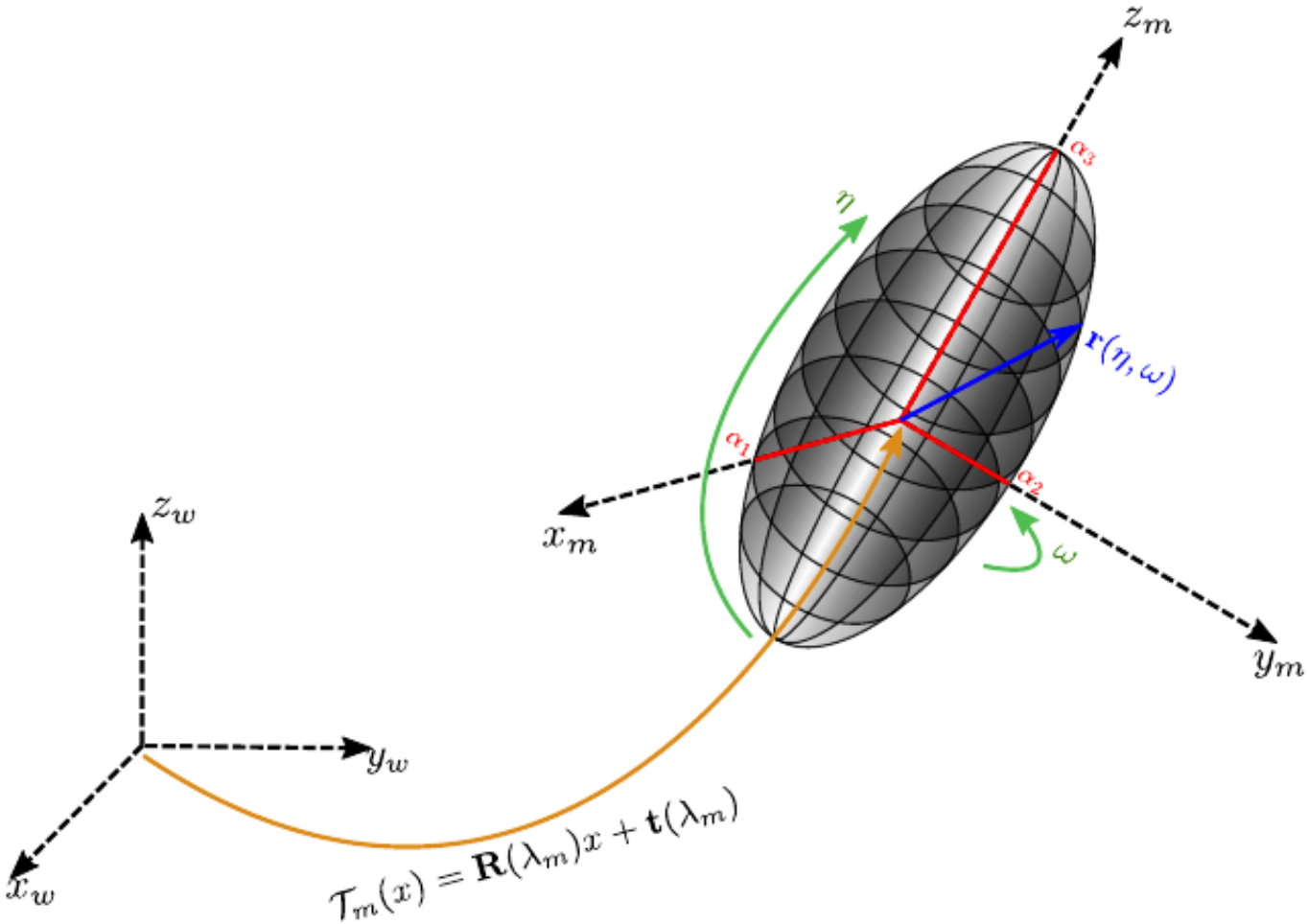
Inside-outside function in general position

$$F(x_w, y_w, z_w) = F(x_w, y_w, z_w; \alpha_1, \alpha_2, \alpha_3, \epsilon_1, \epsilon_2, \phi, \theta, \psi, p_x, p_y, p_z)$$

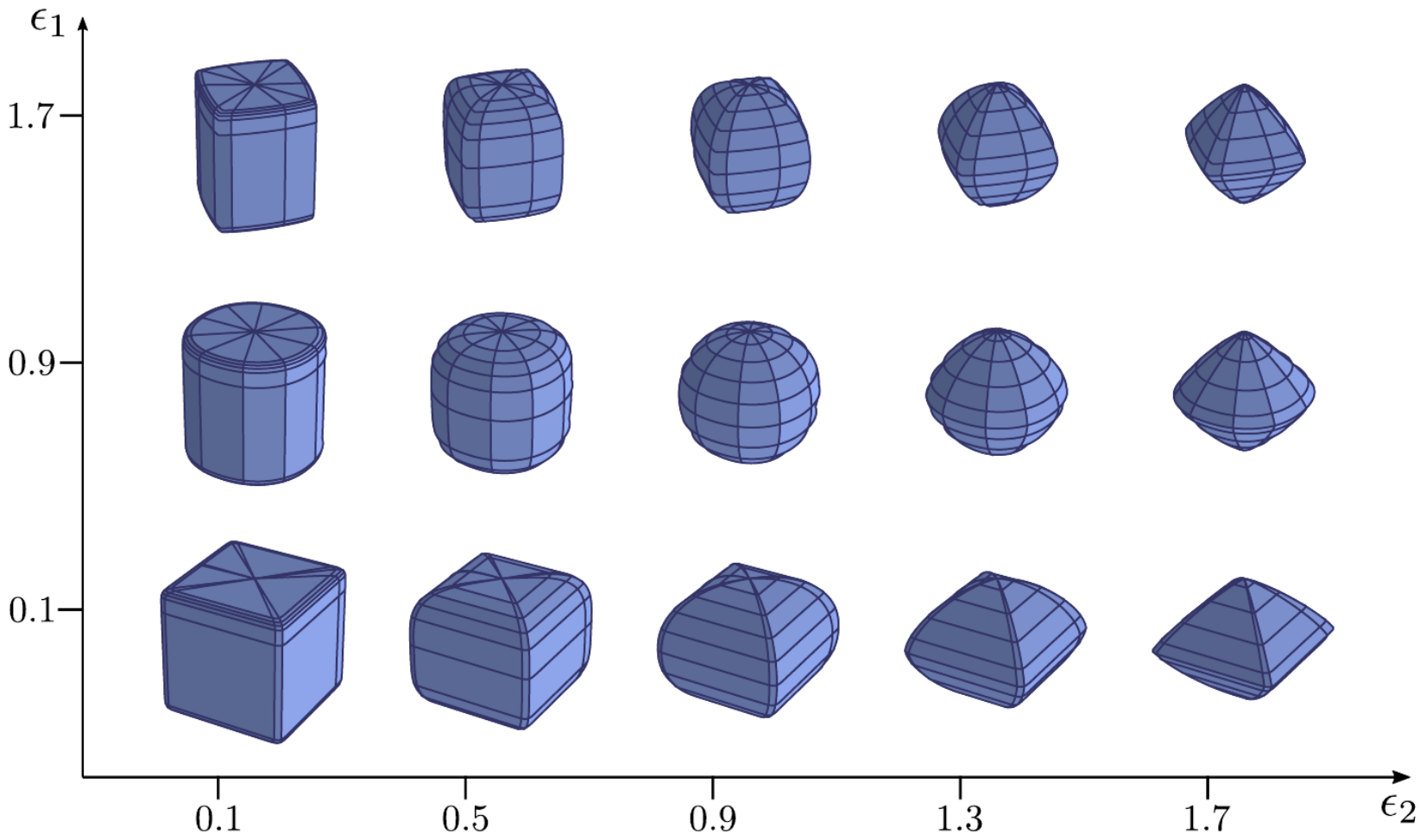
- $\alpha_1, \alpha_2, \alpha_3$: size
- ϵ_1, ϵ_2: shape
- ϕ, θ, ψ: orientation
- p_x, p_y, p_z the position/location in space

Simplification: $\Lambda = \lambda_1, \lambda_2, ..., \lambda_{11}$

Superquadrics



Superquadrics



Superquadrics

Pros

- Small number of parameters
- Wide variety of 3D shapes in a continuous parameter space
- Continuous parameterization $\longrightarrow$ deep learning
- Useful for representing complex geometrical shapes

Cons

- Not always accurate in representation of highly detailed shapes
- Choice of superquadric parameters plays a significant role
- Fitting to 3D data: computationally expensive

3D Reconstruction

- Simplest 3D reconstruction from images: 2.5D depth maps
- 2D Convolutions + post-processing to capture geometry
- Volumetric representations naturally capture 3D geometry
- 3D point sets or 3D meshes

Alternative goal:

- Object decomposition into a parsimonious representation [1]
- Semantic-wise part completion

Learning-based Scene Parsing

D. Paschalidou et al. [1]

Unsupervised learning of primitive-based 3D object representation

Input

- **I** (image, volume, point cloud)
- Oriented point cloud **X** of a target object

Goal

- Estimate the parameters θ of a neural network $\phi_\theta(\mathbf{I})$ that predicts a set of M primitives (primitive representation **P**)
- Every primitive: set of parameters λ_m (shape, size, location and orientation)

Formally: $\phi_\theta : \mathbf{I} \mapsto \mathbf{P}$

Total Loss

$$\mathcal{L}(\mathbf{P}, \mathbf{X}) = \mathcal{L}_D(\mathbf{P}, \mathbf{X}) + \mathcal{L}_\gamma(\mathbf{P})$$

Reconstruction Loss

$$\mathcal{L}_D(\mathbf{P}, \mathbf{X}) = \mathcal{L}_{P\to X}(\mathbf{P}, \mathbf{X}) + \mathcal{L}_{X\to P}(\mathbf{X}, \mathbf{P})$$

Parsimony Loss (Regularizer Loss)

$$\mathcal{L}_\gamma(\mathbf{P}) = \max\left(\alpha - \alpha \sum_{m=1}^{M} \gamma_m, 0\right) + \beta\sqrt{\sum_{m=1}^{M} \gamma_m}$$

Reconstruction Loss

$$\mathcal{L}_D(\mathbf{P}, \mathbf{X}) = \mathcal{L}_{P \to X}(\mathbf{P}, \mathbf{X}) + \mathcal{L}_{X \to P}(\mathbf{X}, \mathbf{P})$$

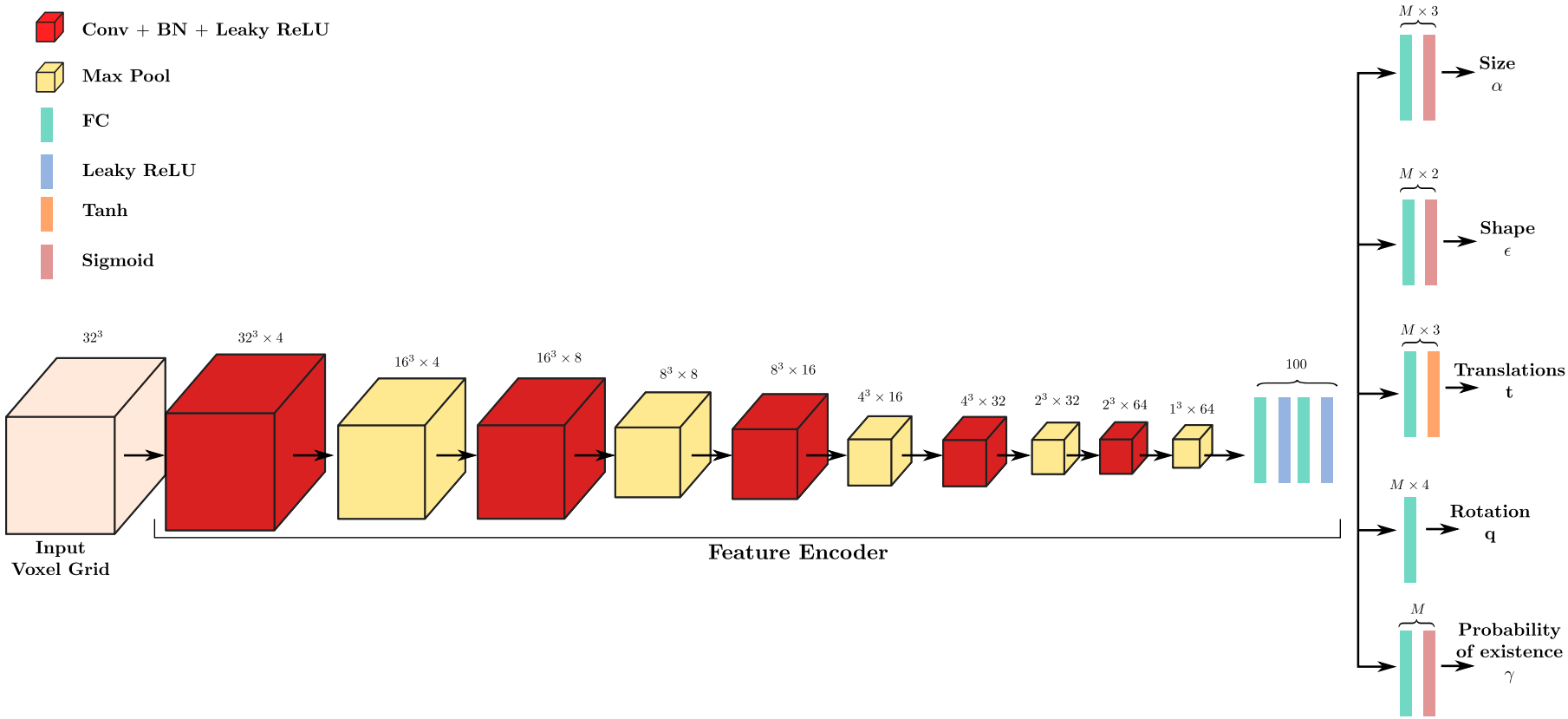
- $\mathcal{L}_{P \to X}$:
  - distance from the predicted primitives $\mathbf{P}$ to the target point cloud $\mathbf{X}$
  - enforces the $\mathbf{P}$ to stay close to $\mathbf{X}$
- $\mathcal{L}_{X \to P}$:
  - distance from the point cloud $\mathbf{X}$ to the primitives $\mathbf{P}$
  - ensures that each observation is explained by at least one primitive
- **Chamfer distance**

Parsimony Loss

$$\mathcal{L}_\gamma(\mathbf{P}) = \max\left(\alpha - \alpha \sum_{m=1}^{M} \gamma_m, 0\right) + \beta\sqrt{\sum_{m=1}^{M} \gamma_m}$$

- Trivial solution: $\mathcal{L}_D(\mathbf{P}, \mathbf{X}) = 0$ for $\gamma_1 = \cdots = \gamma_m = 0$
- Regularizer loss on existence probabilities γ
- Makes sure that at least one primitive is present
- Enforces a parsimonious scene parse
- Weighting factors α and β

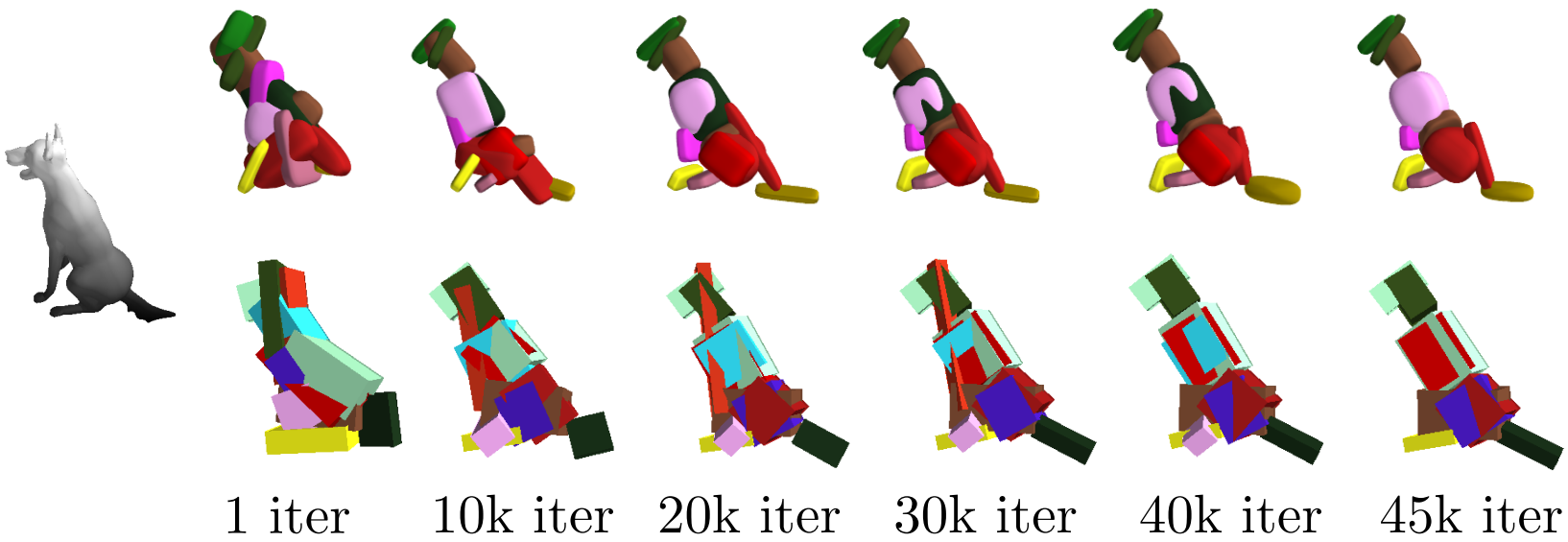
Deep Learning Architecture [1]



Experimentation & Results

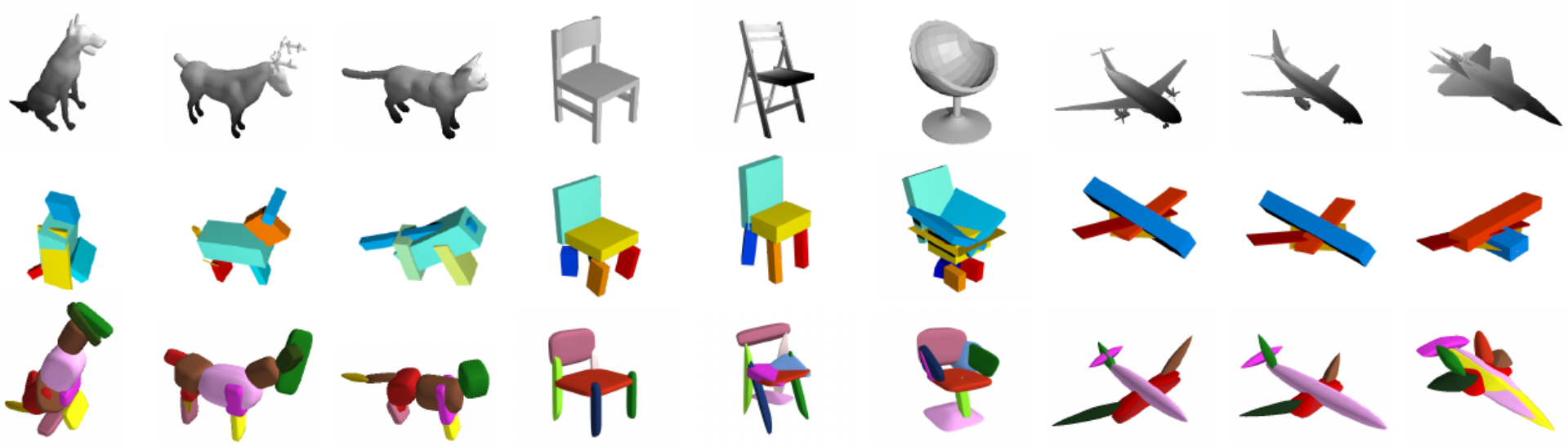
Datasets:

- ShapeNet: Aeroplane, Chair, and Animals
- SURREAL: Humans in various poses



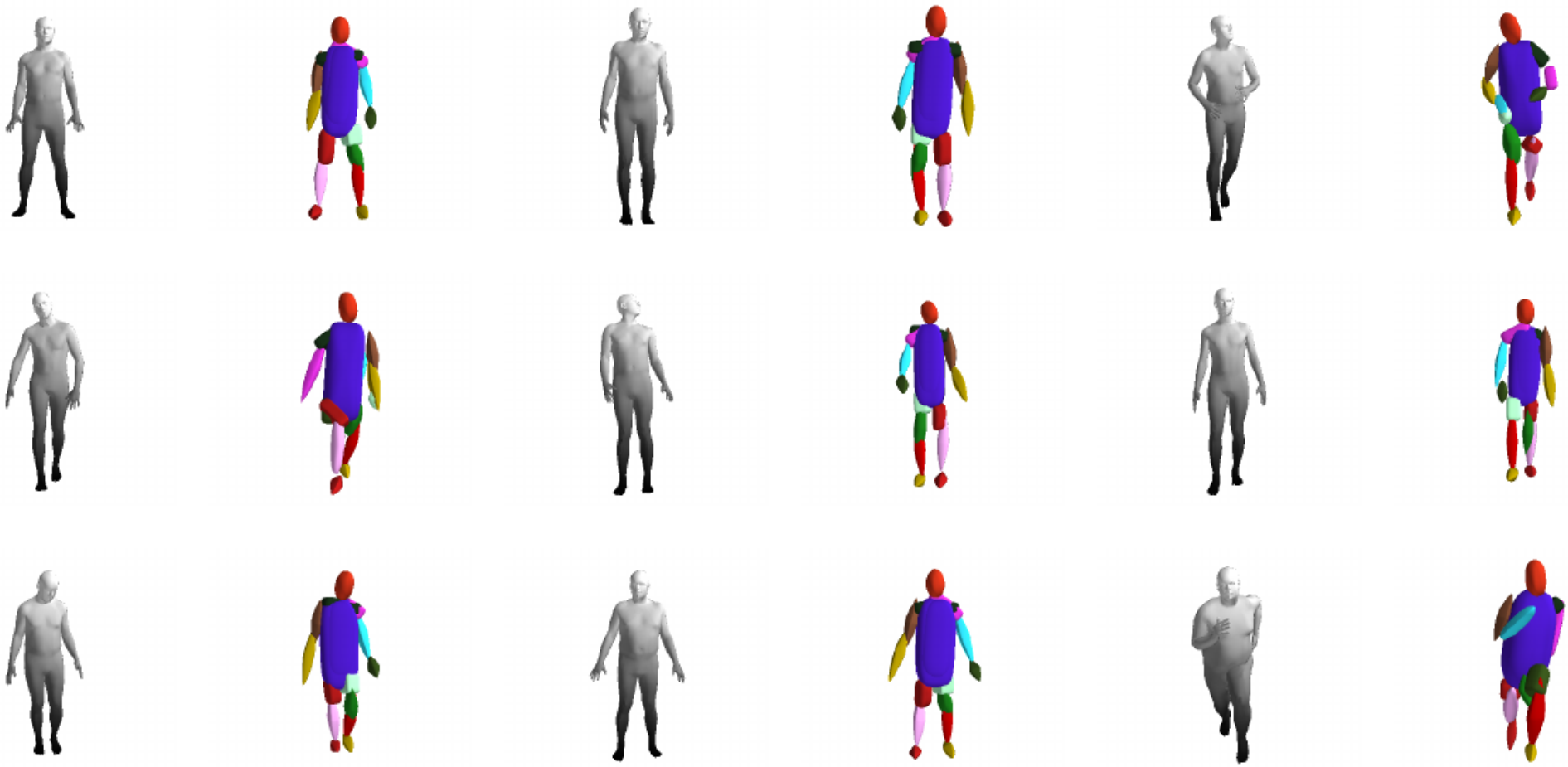
Evolution of Superquadrics (Top) and Cuboids (Bottom). Superquadrics converge faster to more accurate representations.

ShapeNet Results



Top row: Ground-truth mesh. Middle row: Cuboid-based representation. Bottom row: Superquadric-based representation.

SURREAL Results



Conclusion

Shape models in computer vision are chosen based on:

- Degree of uniqueness and compact representation
- Local support
- Expressiveness
- Preservation of information

Superquadrics:

- An extension of quadric surfaces that can model a variety of generic shapes
- Useful for volumetric part representation of natural and manmade objects

Conclusion

D. Paschalidou et al.[1]

- Unsupervised learning to predict superquadric-based representations of complex 3D shapes
- Capture of structure and semantic details of 3D objects
- Input: Occupancy grid or Image
- Feature Encoder + Regressor + Primitive Parameters Layer
- Qualitative results on ShapeNet and SURREAL
- Paves the way for other complex part-based tasks

Potential Research Directions

- Extension into large-scale scenes
- Indoor and outdoor scene parsing
- 3D object recognition
- Pose estimation
- Surface reconstruction
- Shape manipulation

References I

The material in this presentation is mainly taken from the studies below:

[1] D. Paschalidou, A. O. Ulusoy, and A. Geiger, "Superquadrics revisited: Learning 3d shape parsing beyond cuboids," in *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[2] A. H. Barr, "Superquadrics and Angle-Preserving Transformations," 1981.

Additional online sources: `https://cse.buffalo.edu/~jryde/cse673/files/superquadrics.pdf`