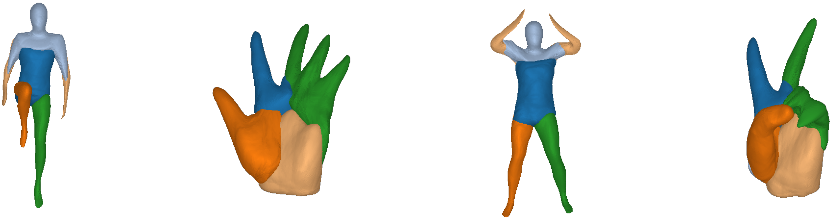
Neural Parts: Learning Expressive 3D Shape Abstractions with Invertible Neural Networks

Despoina Paschalidou

Autonomous Vision Group
Max Planck Institute for Intelligent Systems Tübingen
ETH Zürich

https://paschalidoud.github.io/neural_parts



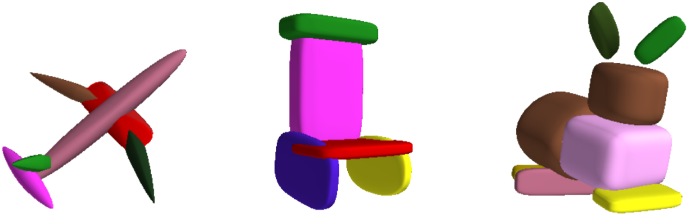
Joint work with

Angelos Katharopoulos

Andreas Geiger

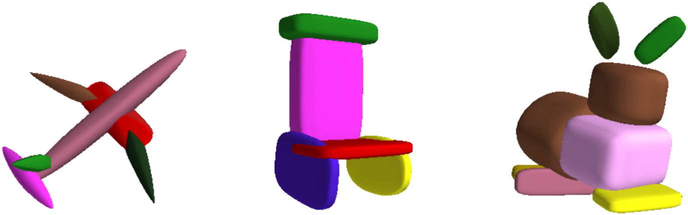
Sanja Fidler

3D Geometric Primitives: Why do we care?



Primitive-based Representations:

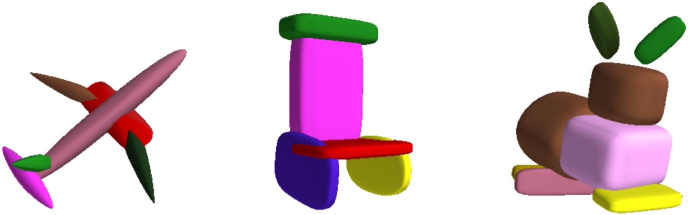
3D Geometric Primitives: Why do we care?



Primitive-based Representations:

- **Parsimonious Description**: Capture the 3D geometry using a small number of primitives.

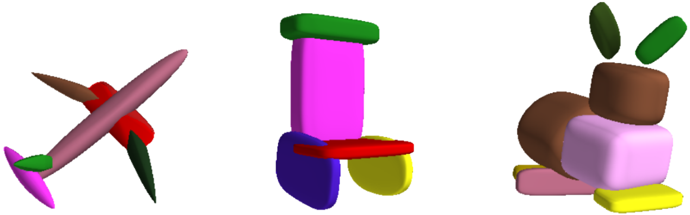
3D Geometric Primitives: Why do we care?



Primitive-based Representations:

- **Parsimonious Description**: Capture the 3D geometry using a small number of primitives.
- Convey semantic information (parts, functionality, etc.)

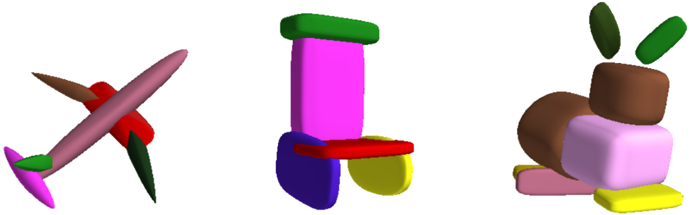
3D Geometric Primitives: Why do we care?



Primitive-based Representations:

- **Parsimonious Description**: Capture the 3D geometry using a small number of primitives.
- Convey semantic information (parts, functionality, etc.)
- **Main Challenges**:

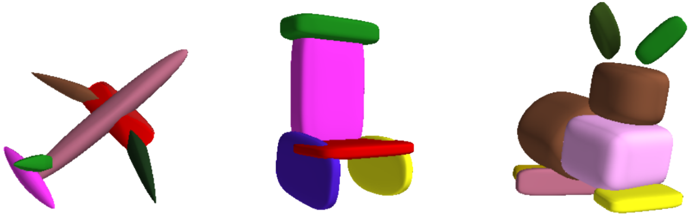
3D Geometric Primitives: Why do we care?



Primitive-based Representations:

- **Parsimonious Description**: Capture the 3D geometry using a small number of primitives.
- Convey semantic information (parts, functionality, etc.)
- **Main Challenges**:
  - Very few annotated datasets

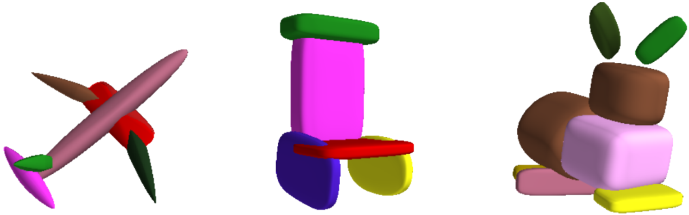
3D Geometric Primitives: Why do we care?



Primitive-based Representations:

- **Parsimonious Description**: Capture the 3D geometry using a small number of primitives.
- Convey semantic information (parts, functionality, etc.)
- **Main Challenges**:
  - Very few annotated datasets
  - Variable number of parts

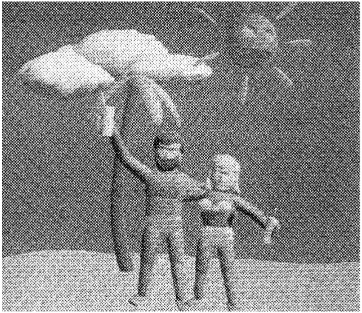
3D Geometric Primitives: Why do we care?

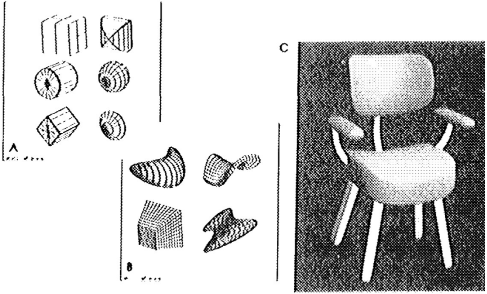


Primitive-based Representations:

- **Parsimonious Description**: Capture the 3D geometry using a small number of primitives.
- Convey semantic information (parts, functionality, etc.)
- **Main Challenges**:
  - Very few annotated datasets
  - Variable number of parts
  - What is really a semantic part?

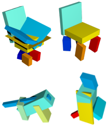
1986: Pentland’s Superquadrics





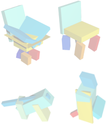
- 1 superquadric can be represented with 11 parameters
- Scene on the left **contructed with 100 primitives** required less than 1000 bytes!
- Early fitting-based approaches did not work robustly

Unsupervised Primitive-based Representations

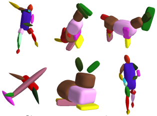


Cuboids
Tulsiani et al. 2017

Unsupervised Primitive-based Representations

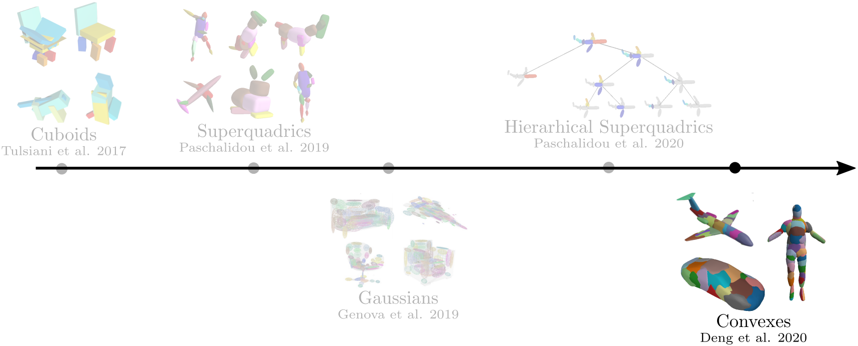


Cuboids
Tulsiani et al. 2017



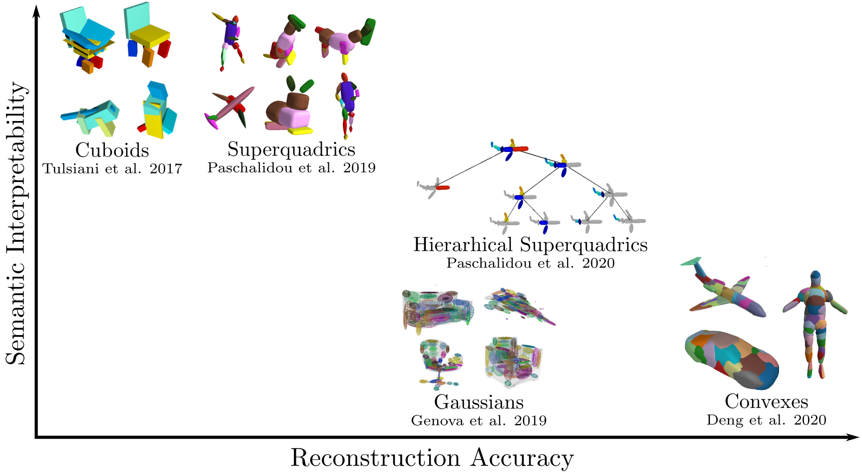
Superquadrics
Paschalidou et al. 2019

Unsupervised Primitive-based Representations



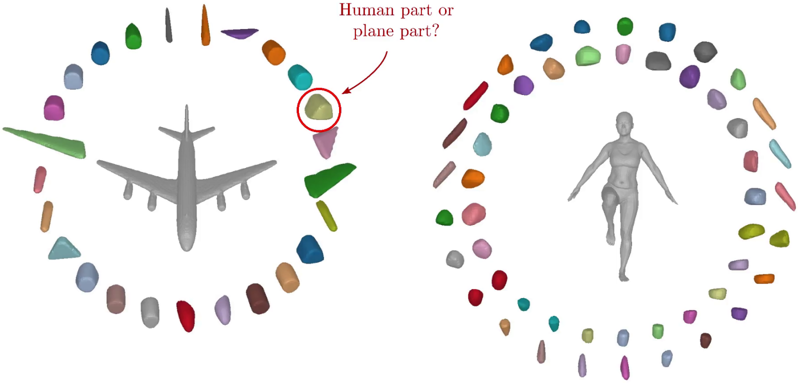
There exists a **trade-off** between the **number of primitives** and the **reconstruction quality** in primitive-based representations.

Primitive Arena

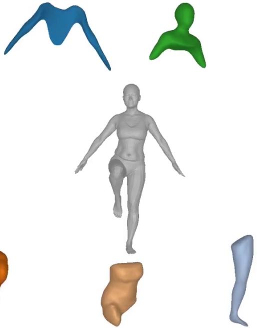


Simple parts require a large number of parts for accurate reconstructions.

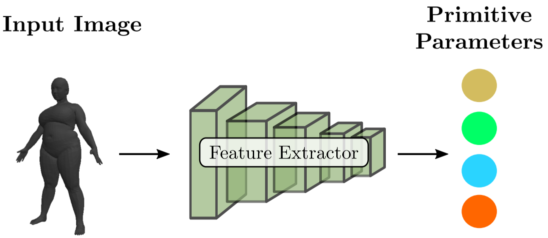
Simple parts require a large number of parts for accurate reconstructions.



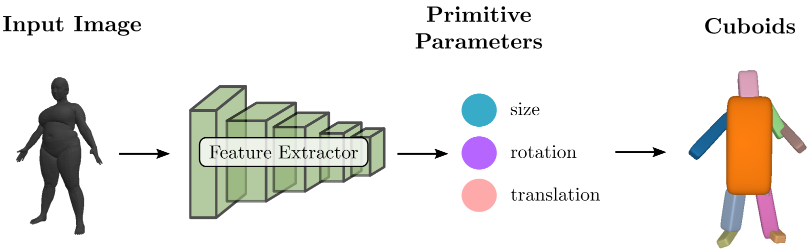
Neural Parts yield accurate and semantic reconstructions using an order of magnitude less parts.



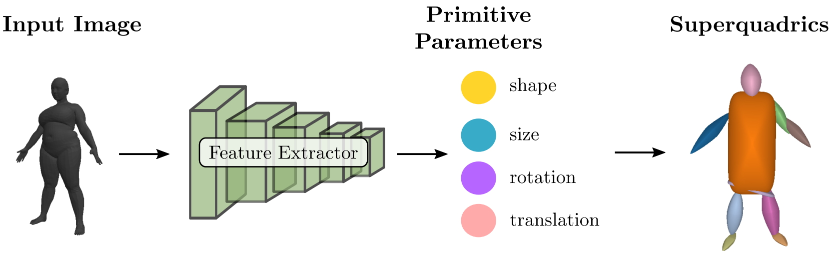
Primitive-based Learning



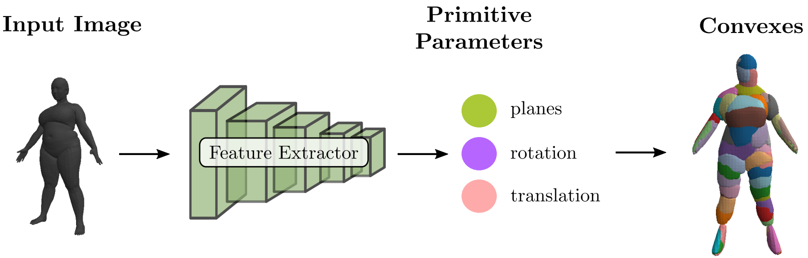
Primitive-based Learning



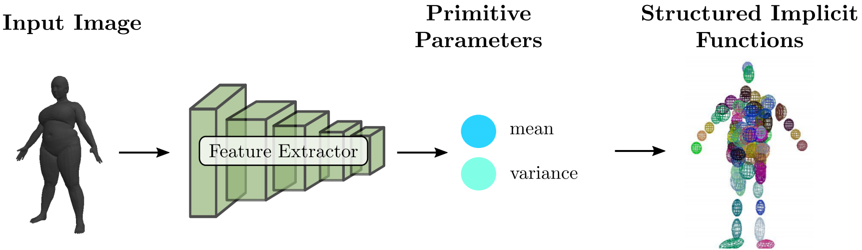
Primitive-based Learning



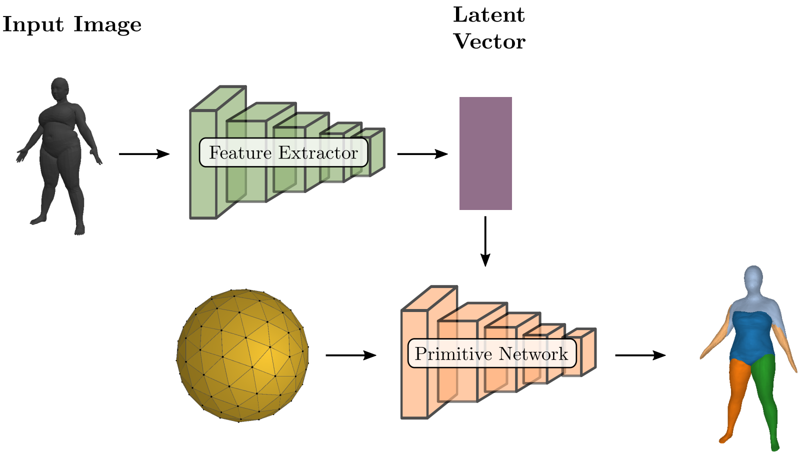
Primitive-based Learning



Primitive-based Learning



Primitive-based Learning



Homeomorphism

A **homeomorphism is a continuous map** between two topological spaces Y and X that preserves all topological properties. In our setup, a homeomorphism $\phi_{\boldsymbol{\theta}} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is

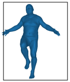
$$\mathbf{x} = \phi_{\boldsymbol{\theta}}(\mathbf{y}) \text{ and } \mathbf{y} = \phi_{\boldsymbol{\theta}}^{-1}(\mathbf{x})$$

where $\mathbf{x}$ and $\mathbf{y}$ are 3D points in X and Y and $\phi_{\boldsymbol{\theta}} : Y \rightarrow X$, $\phi_{\boldsymbol{\theta}}^{-1} : X \rightarrow Y$ are continuous bijections.

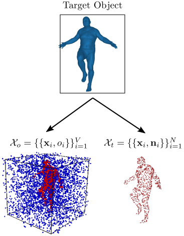
System Overview

Input Image

Target Object

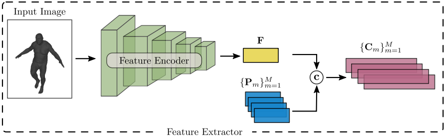


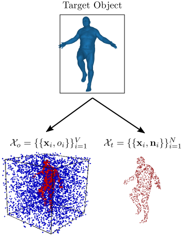
System Overview



- Our **supervision** comes from a watertight mesh of the target object parametrized as **surface samples** $\mathcal{X}_t$ and a set of **occupancy pairs** $\mathcal{X}_o$.

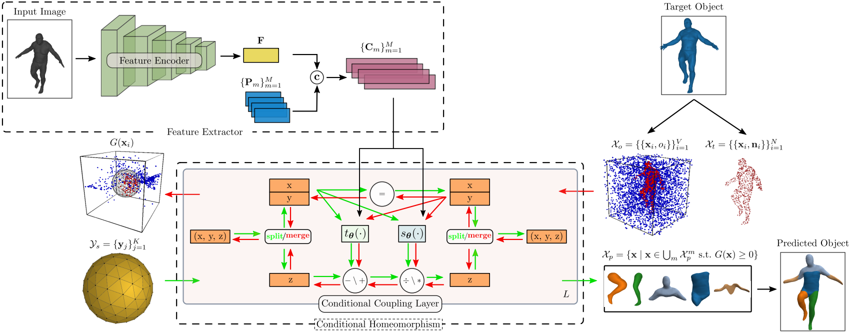
System Overview





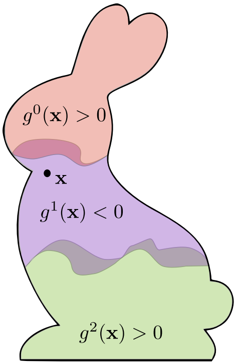
- Our **supervision** comes from a watertight mesh of the target object parametrized as **surface samples** $\mathcal{X}_t$ and a set of **occupancy pairs** $\mathcal{X}_o$.
- The **feature extractor** maps the input image into a **per-primitive shape embedding**.

System Overview



- Our **supervision** comes from a watertight mesh of the target object parametrized as **surface samples** $\mathcal{X}_t$ and a set of **occupancy pairs** $\mathcal{X}_o$.
- The **feature extractor** maps the input image into a **per-primitive shape embedding**.
- The **conditional homeomorphism** deforms a sphere into M primitives and vice-versa.

Implicit and Explicit Representation of Predicted Shape

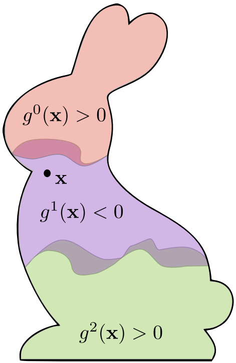


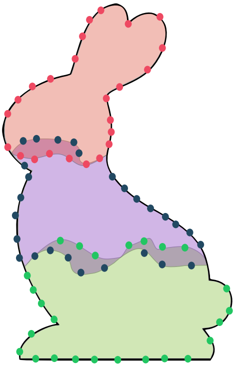
Implicit Representation:

$$G(\mathbf{x}) = \min_{m \in 0 \ldots M} g^m(\mathbf{x})$$

$$g^m(\mathbf{x}) = \|\phi_{\boldsymbol{\theta}}^{-1}(\mathbf{x}; \mathbf{C}_m)\|_2 - r, \ \forall \ \mathbf{x} \in \mathbb{R}^3$$

Implicit and Explicit Representation of Predicted Shape





Implicit Representation:

$$G(\mathbf{x}) = \min_{m \in 0 \ldots M} g^m(\mathbf{x})$$

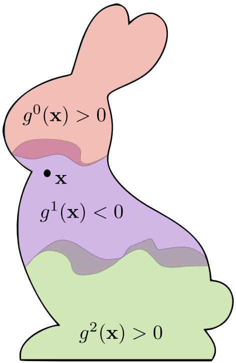
$$g^m(\mathbf{x}) = \|\phi_{\boldsymbol{\theta}}^{-1}(\mathbf{x}; \mathbf{C}_m)\|_2 - r, \ \forall \ \mathbf{x} \in \mathbb{R}^3$$

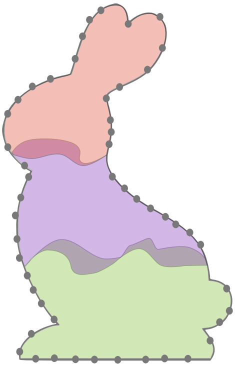
Explicit Representation:

$$\mathcal{X}_p = \{\mathbf{x} \mid \mathbf{x} \in \bigcup_m \mathcal{X}_p^m \text{ s.t. } G(\mathbf{x}) \geq 0\}$$

$$\mathcal{X}_p^m = \{\phi_{\boldsymbol{\theta}}(\mathbf{y}_j; \mathbf{C}_m), \ \forall \ \mathbf{y}_j \in \mathcal{Y}_s\}$$

Implicit and Explicit Representation of Predicted Shape





Implicit Representation:

$$G(\mathbf{x}) = \min_{m \in 0 \ldots M} g^m(\mathbf{x})$$

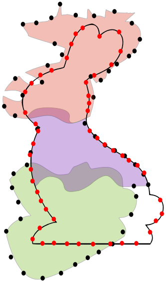
$$g^m(\mathbf{x}) = \|\phi_{\boldsymbol{\theta}}^{-1}(\mathbf{x}; \mathbf{C}_m)\|_2 - r, \ \forall \ \mathbf{x} \in \mathbb{R}^3$$

Explicit Representation:

$$\mathcal{X}_p = \{\mathbf{x} \mid \mathbf{x} \in \bigcup_m \mathcal{X}_p^m \text{ s.t. } G(\mathbf{x}) \geq 0\}$$

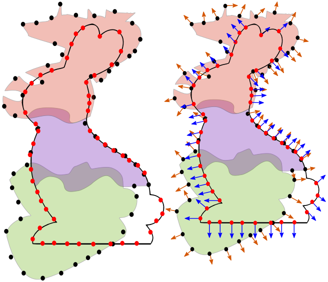
$$\mathcal{X}_p^m = \{\phi_{\boldsymbol{\theta}}(\mathbf{y}_j; \mathbf{C}_m), \ \forall \ \mathbf{y}_j \in \mathcal{Y}_s\}$$

Loss Functions



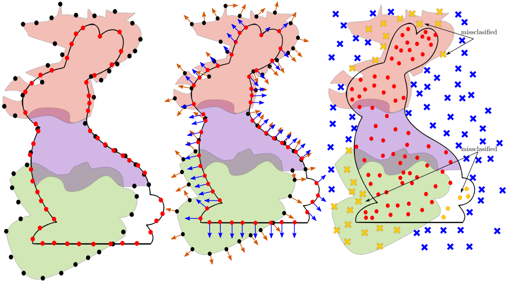
- **Reconstruction Loss**: The surface of the target and the predicted shape should match.

Loss Functions



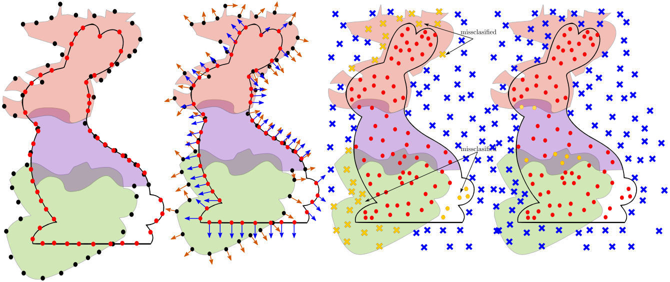
- **Reconstruction Loss**: The surface of the target and the predicted shape should match.
- **Normals Consistency Loss**: The normals of the target and the predicted shape should match.

Loss Functions



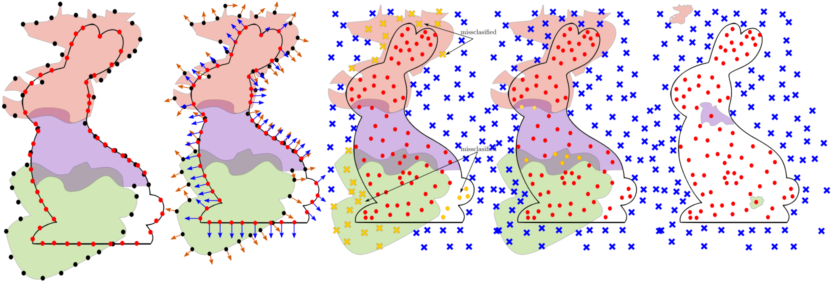
- **Reconstruction Loss**: The surface of the target and the predicted shape should match.
- **Normals Consistency Loss**: The normals of the target and the predicted shape should match.
- **Occupancy Loss**: The volume of the target and the predicted shape should match.

Loss Functions



- **Reconstruction Loss**: The surface of the target and the predicted shape should match.
- **Normals Consistency Loss**: The normals of the target and the predicted shape should match.
- **Occupancy Loss**: The volume of the target and the predicted shape should match.
- **Overlapping Loss**: Prevent overlapping primitives.

Loss Functions

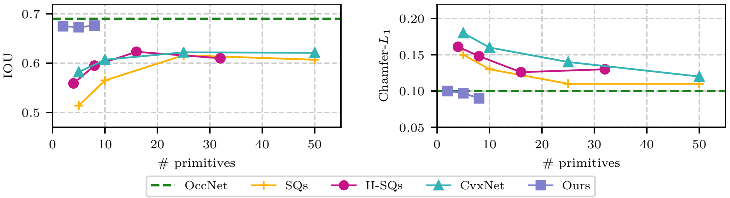


- **Reconstruction Loss**: The surface of the target and the predicted shape should match.
- **Normals Consistency Loss**: The normals of the target and the predicted shape should match.
- **Occupancy Loss**: The volume of the target and the predicted shape should match.
- **Overlapping Loss**: Prevent overlapping primitives.
- **Coverage Loss**: Prevent degenerate primitive arrangements.

How well does it work?

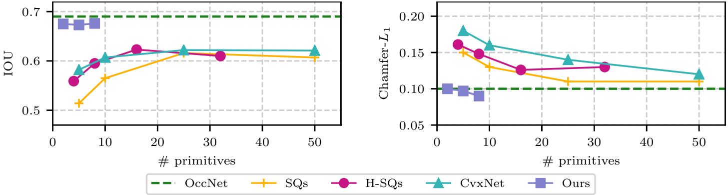
Representation Power of Primitive-based Representations

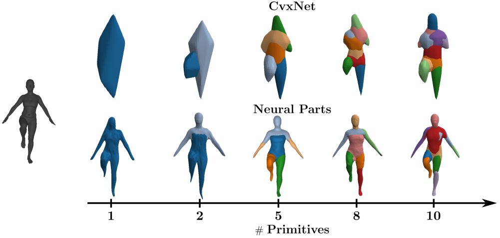
Neural Parts decouple the reconstruction quality from the number of parts.



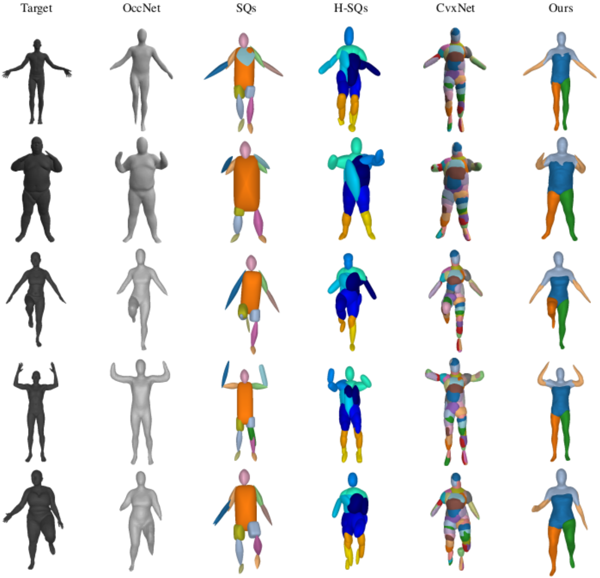
Representation Power of Primitive-based Representations

Neural Parts decouple the reconstruction quality from the number of parts.



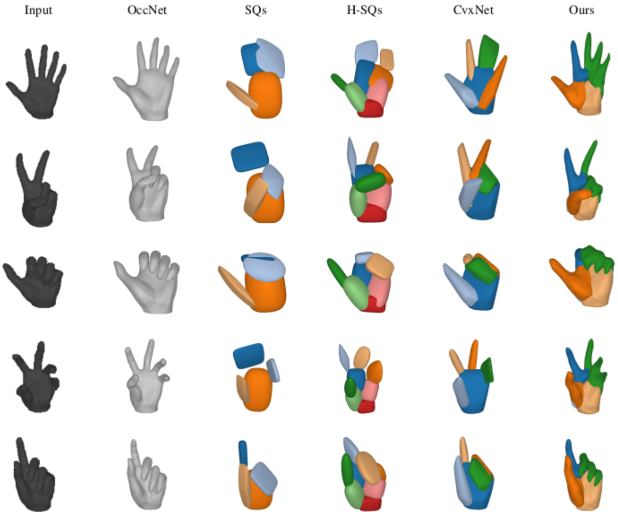


Single-view 3D Reconstruction on D-FAUST

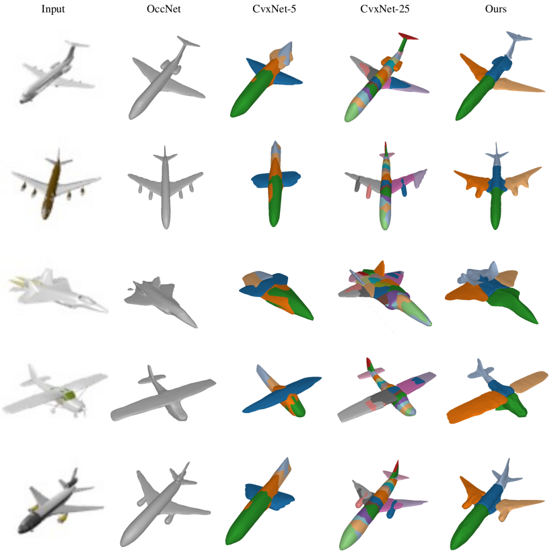


Semantic Consistency

Single-view 3D Reconstruction on FreiHAND

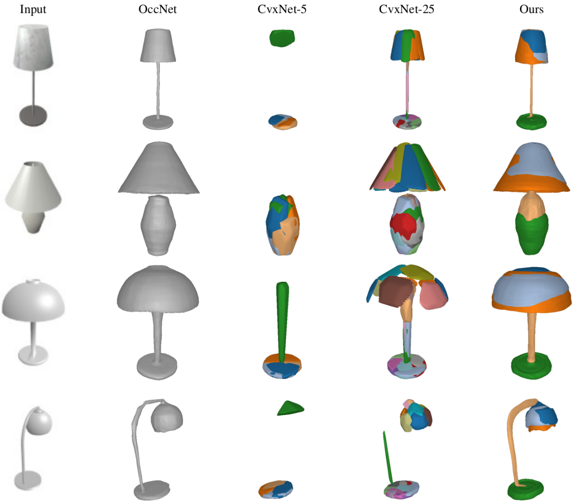


Single-view 3D Reconstruction on ShapeNet



Single-view 3D Reconstruction on ShapeNet

Single-view 3D Reconstruction on ShapeNet



Summary

- We propose Neural Parts, **a novel 3D primitive representation** as the homeomorphic mapping between a sphere and the target shape.

Summary

- We propose Neural Parts, **a novel 3D primitive representation** as the homeomorphic mapping between a sphere and the target shape.
- We demonstrate that **implementing homeomorphims with an INN is better than an MLP**.
- Neural Parts have well defined explicit and implicit formulations.

Summary

- We propose Neural Parts, **a novel 3D primitive representation** as the homeomorphic mapping between a sphere and the target shape.
- We demonstrate that **implementing homeomorphims with an INN is better than an MLP**.
- Neural Parts have well defined explicit and implicit formulations.
- Neural Parts **do not impose any constraint on the shape of the predicted primitive**.

Summary

- We propose Neural Parts, **a novel 3D primitive representation** as the homeomorphic mapping between a sphere and the target shape.
- We demonstrate that **implementing homeomorphims with an INN is better than an MLP**.
- Neural Parts have well defined explicit and implicit formulations.
- Neural Parts **do not impose any constraint on the shape of the predicted primitive**.
- Neural Parts allow to **decouple the reconstruction quality from the number of parts**, thus they yield both geometrically accurate and semantically meaningful shape abstractions.

Summary

- We propose Neural Parts, **a novel 3D primitive representation** as the homeomorphic mapping between a sphere and the target shape.
- We demonstrate that **implementing homeomorphims with an INN is better than an MLP**.
- Neural Parts have well defined explicit and implicit formulations.
- Neural Parts **do not impose any constraint on the shape of the predicted primitive**.
- Neural Parts allow to **decouple the reconstruction quality from the number of parts**, thus they yield both geometrically accurate and semantically meaningful shape abstractions.
- Limitations:

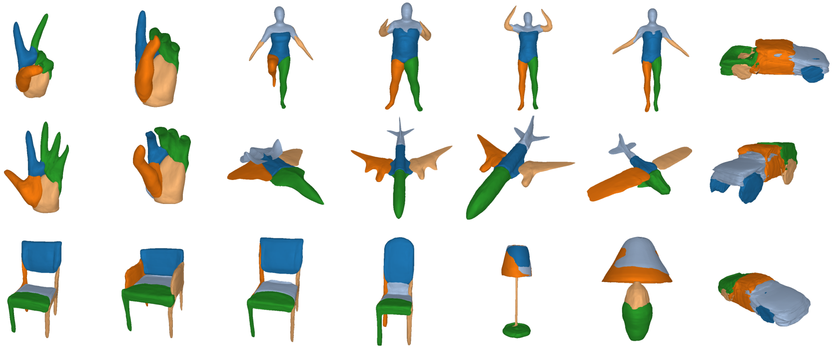
Summary

- We propose Neural Parts, **a novel 3D primitive representation** as the homeomorphic mapping between a sphere and the target shape.
- We demonstrate that **implementing homeomorphims with an INN is better than an MLP**.
- Neural Parts have well defined explicit and implicit formulations.
- Neural Parts **do not impose any constraint on the shape of the predicted primitive**.
- Neural Parts allow to **decouple the reconstruction quality from the number of parts**, thus they yield both geometrically accurate and semantically meaningful shape abstractions.
- Limitations:
  - High computational requirements due to the INN for the case of multiple primitives (e.g. for scenes).

Summary

- We propose Neural Parts, **a novel 3D primitive representation** as the homeomorphic mapping between a sphere and the target shape.
- We demonstrate that **implementing homeomorphims with an INN is better than an MLP**.
- Neural Parts have well defined explicit and implicit formulations.
- Neural Parts **do not impose any constraint on the shape of the predicted primitive**.
- Neural Parts allow to **decouple the reconstruction quality from the number of parts**, thus they yield both geometrically accurate and semantically meaningful shape abstractions.
- Limitations:
  - High computational requirements due to the INN for the case of multiple primitives (e.g. for scenes).
  - Similar to all primitive-based representations, the reconstructed parts are **spatially consistent without necessarily being semantic**.

Thank you for your attention!



https://paschalidoud.github.io/neural_parts