
MeshCoder: LLM-Powered Structured Mesh Code Generation from Point Clouds

Bingquan Dai^{*1,2}, Li Luo^{*1,3}, Qihong Tang^{1,4}, Jie Wang^{1,5}, Xinyu Lian¹, Hao Xu^{1,6}, Minghan Qin², Xudong Xu¹, Bo Dai³, Haoqian Wang^{†2}, Zhaoyang Lyu^{†1}, Jiangmiao Pang¹

¹Shanghai Artificial Intelligence Laboratory ²Tsinghua University ³The University of Hong Kong
⁴Harbin Institute of Technology ⁵Beijing Institute of Technology ⁶AI Thrust, HKUST(GZ)

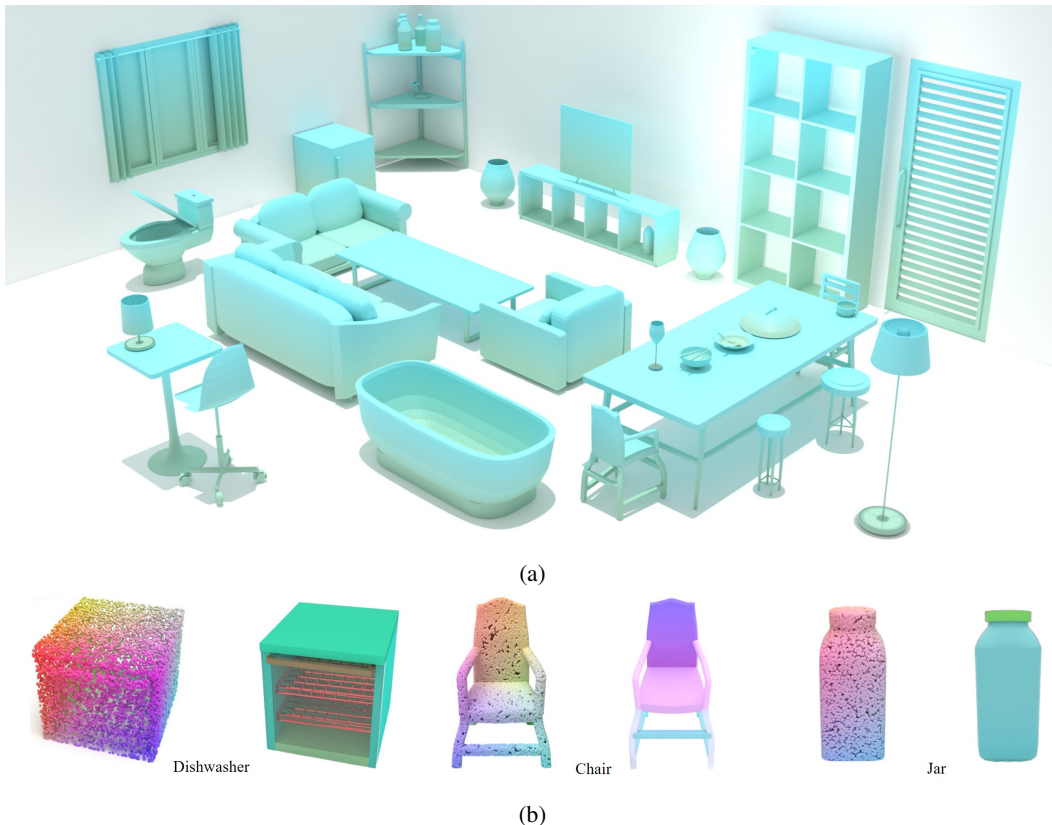


Figure 1: (a) MeshCoder can predict codes and reconstruct 41 categories of objects. (b) MeshCoder takes in point clouds and produce part-segmented meshes by executing the predicted code in Blender. For the dishwasher, we apply transparency to the foremost part to showcase the internal structure.

Abstract

Reconstructing 3D objects into editable programs is pivotal for applications like reverse engineering and shape editing. However, existing methods often rely on limited domain-specific languages (DSLs) and small-scale datasets, restricting their ability to model complex geometries and structures. To address these challenges, we introduce MeshCoder, a novel framework that reconstructs complex

*Equal contribution.

†Corresponding authors.

Project leader: Zhaoyang Lyu

3D objects from point clouds into editable Blender Python scripts. We develop a comprehensive set of expressive Blender Python APIs capable of synthesizing intricate geometries. Leveraging these APIs, we construct a large-scale paired object-code dataset, where the code for each object is decomposed into distinct semantic parts. Subsequently, we train a multimodal large language model (LLM) that translates 3D point cloud into executable Blender Python scripts. Our approach not only achieves superior performance in shape-to-code reconstruction tasks but also facilitates intuitive geometric and topological editing through convenient code modifications. Furthermore, our code-based representation enhances the reasoning capabilities of LLMs in 3D shape understanding tasks. Together, these contributions establish MeshCoder as a powerful and flexible solution for programmatic 3D shape reconstruction and understanding. Project homepage: <https://daibingquan.github.io/MeshCoder>.

1 Introduction

Inferring shape programs from 3D observations is of great importance for reverse engineering, shape editing, and 3D structure understanding. Prior work [1, 2, 3] has explored this problem by defining Domain-Specific Languages (DSLs) to model geometric and structural properties of objects and training neural networks to map 3D observations to shape programs. However, existing methods struggle to generalize to objects with complex geometry and structure. Two key limitations underlie this gap. First, existing DSLs are constrained to modeling simple primitives (e.g., cubes, spheres, cylinders) and cannot represent real-world objects with intricate parts. Second, training shape-to-code inference models demands large-scale paired datasets of 3D objects and their corresponding code, while such datasets are scarce. Prior work often relies on datasets with limited categories, geometric complexity and part count.

To address these challenges, we introduce MeshCoder, a novel framework for generating Blender Python scripts that reconstruct complex 3D objects into their constituent parts. First, we design a set of expressive Blender Python APIs that are capable of synthesizing intricate geometries beyond simple primitives. For instance, our APIs can create complex shapes by translating a 2D section curve along a specified trajectory, bridging section curves of different shapes, adding bevels or applying Boolean operations on basic shapes, repeating a basic shape in one dimension or two dimensions. With these concise yet powerful Blender Python APIs, we can model highly complex shapes, addressing the limitations of prior DSLs.

Second, we present a novel pipeline to construct a large-scale paired object-code dataset. We begin by synthesizing diverse object parts using our APIs with parametrically sampled parameters, yielding a part-level dataset. A part-to-code inference model is then trained on this dataset to predict code for individual parts. Next, we employ this model to construct a holistic object-code dataset. We use Infinigen-Indoor [4] to generate a dataset of objects, and each object is decomposed into its constituent parts. We use the part-to-code inference model to predict code for each part of an object, and then carefully design rules to concatenate code of all parts to obtain code of the object. This process yields a dataset of approximately 1 million objects spanning 41 categories, with objects up to more than 100 parts. Finally, we train a multimodal large language model (LLM) on this dataset to infer code from 3D objects. We use point clouds as 3D shape representations due to their ease of acquisition, and use a triplane-based tokenizer to transform the input point cloud to a set of fixed-length tokens. These tokens are fed into the LLM to generate Blender Python scripts that replicate input geometries in distinct semantic parts.

We evaluate our approach against existing shape-to-code methods, with experimental results and quantitative metrics demonstrating that our framework significantly outperforms prior work. Furthermore, by representing shapes as executable code, our method facilitates intuitive geometric and topological editing through simple code modifications. This capability enables precise alterations to object geometry and mesh topology, enhancing flexibility in downstream applications. Additionally, we conduct experiments on shape structural and geometric understanding tasks, revealing that our code-based representation improves the reasoning capabilities of large language models (LLMs) when interpreting 3D shapes. In summary, our contributions are outlined as follows:

- We have developed a comprehensive set of Blender Python APIs, facilitating the modeling of intricate geometries. This enhanced API suite empowers the procedural generation of complex 3D structures, effectively addressing the limitations of traditional domain-specific languages (DSLs) in representing detailed and varied shapes.
- We propose a pipeline to construct a large-scale paired object-code dataset. Using the dataset we constructed, we can train an shape-to-code inference model.
- We trained MeshCoder, an **Object-to-Code inference framework** that generates Blender Python scripts to reconstruct 3D meshes from point clouds in a structured and editable manner. Our model encodes 3D shapes into part-level code, simplifying mesh editing and enhancing LLMs’ understanding of 3D objects.

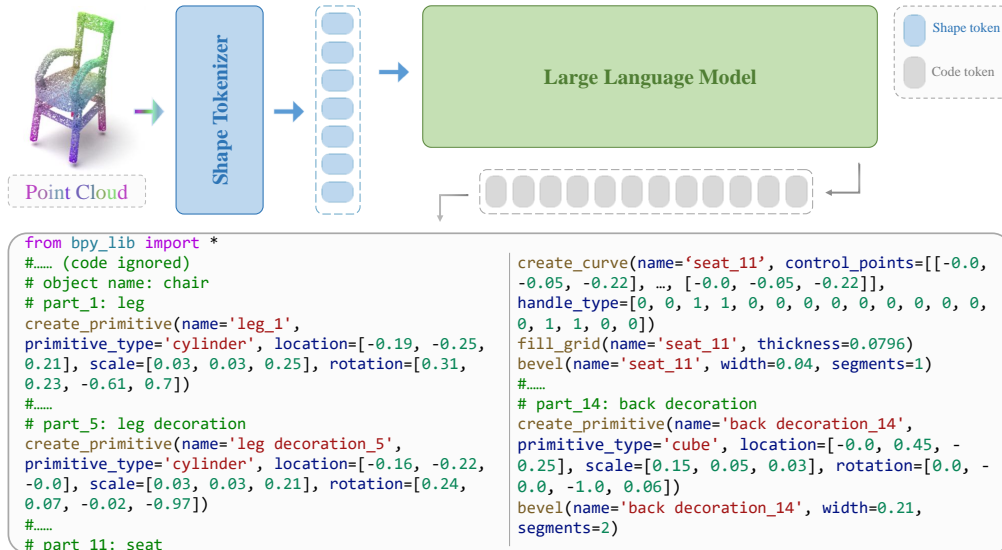


Figure 2: Overview of MeshCoder. The input point cloud is first encoded into shape tokens via a shape tokenizer. These tokens are then fed into a large language model (LLM), which autoregressively generates executable code representing part-based 3D structures. The decoded code specifies object’s name, part identities and names, enabling interpretable and modular reconstruction.

2 Related Work

Shape programs. Shape programs provide a structured and interpretable framework for representing 3D geometry by utilizing domain-specific languages to describe the generative processes of shapes. Early work such as ShapeAssembly [5] introduced explicit shape programs that capture the hierarchical and part-based organization of objects. Subsequent methods, including ShapeCoder [6], PLAD [2], and ShapeLib [7], progressively improved program abstraction, learning efficiency, and scalability with large language models. Other approaches, such as those proposed by Liang [3] and Tian et al. [1], incorporate differentiable rendering or neuro-symbolic reasoning to enhance program inference and execution. While these methods exhibit strong generalization capabilities in composing simple geometric elements like boxes and cylinders, they often struggle to model complex part geometries or generate artist-grade quad meshes, which restricts their application in high-fidelity asset creation. In addition, a range of methods in CAD program generation [8, 9, 10, 11, 12, 13] have explored synthesizing code representations for individual CAD parts. However, these approaches are limited to isolated component generation and lack the capability to model complete multi-part objects with coherent structural relationships.

Part-based Representation. Part-based representations have proven highly valuable in 3D shape analysis and synthesis. Some approaches [14, 15, 16, 17, 18, 19, 20, 21, 22, 23] take a generative approach, assembling objects by combining predefined or learned parts into complete 3D structures. Other methods [24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35] focus on segmenting 3D objects

into individual parts, enabling more modular and flexible manipulation of shapes. For instance, SAMPart3D [24] introduces a scalable zero-shot 3D part segmentation framework that segments any 3D object into semantic parts at multiple granularities without requiring predefined part label sets as text prompts. PartSLIP [28] explores low-shot part segmentation of 3D point clouds by leveraging a pretrained image-language model, GLIP, transferring rich knowledge from 2D to 3D through GLIP-based part detection on point cloud rendering and a novel 2D-to-3D label lifting algorithm. SATR [26] performs zero-shot 3D shape segmentation via text descriptions by using a zero-shot 2D object detector, inferring 3D segmentation from multi-view 2D bounding box predictions by exploiting the topological properties of the underlying surface. Despite these advancements in part segmentation and reconstruction, these methods do not translate segmented parts into executable code representations, limiting their integration into code-driven design workflows.

3 Methodology

As shown in Figure 2, we aim to train an object-to-code inference model that takes in a point cloud of an object, and then predict the Blender python scripts of each part of the object. When executing the python scripts in Blender, we can obtain the same object in separated parts. To train such an object-to-code inference model, we need a dataset of paired objects and the corresponding codes. To obtain such a dataset, we first train a part-to-code inference model that predicts code for a single part on our synthetic dataset of paired parts and the corresponding codes. Then, given a dataset of objects separated in different parts, we use the trained part-to-code inference model to predict code for every part of an object. Finally, we concatenate the codes of every part of the object and obtain the code of the object. Now, we have a dataset of paired objects and the corresponding codes, and are ready to train the object-to-code inference model.

We explain the key steps described above in details in the following sections. First, we explain how to synthesize a dataset of paired parts and the corresponding codes in Section 3.1. Then, we describe the training procedure of the part-to-code inference model in Section 3.2. Next, we use the part-to-code inference model to obtain the code of an entire object in Section 3.3. Finally, we train the object-to-code inference model in Section 3.4.

3.1 Part Dataset

We aim to generate a dataset of paired part shapes and codes. To do so, we implement probabilistic programs to generate Blender Python scripts, and obtain the corresponding shape by executing the code in Blender. We carefully design these probabilistic programs and ensure that the shapes generated are within the range $[-1, 1]^3$. There are several types of shapes that we generate, as illustrated in Figure 3. We explain them in the following paragraphs.

Primitive. Primitives are a set of fundamental geometric shapes, consistent with those defined in Blender. Specifically, we consider five basic shapes: cube, cylinder, UV sphere, cone, and torus. Each primitive is parameterized by three attributes: `location` ($\mathbf{location} \in \mathbb{R}^3$), `rotation` ($\mathbf{rotation} \in \mathbb{H}$), and `scale` ($\mathbf{scale} \in \mathbb{R}^3$), where \mathbb{H} denotes the space of unit quaternions. The `location` specifies the shape’s position in 3D space, `rotation` defines its orientation via quaternions, and `scale` determines the shape’s size along its local axes. Examples of Primitives can be found in the first row of Figure 3.

Translation. Translation is defined as the geometry obtained by sweeping a 2D cross-sectional shape along a 3D trajectory curve, which is equivalent to Sweep in CAD. As illustrated in the second row of Figure 3, during this translation process, the tangent direction of the 3D trajectory remains perpendicular to the 2D shape, and the size of the section shape can change along the 3D trajectory. For a more detailed explanation, please refer to A.1. To implement this, we first define a 2D shape using a set of control points (i.e., spatial coordinates), and then specify a 3D trajectory curve in a similar manner. Specifically, our experiments consider five types of cross-sectional shapes: rectangles, circles, circular arcs, polygons, and Bézier curves. For the trajectories, we define six forms: straight lines, polylines, circles, circular arcs, rectangles, and Bézier curves. Notably, this method also allows a 2D shape to rotate around an axis to form a solid of revolution, making it suitable for modeling objects such as bottles and plates.

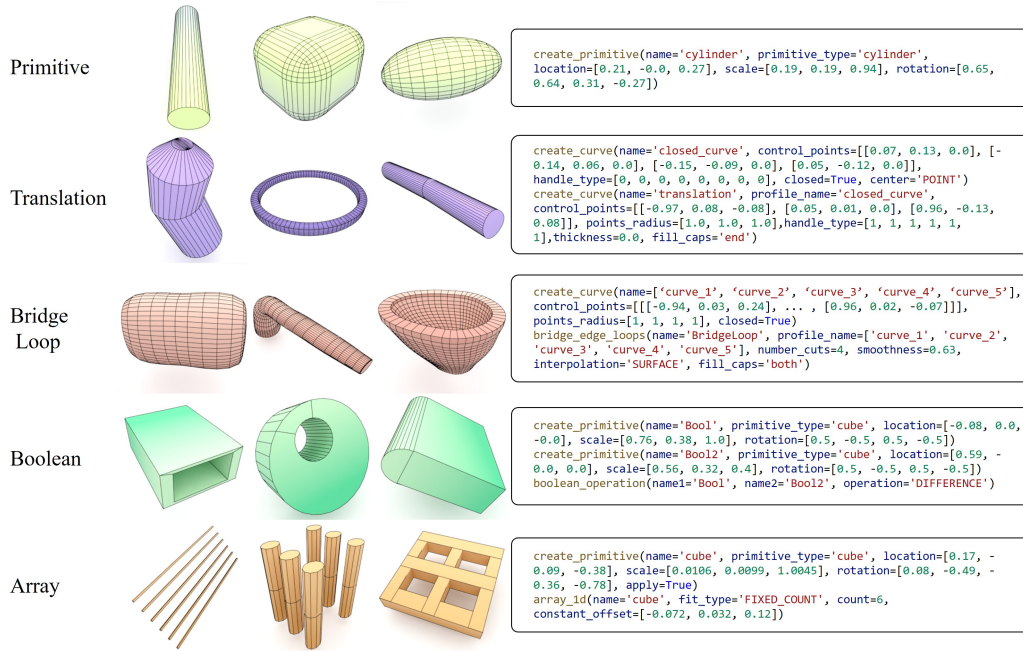


Figure 3: Visualization of basic geometric shape types and their corresponding code. For each shape category, the code shown corresponds to the first example.

Bridge loop. Although the Translation method is capable of generating certain complex objects, it remains constrained in several ways. For instance, in the Translation operation, the 2D cross-sectional shape is always orthogonal to the tangent direction of the trajectory. Moreover, the section shape is allowed to change only in scale, without any deformation in its geometry. To address this limitation, we introduce an alternative method for constructing geometries, namely the Bridge Loop. This geometry is constructed by first generating a sequence of 2D shapes and then connecting their corresponding vertices to form a continuous 3D geometry, which is equivalent to Loft in CAD. Some cases can be seen in the third row of Figure 3. The Bridge Loop approach enables the creation of more complex geometries compared to those achievable via Translation alone. For a more detailed explanation, please refer to A.1.

Boolean. Boolean geometries refer to geometries formed by applying Boolean operations—namely union, intersection, and difference—to two or more of the fundamental shape categories defined in Section 3.1. The union operation enables the construction of complex composite geometries, while the difference operation is used to generate geometries with holes or indentations. We can see some examples and their corresponding codes in the fourth row of Figure 3.

Array. When a particular type of primitive geometry appears repeatedly in a regular pattern, we do not invoke the construction function for each primitive individually, as this would result in lengthy code. Instead, we employ an Array method to construct the entire structure collectively. Specifically, we define two types of Arrays: 1D Arrays, where a geometry is repeatedly instantiated along a curve, and 2D Arrays, where repetition occurs across a plane. Cases of this type can be seen in the last row of Figure 3.

It is important to note that when designing our functions, the function parameters include two types: array-based parameters and individual parameters. During dataset construction, we select the parameter format based on the most suitable fit for the object: individual parameters are used for simpler parts, while array-based parameters are adopted when a part contains multiple similar shapes. For instance, the translation example in Figure 3 uses individual parameters, whereas the bridge loop example employs array-based parameters.

After designing the five fundamental types of template functions described above, we perform probabilistic random sampling over these functions and their parameters to generate a series of function code snippets. For each sampled code snippet y , we execute it to obtain the corresponding mesh M . In this way, we construct a dataset of paired code y and mesh M .

3.2 Part-to-code Inference Model

After constructing the dataset of paired code \mathbf{y} and mesh M , we sample a point cloud $\mathbf{x} \in \mathbb{R}^{N \times 3}$ from each mesh M , where N is the number of points in the point cloud. We train a part-to-code inference model h that takes in a point cloud \mathbf{x} and predict the corresponding code \mathbf{y} . The inference model consists of two modules: The shape tokenizer model and a finetuned LLM. The tokenizer model takes in the point cloud \mathbf{x} and outputs a set of fixed length tokens $\mathbf{z} \in \mathbb{R}^{L \times D}$, where L is the number of shape tokens and D is the dimension of each token. We set D to the same dimension as the word embeddings in the LLM. Thereafter, the LLM takes in the shape tokens \mathbf{z} and then predict \mathbf{y} , the code of the point cloud \mathbf{x} . We train the shape tokenizer model and finetune the LLM at the same time using the cross-entropy loss for the prediction of the next token in the shape code \mathbf{y} . We use Llama-3.2-1B as the base LLM and finetune it using LoRA.

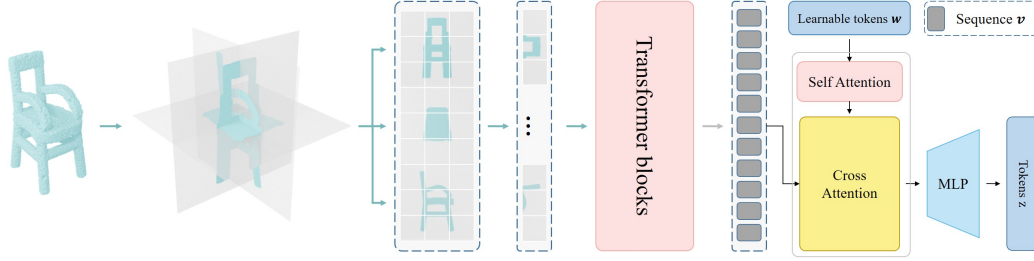


Figure 4: Architecture of the shape tokenizer. We first project the point cloud into the triplane and obtain triplane features. The triplane features are patchified and reshaped into a 1D sequence, and fed into transformer blocks to obtain triplane tokens. Finally, we use a set of learnable tokens to aggregate information from triplane tokens via cross-attention.

The shape tokenizer model. We explain the detailed structure of the shape tokenizer model. As shown in Figure 4, the shape tokenizer model transforms a point cloud $\mathbf{x} \in \mathbb{R}^{N \times 3}$ to a set of fixed length tokens $\mathbf{z} \in \mathbb{R}^{L \times D}$. We first project the point cloud \mathbf{x} to a triplane and obtain triplane feature $\mathbf{u} \in \mathbb{R}^{3 \times H \times W \times D_1}$, where H, W are the height and width of the planes, and D_1 is the dimension of the triplane feature. The coordinates of each point is fed to a shared MLP and a feature of dimension D_1 is obtained. We project each point’s feature to the three perpendicular planes according to the point’s position. Features projected to the same pixel are aggregated by max-pooling. Pixels that do not correspond to any point are filled with zeros. After obtaining the triplane feature \mathbf{u} , we patchify it and reshape it into a 1D sequence $\mathbf{v} \in \mathbb{R}^{(3 \cdot H/f \cdot W/f) \times D_1}$, where f is the patch size. We then feed the sequence \mathbf{v} to a set of transformer blocks and outputs $\mathbf{v}' \in \mathbb{R}^{(3 \cdot H/f \cdot W/f) \times D_1}$. Next, to compress the number of tokens fed into the LLM, we use a learnable set of tokens $\mathbf{w} \in \mathbb{R}^{L \times D_2}$ to aggregate information from \mathbf{v}' using cross attention:

$$\text{CrossAttn}(\text{Transformer}(\mathbf{w}), \mathbf{v}', \mathbf{v}'), \quad (1)$$

where Transformer denotes a transformer block, $\text{CrossAttn}(Q, K, V)$ denotes a cross attention block, and Q, K, V are query, key, value, respectively. By feeding \mathbf{w} to a set of these cross attention blocks, we obtain tokens $\mathbf{w}' \in \mathbb{R}^{L \times D_2}$ that contain information about the point cloud \mathbf{x} . Finally, we use an MLP to transform the dimension of \mathbf{w}' from D_2 to D and obtain shape tokens $\mathbf{z} \in \mathbb{R}^{L \times D}$, where D is the dimension of the word embeddings in the LLM. Now, the shape tokens \mathbf{z} can be readily fed to the LLM and predict the code corresponding to the point cloud \mathbf{x} .

3.3 Assemble Parts to Objects

After training the part-to-code inference model h , we can use it to obtain the code of an object. Given a dataset of objects, in which each object \mathcal{O} is separated into its constituent parts $\mathcal{O} = \{\mathbf{q}_i | i = 1, 2, \dots, M\}$, where \mathbf{q}_i is the i -th part of object \mathcal{O} , and M is the number of parts of the object \mathcal{O} . We also assume that each part \mathbf{q}_i has its semantic label. We can use the part-to-code inference model h to obtain the code of each part. Specifically, we first normalize each part \mathbf{q}_i to the cube $[-1, 1]^3$ according to its minimum bounding box and obtain the shape \mathbf{q}'_i . Then we use the part-to-code inference model h to obtain its code $\mathbf{y}'_i = h(\mathbf{q}'_i)$. We then implement algorithms to transform the relevant numerical parameters in the code \mathbf{y}'_i to the original location, scale, and pose of \mathbf{q}'_i and obtain the code \mathbf{y}_i of the original shape \mathbf{q}_i . Finally, we concatenate the codes of all parts of

the object, add semantic information to the code for each part, and obtain the code of the object $\mathbf{y} = \{\mathbf{y}_i | i = 1, 2, \dots, M\}$. When concatenating the code, we sort each part based on its spatial position. Specifically, we assign an index to each part following a spatial order from bottom to top, left to right, and front to back. An overview of this pipeline is illustrated in Figure 5. During code inference, the part point cloud \mathbf{q}_i is first transformed into a canonical space using a rotation matrix \mathbf{R} , translation matrix \mathbf{T} , and scaling factor s , resulting in \mathbf{q}'_i . The trained part-to-code inference model \mathbf{h} generates the code \mathbf{y}'_i of \mathbf{q}'_i . \mathbf{y}'_i is then transformed back to the original pose and scale using the inverse of \mathbf{R} , \mathbf{T} , and s , and we obtain the code \mathbf{y}_i of \mathbf{q}_i .

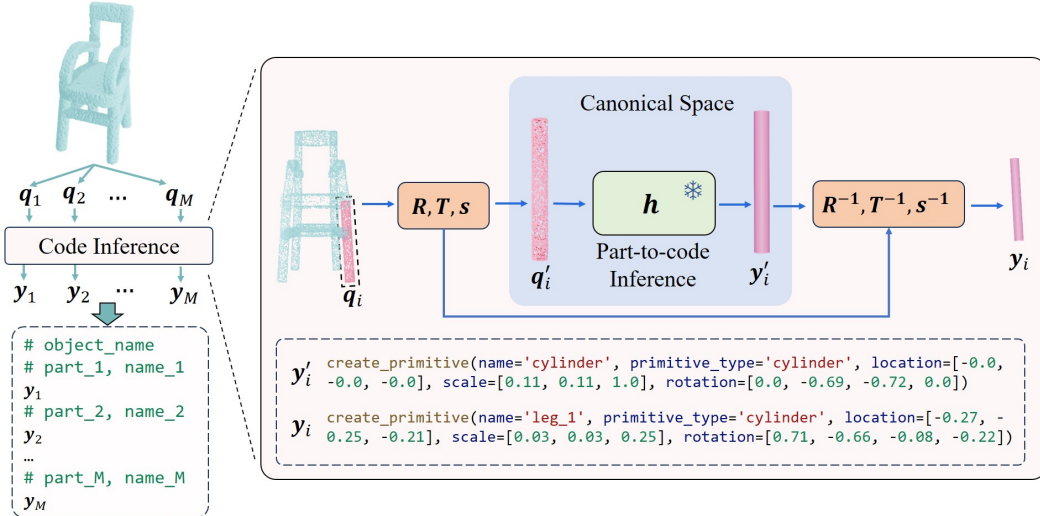


Figure 5: Pipeline of object-level code dataset construction using the part-to-code inference model. For each part point cloud \mathbf{q}_i , the code inference module independently predicts its corresponding code \mathbf{y}_i . All part codes \mathbf{y}_i are then concatenated to form the complete object code. We also add meaningful semantic information to the object code following the template shown in the figure. The complete code of the example chair is shown in Figure 2.

3.4 Object-to-code Inference Model

After obtaining the code \mathbf{y} of each object \mathcal{O} in the dataset, we can use them to train an object-to-code inference model. Our object-to-code inference model has the same structure as the part-to-code inference model described in Section 3.2. We initialize the weights of the object-to-code inference model as the weights of the trained part-to-code inference model, and use the same training method in Section 3.2 to train the object-to-code inference model. It is worth noting semantic information in the ground-truth code of objects enables the object-to-code inference model to learn the semantic structure of objects, and facilitate 3D shape understanding.

4 Experiment

4.1 Datasets

4.1.1 Synthetic Part Dataset

To facilitate the training of our part-to-inference model, we first constructed a synthetic part dataset. Specifically, we utilized functions from our basic shape code library, randomly sampling their parameters based on manually defined distributions to generate paired data of synthetic parts and corresponding code. This process yielded 1.5 million point cloud–code pairs for primitive shapes, 3 million for Translation-based parts, 1.5 million for Bridge Loop structures, 1.5 million for Boolean operations, and 2.4 million for Array-based constructions. In total, our constructed part dataset comprises around 10 million point cloud–code pairs. We partitioned the dataset into 70% for training, 15% for validation, and 15% for testing.

4.1.2 Object Dataset

We trained our model on the Infinigen Indoor [4] dataset. **Infinigen Indoor** is a procedural framework for generating synthetic 3D indoor objects, where each generated instance is automatically composed by its corresponding parts. We have made extensive modifications to the original Infinigen codebase to enable it to produce both individual components and their complete assemblies. Using this framework, we constructed a synthetic dataset comprising 41 common object categories, generating 1 million object-code pairs in total. We partitioned the dataset into training, validation, and test sets, following the same split strategy as the Synthetic Part Dataset. For more details, please refer to the A.1.

4.2 Implementation Details

We conduct training and evaluation on the Infinigen Indoor datasets [36]. For the part-to-code reconstruction model, we adopt the AdamW optimizer and train it for 20 epochs on NVIDIA A100 GPUs with a batch size of 512, and a learning rate of 10^{-4} . We evaluate the model at every epoch and select the checkpoint with the lowest L_2 Chamfer Distance (CD) loss. Then we initialize the weights of the object-to-code reconstruction model with the weights of the trained part-to-code reconstruction model, and train the model on Infinigen Indoor dataset for 10 epochs, with a batch size of 256, and a learning rate of 10^{-4} . The checkpoint with the lowest CD loss is selected. For additional training details and the parameter settings of the models, please refer to A.3 and A.2.

4.3 Reconstruction Performance

For reconstruction performance, we compare our method with two representative shape-to-code baselines, Shape2Prog [1] and PLAD [2]. Figure 6 illustrates visualization comparisons of results. We adopt IoU and L_2 CD as our evaluation metrics. Specifically, we voxelize the model’s predicted outputs into 32^3 grids and compute the IoU between the predicted and ground truth voxel grids. In parallel, we sample point clouds from both the predicted outputs and the ground truth, and calculate the Chamfer Distance between the two point clouds. Regarding the number of points and normalization, please refer to the appendix A.4. In Table 1, we present reconstruction metrics for some specific object categories as well as the overall performance across the entire dataset. It can be observed that our method consistently outperforms the baselines in both IoU and CD metrics. Complete results for all categories in each dataset are provided in A.4. We conducted a series of ablation studies to evaluate the impact of various components within our model. For comprehensive details on these experiments, please refer to A.4.

Table 1: Quantitative comparison of reconstruction performance between MeshCoder and baselines.

Method	CD($\times 10^{-2}$) \downarrow						IoU (%) \uparrow					
	Lamp	Chair	Sofa	TableDining	Toilet	All	Lamp	Chair	Sofa	TableDining	Toilet	All
Shape2Prog	25.44	1.30	2.14	1.03	7.51	6.01	16.96	49.68	65.29	71.26	51.14	45.03
PLAD	1.40	2.26	1.52	5.52	2.30	1.87	69.58	40.93	81.33	58.43	62.61	67.62
MeshCoder	0.004	0.060	0.027	0.024	0.022	0.063	86.23	81.87	93.81	88.14	89.10	86.75

4.4 Shape Editing

MeshCoder facilitates the transformation of 3D shapes into high-level, human-readable code representations, significantly enhancing the interpretability and editability of complex geometries. This capability enables intuitive and precise modifications through code-based interventions. Our shape editing encompasses two primary categories: geometric editing and topological editing. As illustrated in Figure 7, geometric editing can be performed by adjusting function calls or modifying specific parameters within the generated code. For instance, we can adjust the parameters of the code to convert a square tabletop into a larger circular one. Additionally, topological editing, which is illustrated in Figure 8 such as adjusting mesh resolution, can be achieved by modifying designated parameters within the code, allowing for control over the mesh’s complexity and surface detail. This code-centric approach streamlines the process of modifying 3D models, making it more accessible and efficient for applications requiring iterative design and customization. Additionally, it empowers users to adjust the model resolution according to their desired balance between storage requirements and mesh quality. For additional results and details, please refer to A.5 in the appendix.



Figure 6: Qualitative comparison of reconstruction performance between MeshCoder and baselines. MeshCoder can accurately reconstruct objects with intricate parts and complex structures.

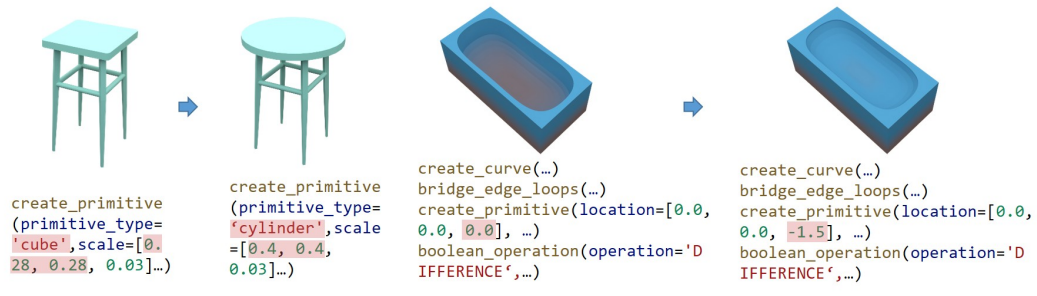


Figure 7: Parameter modification in the code conveniently to alter the geometric shape. Left: Change tabletop from square to circular. Right: Make the bathtub shallower.

4.5 Shape Understanding

MeshCoder is capable of predicting object codes with rich semantic information. These codes effectively capture structural and geometric details, making them valuable for shape understanding. By inputting the predicted codes into GPT, we can assist it in comprehending object structures. We conduct experiments on shape understanding, with an example illustrated in Figure 9. Additional results and details are given in A.6 in the appendix.

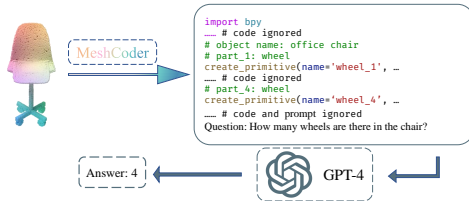


Figure 9: The pipeline of conducting experiments on shape understanding.

5 Limitations

Although our method achieves significant advancements in category diversity, geometric complexity, and reconstruction accuracy compared to existing approaches, it primarily targets human-made objects. The applicability of code-based representations to organic forms, such as animals and humans, remains underdeveloped. We reserve this as a direction for future research.

6 Conclusion

In this work, we present MeshCoder, a comprehensive framework that translates 3D point cloud data into editable Blender Python scripts, enabling detailed reconstruction and intuitive editing of complex 3D objects. By developing a robust set of Blender Python APIs, we facilitate the modeling of intricate geometries. Leveraging these APIs, we constructed a large-scale dataset pairing 3D objects with their corresponding code representations, decomposed into semantic parts. Subsequently, we

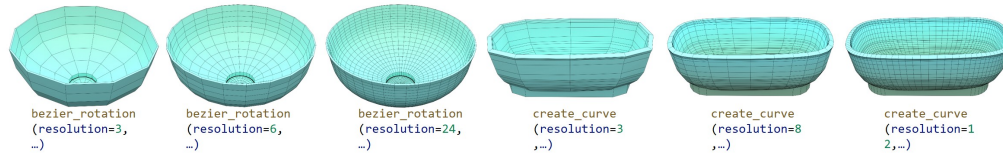


Figure 8: Mesh resolution adjustment by modifying the resolution parameters in the code. The figure depicts results with progressively increasing resolution from left to right.

trained a multimodal large language model (LLM) capable of generating executable Blender scripts from point cloud inputs. Our approach not only achieves superior performance in shape-to-code reconstruction tasks but also enhances the reasoning capabilities of LLMs in 3D shape understanding. By representing shapes as structured code, MeshCoder offers a flexible and powerful solution for programmatic 3D shape reconstruction and editing, paving the way for advanced applications in reverse engineering, design, and analysis.

7 Acknowledgement

This work is funded by the Shenzhen Science and Technology Project (Grants JCYJ20220818101001004 and KJZD20240903103210014), the National Key R&D Program of China (Grant 2022ZD0160201), Shanghai Artificial Intelligence Laboratory, and in part by the HKU Startup Fund.

References

- [1] Yonglong Tian, Andrew Luo, Xingyuan Sun, Kevin Ellis, William T. Freeman, Joshua B. Tenenbaum, and Jiajun Wu. Learning to infer and execute 3d shape programs. In *International Conference on Learning Representations*, 2019.
- [2] R Kenny Jones, Homer Walke, and Daniel Ritchie. Plad: Learning to infer shape programs with pseudo-labels and approximate distributions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9871–9880, 2022.
- [3] Yichao Liang. Learning to infer 3d shape programs with differentiable renderer. *arXiv preprint arXiv:2206.12675*, 2022.
- [4] Alexander Raistrick, Lingjie Mei, Karhan Kayan, David Yan, Yiming Zuo, Beining Han, Hongyu Wen, Meenal Parakh, Stamatis Alexandropoulos, Lahav Lipson, Zeyu Ma, and Jia Deng. Infinigen indoors: Photorealistic indoor scenes using procedural generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21783–21794, June 2024.
- [5] R Kenny Jones, Theresa Barton, Xianghao Xu, Kai Wang, Ellen Jiang, Paul Guerrero, Niloy J Mitra, and Daniel Ritchie. Shapeassembly: Learning to generate programs for 3d shape structure synthesis. *ACM Transactions on Graphics (TOG)*, 39(6):1–20, 2020.
- [6] R Kenny Jones, Paul Guerrero, Niloy J Mitra, and Daniel Ritchie. Shapecoder: Discovering abstractions for visual programs from unstructured primitives. *ACM Transactions on Graphics (TOG)*, 42(4):1–17, 2023.
- [7] R Kenny Jones, Paul Guerrero, Niloy J Mitra, and Daniel Ritchie. Shapelib: designing a library of procedural 3d shape abstractions with large language models. *arXiv preprint arXiv:2502.08884*, 2025.
- [8] Pradeep Kumar Jayaraman, J. Lambourne, Nishkrit Desai, Karl D. D. Willis, Aditya Sanghi, and Nigel Morris. Solidgen: An autoregressive model for direct b-rep synthesis. *ArXiv*, abs/2203.13944, 2022. URL <https://api.semanticscholar.org/CorpusID:247761924>.

- [9] Mohammad Sadil Khan, Elona Dupont, Sk Aziz Ali, Kseniya Cherenkova, Anis Kacem, and Djamila Aouada. Cad-signet: Cad language inference from point clouds using layer-wise sketch instance guided attention. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4713–4722, June 2024.
- [10] Rundi Wu, Chang Xiao, and Changxi Zheng. Deepcad: A deep generative network for computer-aided design models. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6752–6762, 2021. doi: 10.1109/ICCV48922.2021.00670.
- [11] Jingwei Xu, Zibo Zhao, Chenyu Wang, Wen Liu, Yi Ma, and Shenghua Gao. Cad-mllm: Unifying multimodality-conditioned cad generation with mllm, 2024.
- [12] Xiang Xu, Pradeep Kumar Jayaraman, Joseph G Lambourne, Karl DD Willis, and Yasutaka Furukawa. Hierarchical neural coding for controllable cad model generation. In *International Conference on Machine Learning*, pages 38443–38461, 2023.
- [13] Jianyu Wu, Yizhou Wang, Xiangyu Yue, Xinzhu Ma, Jingyang Guo, Dongzhan Zhou, Wanli Ouyang, and Shixiang Tang. Cmt: A cascade mar with topology predictor for multimodal conditional cad generation, 2025.
- [14] Juil Koo, Seungwoo Yoo, Minh Hieu Nguyen, and Minhyuk Sung. Salad: Part-level latent diffusion for 3d shape generation and manipulation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14441–14451, 2023.
- [15] Anran Liu, Cheng Lin, Yuan Liu, Xiaoxiao Long, Zhiyang Dou, Hao-Xiang Guo, Ping Luo, and Wenping Wang. Part123: part-aware 3d reconstruction from a single-view image. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–12, 2024.
- [16] Minghao Chen, Roman Shapovalov, Iro Laina, Tom Monnier, Jianyuan Wang, David Novotny, and Andrea Vedaldi. Partgen: Part-level 3d generation and reconstruction with multi-view diffusion models. *arXiv preprint arXiv:2412.18608*, 2024.
- [17] Yuhang Huang, SHilong Zou, Xinwang Liu, and Kai Xu. Part-aware shape generation with latent 3d diffusion of neural voxel fields. *arXiv preprint arXiv:2405.00998*, 2024.
- [18] Lin Gao, Jie Yang, Tong Wu, Yu-Jie Yuan, Hongbo Fu, Yu-Kun Lai, and Hao Zhang. Sdm-net: Deep generative network for structured deformable mesh. *ACM Transactions on Graphics (TOG)*, 38(6):1–15, 2019.
- [19] Zhijie Wu, Xiang Wang, Di Lin, Dani Lischinski, Daniel Cohen-Or, and Hui Huang. Sagnet: Structure-aware generative network for 3d-shape modeling. *ACM Transactions on Graphics (TOG)*, 38(4):1–14, 2019.
- [20] Kaichun Mo, Paul Guerrero, Li Yi, Hao Su, Peter Wonka, Niloy Mitra, and Leonidas J Guibas. Structurenets: Hierarchical graph networks for 3d shape generation. *arXiv preprint arXiv:1908.00575*, 2019.
- [21] Rundi Wu, Yixin Zhuang, Kai Xu, Hao Zhang, and Baoquan Chen. Pq-net: A generative part seq2seq network for 3d shapes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 829–838, 2020.
- [22] George Kiyohiro Nakayama, Mikaela Angelina Uy, Jiahui Huang, Shi-Min Hu, Ke Li, and Leonidas Guibas. Diffacto: Controllable part-based 3d point cloud generation with cross diffusion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14257–14267, 2023.
- [23] Dmitry Petrov, Matheus Gadelha, Radomír Měch, and Evangelos Kalogerakis. Anise: Assembly-based neural implicit surface reconstruction. *IEEE Transactions on Visualization and Computer Graphics*, 2023.
- [24] Yunhan Yang, Yukun Huang, Yuan-Chen Guo, Liangjun Lu, Xiaoyang Wu, Edmund Y Lam, Yan-Pei Cao, and Xihui Liu. Sampart3d: Segment any part in 3d objects. *arXiv preprint arXiv:2411.07184*, 2024.

- [25] Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip HS Torr, and Vladlen Koltun. Point transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 16259–16268, 2021.
- [26] Ahmed Abdelreheem, Ivan Skorokhodov, Maks Ovsjanikov, and Peter Wonka. Satr: Zero-shot semantic segmentation of 3d shapes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15166–15179, 2023.
- [27] Zongji Wang and Feng Lu. Voxsegnet: Volumetric cnns for semantic part segmentation of 3d shapes. *IEEE transactions on visualization and computer graphics*, 26(9):2919–2930, 2019.
- [28] Minghua Liu, Yin hao Zhu, Hong Cai, Shizhong Han, Zhan Ling, Fatih Porikli, and Hao Su. Partslip: Low-shot part segmentation for 3d point clouds via pretrained image-language models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 21736–21746, 2023.
- [29] Yuchen Zhou, Jiayuan Gu, Xuanlin Li, Minghua Liu, Yunhao Fang, and Hao Su. Partslip++: Enhancing low-shot 3d part segmentation via multi-view instance segmentation and maximum likelihood estimation. *arXiv preprint arXiv:2312.03015*, 2023.
- [30] Yuheng Xue, Nenglun Chen, Jun Liu, and Wenyun Sun. Zerops: High-quality cross-modal knowledge transfer for zero-shot 3d part segmentation. *arXiv preprint arXiv:2311.14262*, 2023.
- [31] Ardian Umam, Cheng-Kun Yang, Min-Hung Chen, Jen-Hui Chuang, and Yen-Yu Lin. Partdistill: 3d shape part segmentation by vision-language model distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3470–3479, 2024.
- [32] George Tang, William Zhao, Logan Ford, David Benhaim, and Paul Zhang. Segment any mesh: Zero-shot mesh part segmentation via lifting segment anything 2 to 3d. *arXiv preprint arXiv:2408.13679*, 2024.
- [33] Anh Thai, Weiyao Wang, Hao Tang, Stefan Stojanov, James M Rehg, and Matt Feiszli. 3×2 : 3d object part segmentation by 2d semantic correspondences. In *European Conference on Computer Vision*, pages 149–166. Springer, 2024.
- [34] Ziming Zhong, Yanyu Xu, Jing Li, Jiale Xu, Zhengxin Li, Chaohui Yu, and Shenghua Gao. Meshsegmenter: Zero-shot mesh semantic segmentation via texture synthesis. In *European Conference on Computer Vision*, pages 182–199. Springer, 2024.
- [35] Xiangyang Zhu, Renrui Zhang, Bowei He, Ziyu Guo, Ziyao Zeng, Zipeng Qin, Shanghang Zhang, and Peng Gao. Pointclip v2: Prompting clip and gpt for powerful 3d open-world learning. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 2639–2650, 2023.
- [36] Kaichun Mo, Shilin Zhu, Angel X. Chang, Li Yi, Subarna Tripathi, Leonidas J. Guibas, and Hao Su. PartNet: A large-scale benchmark for fine-grained and hierarchical part-level 3D object understanding. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [37] Tencent Hunyuan3D Team. Hunyuan3d 2.0: Scaling diffusion models for high resolution textured 3d assets generation, 2025.
- [38] Jianfeng Xiang, Zelong Lv, Sicheng Xu, Yu Deng, Ruicheng Wang, Bowen Zhang, Dong Chen, Xin Tong, and Jiaolong Yang. Structured 3d latents for scalable and versatile 3d generation. *arXiv preprint arXiv:2412.01506*, 2024.
- [39] Matt Deitke, Dustin Schwenk, Jordi Salvador, Luca Weihs, Oscar Michel, Eli VanderBilt, Ludwig Schmidt, Kiana Ehsani, Aniruddha Kembhavi, and Ali Farhadi. Objaverse: A universe of annotated 3d objects. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13142–13153, 2022. URL <https://api.semanticscholar.org/CorpusID:254685588>.

- [40] Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. *ArXiv*, abs/2305.18290, 2023. URL <https://api.semanticscholar.org/CorpusID:258959321>.
- [41] Qihao Zhu Runxin Xu Junxiao Song Mingchuan Zhang Y.K. Li Y. Wu Daya Guo Zhihong Shao, Peiyi Wang. Deepseekmath: Pushing the limits of mathematical reasoning in open language models, 2024. URL <https://arxiv.org/abs/2402.03300>.
- [42] Ruowen Zhao, Junliang Ye, Zhengyi Wang, Guangce Liu, Yiwen Chen, Yikai Wang, and Jun Zhu. Deepmesh: Auto-regressive artist-mesh creation with reinforcement learning. *arXiv preprint arXiv:2503.15265*, 2025.

NeurIPS Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [\[Yes\]](#)

Justification: Yes, they can reflect the paper's contributions and scope.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [\[Yes\]](#)

Justification: We do discuss the limitations of the work.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [\[Yes\]](#)

Justification: We have discussed them in our paper.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: We have elaborated on our work in detail in our paper.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
 - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Yes, we will provide open access to the data and code.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: We specify all the training and test details in the paper.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: Yes, we mentioned some limitations and indicators.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: Yes, we mentioned that in our paper.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics [https://neurips.cc/public/EthicsGuidelines?](https://neurips.cc/public/EthicsGuidelines)

Answer: [Yes]

Justification: We obey the NeurIPS Code of Ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: We mentioned the real-world impact of the work.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to

generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.

- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [\[Yes\]](#)

Justification: We declare that what we are using is a public dataset.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [\[Yes\]](#)

Justification: We have done that.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their [licensing guide](#) can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. New assets

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: We will make our data public.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. Crowdsourcing and research with human subjects

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [No]

Justification: We did not conduct such experiments.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. Institutional review board (IRB) approvals or equivalent for research with human subjects

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [No]

Justification: We did it on open data, there are no potential dangers.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. Declaration of LLM usage

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [Yes]

Justification: LLM is part of our model, and we've fine-tuned it.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.

A Appendix of MeshCoder: LLM-Powered Structured Mesh Code Generation from Point Clouds

A.1 Datasets

A.1.1 The principles of Translation and Bridge Loop

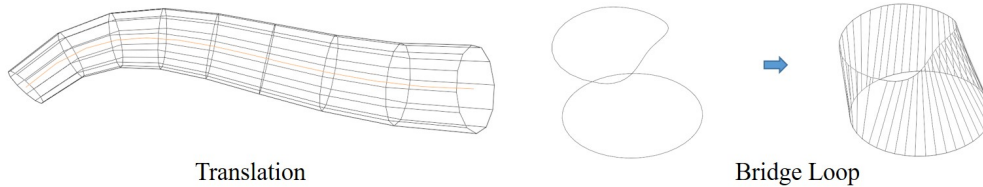


Figure 10: A schematic illustration of the principles of Translation and Bridge Loop. In the Translation module, the wireframe of the resulting mesh is shown as a cross-sectional circle is translated along a yellow trajectory. In the Bridge Loop module, the wireframe of the mesh is constructed by connecting the vertices of two 2D shapes.

As illustrated in the figure 10, in the Translation operation, a 2D cross-sectional shape (a circle in this example) and a 3D trajectory curve must first be defined. The Translation process generates a mesh by sweeping the 2D shape along the 3D trajectory. During this sweep, the cross-section remains perpendicular to the tangent direction of the trajectory at all times, and only uniform scaling (either enlargement or reduction) of the cross-section is permitted.

In contrast, the Bridge Loop operation begins with two predefined 2D shapes. By connecting the corresponding vertices of these two shapes, a mesh can be constructed. This method places no constraints on the types of 2D shapes used—meaning the two shapes can differ, such as a circle and an irregular closed shape in this example. Moreover, it imposes no restrictions on the relative orientations of the shapes. As a result, Bridge Loop overcomes the limitations of Translation, which requires the cross-section to align with the trajectory’s tangent direction. This enables Bridge Loop to generate more complex geometries that Translation cannot produce.

A.1.2 Part datasets

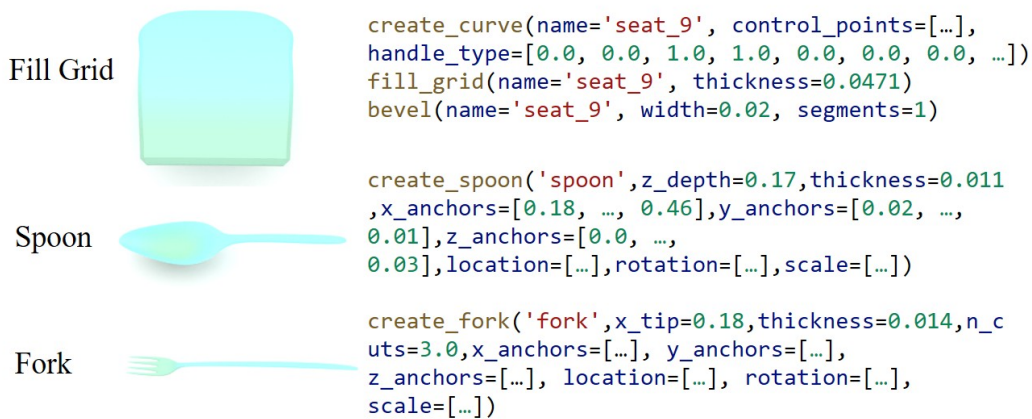


Figure 11: The Fill Grid type, Spoon type and Fork type in basic shape code library

For certain shapes that are difficult to represent using the method we defined in Section 3.1, we introduce three additional categories: the Fill Grid type, Spoon type and Fork type. As illustrated in the Figure 11. For the Fill Grid type, we first construct a closed 3D shape (as opposed to the 2D cross-sectional shape used in Translation), fill it to form a surface, and then extrude it along its

normal direction to generate the final mesh. For the Spoon and Fork type, we draw inspiration from the implementation in Infinigen Indoor [4] and design dedicated procedural functions tailored for their generation.

We present two core functions from our codebase: the complete implementation for creating primitives (Figure 22) and the complete implementation for creating curves (Figure 23). The full codebase can be found in the supplementary materials.

More examples of parts and their corresponding complete code implementations are provided in Figures 12, 13, 14, 15, and 16.

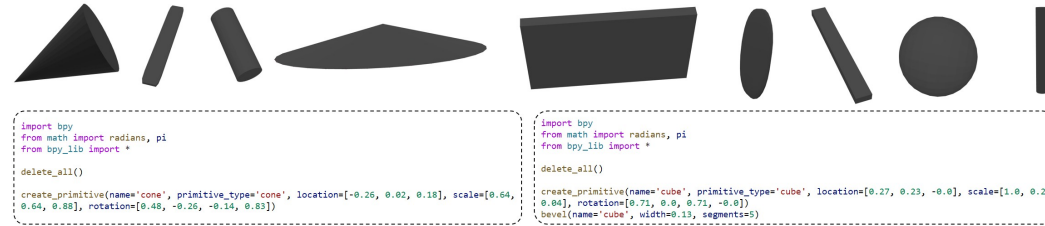


Figure 12: Examples of Primitive and complete code. And the code corresponds to the first two objects shown in the figure.

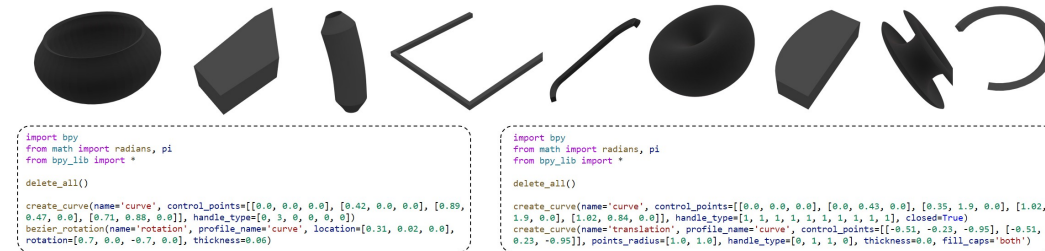


Figure 13: Examples of Translation and complete code. And the code corresponds to the first two objects shown in the figure.

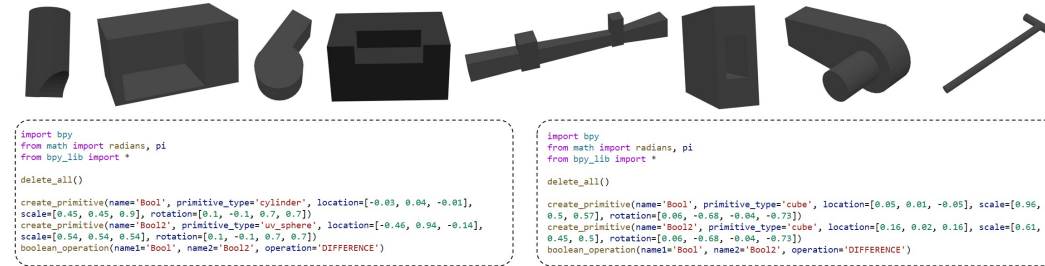


Figure 14: Examples of Boolean and complete code. And the code corresponds to the first two objects shown in the figure.

Taking the Primitive type as an example, we describe how to use functions from the basic shape code library to generate a synthetic part dataset. We begin by randomly selecting the type of primitive to generate (e.g., cube, cylinder, etc.). Next, for each axis, we independently uniform sample a value x from the range $[-2, 2]$, and then set the corresponding scale as 10^x . To determine the orientation of the shape, we uniformly sample a direction from a unit sphere and a roll angle from a uniform distribution. Once the orientation is fixed, we scale the shape uniformly along all three axes based on the size of its bounding box. Specifically, we ensure that the longest edge of the bounding box lies within the range $[1, 2]$. Finally, we assign the shape a random position within the 3D space such that the entire shape remains within the $[-1, 1]$ bounds. For other shape types beyond Primitive, we follow a similar approach by randomly assigning values to the relevant parameters.

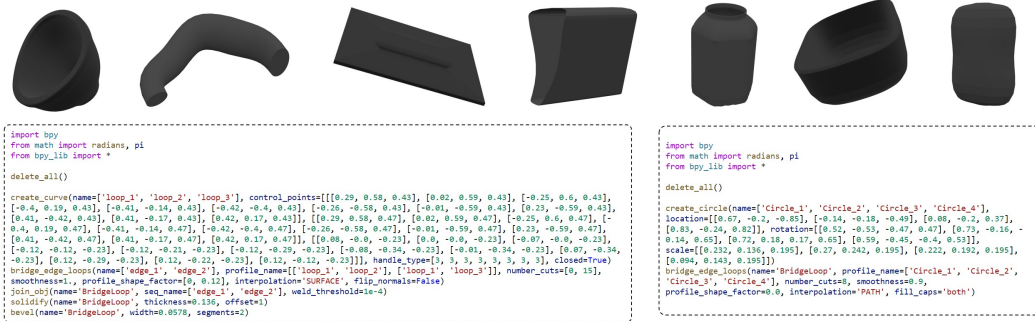


Figure 15: Examples of Bridge Loop and complete code. And the code corresponds to the first two objects shown in the figure.

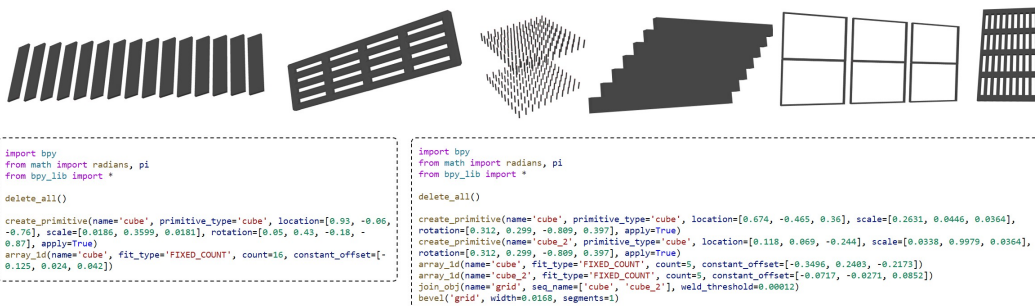


Figure 16: Examples of Array and complete code. And the code corresponds to the first two objects shown in the figure.

A.1.3 Object datasets

For assembling part codes into a complete program, we provide a full example containing the complete code, as shown in Figure 17. Regarding the ordering strategy used when assembling parts into a complete object, we adopt a consistent spatial heuristic to determine part sequence. Specifically, parts are arranged from bottom to top, left to right, and front to back. To implement this, we divide the 3D space into a $32 \times 32 \times 32$ grid and assign each part a characteristic grid cell that serves as the basis for sorting. The characteristic grid cell of a part is defined as follows: among all grid cells that the part occupies, we first select the one with the smallest z -coordinate. If multiple candidates share the same z -value, we choose the one with the smallest x -coordinate. If a tie still exists, we select the one with the smallest y -coordinate. Parts are then sorted based on the lexicographic order of these characteristic grid cells, which determines their final sequence within the object.

It is important to note that for each object, the prerequisite for successfully constructing its corresponding code lies in the ability of our part-to-code inference model to accurately infer all of its individual parts. We consider a part to be successfully inferred if the Chamfer Distance (CD) between the predicted point cloud and the ground truth is below 5×10^{-3} . Therefore, when constructing the object-code pairs dataset, we only include objects for which **all** constituent parts meet this criterion. Objects with any part failing to meet this standard are discarded. As a result, the number of successfully constructed object-code pairs is smaller than the total number of objects in the original Infigen dataset. In fact, the original Infigen dataset we use contains 1.57 million object instances, from which we successfully construct 1 million shape-code pairs. For training and evaluation, we split the full Infigen dataset into 70% for training, 15% for testing, and 15% for validation. Accordingly, MESHCODER is trained only on the subset of the shape-code pairs that fall within the training portion of the Infigen dataset. In contrast, the baseline models are trained on the full set of objects in the training split of the original Infigen dataset. Importantly, all evaluation results for our method and the baselines are reported on the same test set, i.e., the testing split of the complete Infigen dataset.

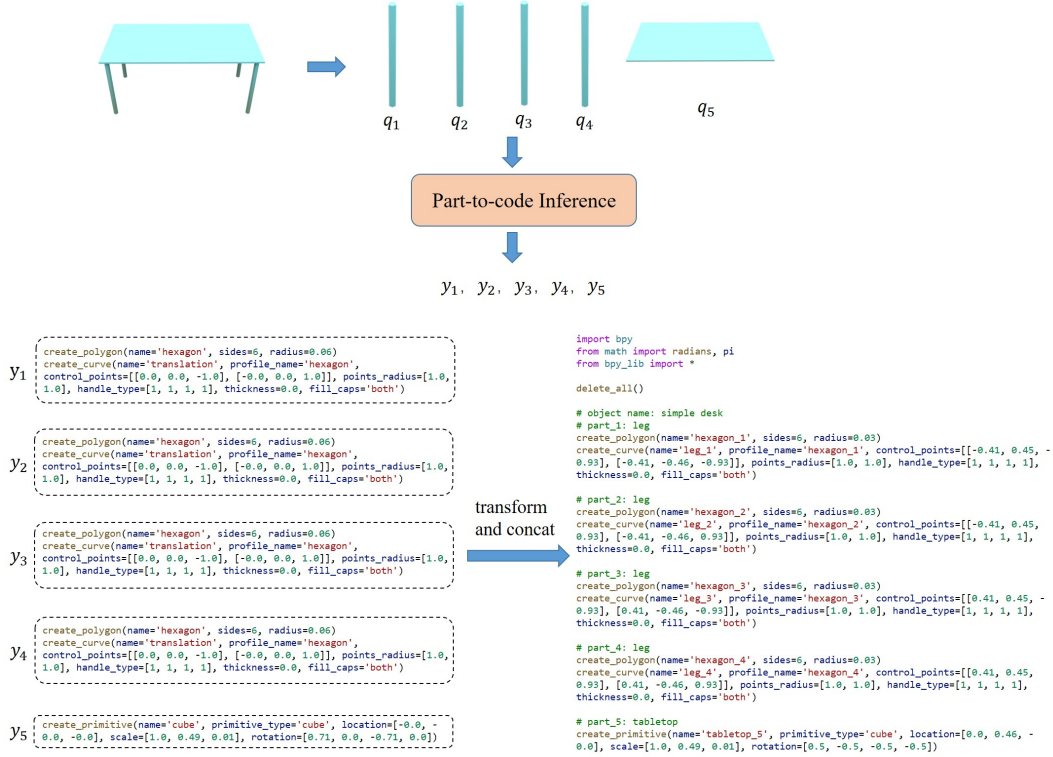


Figure 17: A complete code example of converting part codes into a full object program.

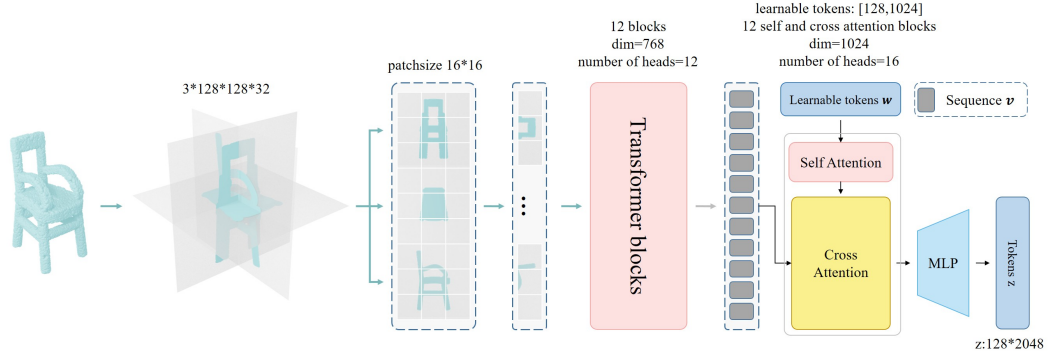


Figure 18: Detailed configuration of the shape tokenizer.

A.2 Model architecture

We explain the detailed structure of the shape tokenizer. As illustrated in the Figure 18, we first project the input point cloud of shape $\mathbb{R}^{n \times 3}$ onto three orthogonal planes to obtain tri-plane features with shape $\mathbb{R}^3 \times 256 \times 256 \times 32$. With a patch size set to 16×16 , these tri-plane features are encoded into tokens and fed into Transformer blocks, where the resulting representation is mapped to v and used as the key and value (K, V) inputs. Meanwhile, a set of learnable tokens with shape $\mathbb{R}^{128 \times 1024}$ are used as queries in a self and cross attention module. After passing through 12 layers of self and cross attention, we obtain output tokens of shape $\mathbb{R}^{128 \times 1024}$, which are then projected to the final representation of shape $\mathbb{R}^{128 \times 2048}$ via an MLP.

A.3 More training details

For the part-to-code reconstruction model, we adopt the AdamW optimizer and train it for 20 epochs on 64 NVIDIA A100 GPUs for about a week with a batch size of 512, and a learning rate of 10^{-4} . We evaluate the model at every epoch and select the checkpoint with the lowest L_2 Chamfer Distance (CD) loss. Then we initialize the weights of the object-to-code reconstruction model with the weights of the trained part-to-code reconstruction model, and train the model on Infinigen Indoor dataset for 10 epochs, with a batch size of 256, and a learning rate of 10^{-4} . It is trained on 64 NVIDIA A100 GPUs for about 2 days. The checkpoint with the lowest CD loss is selected.

To further enhance the robustness and generalization ability of the object-to-code inference model, we apply data augmentation techniques. Specifically, we perform random rotation and scaling on the objects. Additionally, during training, we randomly sample the number of points in each point cloud within the range of 4096 to 16384, and add Gaussian noise to further perturb the input. MeshCoder is trained and evaluated on a unified dataset that aggregates all object categories.

A.4 Complete experiment result of Shape Reconstruction

For **MeshCoder**, during inference, each object is represented by a point cloud containing 16,384 points. Given the input point cloud, the object-to-code inference model is able to predict the corresponding Blender Python script code. The resulting code is then executed to generate a corresponding mesh. We uniformly sample 100,000 points from the generated mesh and compute the Chamfer Distance (CD) to the input point cloud using the L_2 norm.

Given two point sets P and Q , each of size 100,000, the L_2 Chamfer Distance is defined as:

$$\text{CD}(P, Q) = \frac{1}{|P|} \sum_{x \in P} \min_{y \in Q} \|x - y\|_2 + \frac{1}{|Q|} \sum_{y \in Q} \min_{x \in P} \|y - x\|_2.$$

To evaluate IoU, we voxelize both the ground-truth mesh and the predicted mesh into grids of resolution 32^3 , and compute the voxel-based Intersection-over-Union (IoU) as:

$$\text{IoU} = \frac{|\mathcal{V}_{\text{pred}} \cap \mathcal{V}_{\text{gt}}|}{|\mathcal{V}_{\text{pred}} \cup \mathcal{V}_{\text{gt}}|},$$

where $\mathcal{V}_{\text{pred}}$ and \mathcal{V}_{gt} denote the sets of occupied voxels in the predicted and ground-truth voxel grids, respectively.

For **baseline methods**, which take voxel grids as input and output voxel grids, we first voxelize the ground-truth mesh into a 32^3 grid and feed it into the baseline models. The predicted voxel grid is then compared to the input voxelized ground truth to compute IoU. Additionally, we extract a mesh from the predicted voxel grid using the Marching Cubes algorithm and uniformly sample 100,000 points from the resulting mesh surface. These sampled points, along with the ground-truth point cloud, are then both uniformly scaled to fit within the $[-1, 1]^3$ volume. Finally, the Chamfer Distance is computed between the two normalized point clouds using the L_2 norm.

It’s noticed that for each object category, we independently train the baseline models, according to their official code, resulting in category-specific checkpoints. These models are then evaluated on the corresponding test sets for each category.

The quantitative comparison of reconstruction metrics between MeshCoder and baseline methods across all object categories is summarized in Table 2 and Table 3. Some additional examples of object reconstruction results and their complete code can be referred to Figure 24, 25, 26.

In addition to evaluating our object-to-code inference model, we also perform a quantitative assessment of our part-to-code inference model. Specifically, for each category described in Section 3.1, we construct a test set consisting of 10,000 samples. We evaluate the model’s performance using the CD and voxel IoU metrics on these test sets. The results, shown in Table 4, demonstrate strong performance across all categories, with low CD values and high IoU scores, indicating that our part-to-code inference model is highly effective in generating accurate code representations for individual parts.

Table 2: Comparison of reconstruction metrics across all categories. Chamfer Distance (CD) and IoU is shown in percentage (%).

Category	L2 CD($\times 10^{-2}$)			Voxel IoU (%)		
	MeshCoder	PLAD	Shape2prog	MeshCoder	PLAD	Shape2prog
ArmChair	0.04	2.31	4.44	94.33	78.79	62.74
BarChair	0.03	2.23	2.55	88.73	74.96	58.23
Bathtub	0.09	1.22	2.45	78.70	74.50	42.94
BeverageFridge	0.22	1.12	12.63	88.03	82.13	39.13
Bottle	0.01	1.08	6.34	88.65	65.58	40.24
Bowl	0.02	1.43	6.29	89.93	60.02	25.60
CeilingClassicLamp	0.02	1.98	3.94	96.13	76.01	59.07
CeilingLight	0.03	3.46	1.32	65.83	40.61	44.97
CellShelf	0.01	1.93	9.40	94.67	59.02	22.30
Chair	0.06	2.26	1.30	81.87	40.93	49.68
Chopsticks	0.03	1.38	21.06	82.24	55.68	11.25
Cup	0.06	1.40	7.35	85.96	62.03	29.47
DeskLamp	0.02	1.76	8.77	80.28	64.31	25.35
Dishwasher	0.13	1.44	3.01	88.37	84.44	46.69
FloorLamp	0.00	2.13	22.97	85.96	66.89	17.16
Fork	0.14	0.34	8.40	58.86	89.28	11.03
Hardware	0.01	0.62	8.45	89.87	83.96	23.56
Jar	0.03	0.76	1.39	79.12	69.67	41.51
Lamp	0.00	1.40	25.44	86.23	69.58	16.96
LargeShelf	0.02	0.82	5.15	88.08	60.70	16.81
Lid	0.05	1.83	2.39	73.22	63.47	50.11
LiteDoor	0.03	1.36	5.75	94.75	36.91	18.71
LouverDoor	0.07	1.40	16.17	89.46	37.43	20.94
Microwave	0.07	1.44	11.04	91.72	55.65	49.38
OfficeChair	0.03	1.44	2.63	78.41	55.65	46.91
PanelDoor	0.04	1.31	6.50	94.60	37.18	20.94
Plate	0.04	0.96	1.07	72.70	70.72	60.05
SidetableDesk	0.01	0.67	4.50	93.23	91.75	35.75
SimpleBookcase	0.03	1.78	2.89	92.14	65.14	33.79
SimpleDesk	0.01	2.12	25.39	88.68	93.80	45.79
Sofa	0.03	1.52	2.14	93.81	81.33	65.29
Spoon	0.67	0.37	4.09	74.00	87.04	18.92
TableCocktail	0.02	2.59	5.93	88.47	60.49	25.19
TableDining	0.02	5.52	1.03	88.14	58.43	71.26
Toilet	0.02	2.30	7.51	89.10	62.61	51.14
TriangleShelf	0.01	2.30	12.61	88.75	62.61	30.59
TV	0.04	1.53	3.41	87.80	72.69	34.14
TVStand	0.01	0.78	13.50	91.26	73.78	22.57
Vase	0.30	0.73	19.10	72.26	89.95	60.94
Window	0.14	0.59	3.73	87.36	84.21	64.64
Wineglass	0.06	0.98	6.83	88.36	73.96	28.56
All (Avg.)	0.06	1.87	6.00	86.75	67.62	45.03

A.4.1 Ablation Study

We conduct four ablation studies to evaluate the impact of key design choices in our framework.

Triplane Resolution and the Number of Learnable Tokens. The first ablation investigates the effect of varying the resolution of the triplane representation and the number of learnable tokens. As shown in Table 5, we observe that increasing both the triplane resolution and the number of learnable tokens consistently improves the performance of the object-to-code inference model. This suggests that a finer-grained spatial encoding and a richer set of token representations enable the model to better capture the underlying 3D structure of objects.

Table 3: Comparison of standard deviation of reconstruction metrics across all categories.

Category	CD			IoU		
	MeshCoder	PLAD	Shape2prog	MeshCoder	PLAD	Shape2prog
ArmChair	1.51×10^{-3}	9.35×10^{-3}	1.51×10^{-2}	4.62×10^{-2}	6.48×10^{-2}	5.28×10^{-2}
BarChair	1.82×10^{-4}	1.18×10^{-2}	9.90×10^{-3}	8.19×10^{-2}	1.00×10^{-1}	8.30×10^{-2}
Bathhtub	6.93×10^{-4}	1.02×10^{-2}	6.70×10^{-3}	1.31×10^{-1}	1.91×10^{-1}	8.57×10^{-2}
BeverageFridge	4.84×10^{-3}	2.81×10^{-3}	3.44×10^{-2}	1.13×10^{-1}	6.23×10^{-2}	6.64×10^{-2}
Bottle	7.56×10^{-5}	6.80×10^{-3}	6.00×10^{-2}	1.13×10^{-1}	7.05×10^{-2}	6.90×10^{-2}
Bowl	4.83×10^{-5}	5.13×10^{-3}	8.78×10^{-3}	8.11×10^{-2}	6.24×10^{-2}	2.35×10^{-2}
CeilingClassicLamp	7.33×10^{-7}	7.66×10^{-4}	9.86×10^{-4}	3.39×10^{-5}	2.96×10^{-3}	3.82×10^{-2}
CeilingLight	1.79×10^{-6}	3.90×10^{-3}	4.44×10^{-3}	3.10×10^{-2}	5.08×10^{-2}	6.93×10^{-2}
CellShelf	3.37×10^{-5}	1.94×10^{-2}	6.95×10^{-2}	9.65×10^{-2}	1.34×10^{-1}	9.45×10^{-2}
Lamp	2.20×10^{-5}	9.05×10^{-3}	2.74×10^{-1}	1.56×10^{-1}	6.87×10^{-2}	1.18×10^{-1}
Chair	1.09×10^{-3}	1.04×10^{-2}	4.52×10^{-3}	1.05×10^{-1}	9.17×10^{-2}	6.72×10^{-2}
Chopsticks	3.64×10^{-3}	1.31×10^{-2}	1.85×10^{-1}	1.87×10^{-1}	1.00×10^{-1}	1.01×10^{-1}
Cup	1.59×10^{-3}	5.79×10^{-3}	3.60×10^{-2}	9.84×10^{-2}	6.80×10^{-2}	6.98×10^{-2}
DeskLamp	7.62×10^{-4}	8.60×10^{-3}	4.55×10^{-2}	1.30×10^{-1}	7.21×10^{-2}	6.01×10^{-2}
Dishwasher	9.66×10^{-3}	2.69×10^{-3}	2.39×10^{-2}	1.27×10^{-1}	4.74×10^{-2}	8.82×10^{-2}
FloorLamp	1.23×10^{-4}	2.09×10^{-2}	2.54×10^{-1}	1.68×10^{-1}	4.92×10^{-2}	1.12×10^{-1}
Fork	8.81×10^{-3}	2.14×10^{-3}	8.57×10^{-2}	2.14×10^{-1}	1.25×10^{-1}	6.55×10^{-2}
Hardware	2.20×10^{-4}	3.07×10^{-3}	4.48×10^{-2}	1.21×10^{-1}	1.02×10^{-1}	1.34×10^{-1}
Jar	1.40×10^{-4}	2.44×10^{-3}	6.11×10^{-3}	1.44×10^{-1}	6.31×10^{-2}	8.98×10^{-2}
LargeShelf	1.79×10^{-4}	4.65×10^{-3}	5.12×10^{-2}	1.53×10^{-1}	8.67×10^{-2}	7.09×10^{-2}
Lid	8.89×10^{-4}	1.09×10^{-2}	1.95×10^{-2}	1.55×10^{-1}	1.22×10^{-1}	1.23×10^{-1}
LiteDoor	5.79×10^{-3}	4.39×10^{-3}	2.88×10^{-2}	1.44×10^{-1}	6.32×10^{-2}	9.69×10^{-2}
LouverDoor	4.67×10^{-3}	4.84×10^{-3}	9.23×10^{-2}	1.65×10^{-1}	6.82×10^{-2}	1.53×10^{-1}
Microwave	3.92×10^{-3}	2.43×10^{-2}	3.15×10^{-2}	7.26×10^{-2}	1.34×10^{-1}	1.65×10^{-1}
OfficeChair	1.72×10^{-4}	7.35×10^{-3}	2.95×10^{-2}	8.97×10^{-2}	1.06×10^{-1}	1.05×10^{-1}
PanelDoor	9.05×10^{-3}	4.79×10^{-3}	3.74×10^{-2}	1.50×10^{-1}	7.17×10^{-2}	1.09×10^{-1}
Plate	1.73×10^{-4}	6.40×10^{-3}	5.78×10^{-3}	1.70×10^{-1}	1.29×10^{-1}	1.74×10^{-1}
SidetableDesk	5.11×10^{-5}	3.52×10^{-3}	5.37×10^{-2}	9.64×10^{-2}	5.83×10^{-2}	1.23×10^{-1}
SimpleBookcase	3.66×10^{-3}	6.54×10^{-3}	7.01×10^{-3}	1.08×10^{-1}	9.62×10^{-2}	6.06×10^{-2}
SimpleDesk	4.29×10^{-5}	9.90×10^{-2}	1.72×10^{-1}	1.68×10^{-1}	6.43×10^{-2}	8.00×10^{-2}
Sofa	1.35×10^{-3}	5.78×10^{-3}	6.99×10^{-3}	6.61×10^{-2}	7.32×10^{-2}	6.37×10^{-2}
Spoon	5.64×10^{-2}	1.63×10^{-3}	4.59×10^{-2}	2.26×10^{-1}	8.26×10^{-2}	9.81×10^{-2}
TableCocktail	2.03×10^{-4}	2.85×10^{-2}	3.10×10^{-2}	1.09×10^{-1}	2.11×10^{-1}	8.68×10^{-2}
TableDining	3.31×10^{-3}	7.18×10^{-2}	6.16×10^{-3}	1.55×10^{-1}	1.64×10^{-1}	9.41×10^{-2}
Toilet	1.09×10^{-4}	8.44×10^{-3}	1.99×10^{-2}	4.22×10^{-2}	5.24×10^{-2}	5.41×10^{-2}
TriangleShelf	3.60×10^{-5}	9.30×10^{-3}	9.47×10^{-2}	1.03×10^{-1}	6.63×10^{-2}	1.08×10^{-1}
TV	6.25×10^{-4}	1.74×10^{-3}	1.02×10^{-2}	1.66×10^{-1}	2.84×10^{-2}	6.68×10^{-2}
TVStand	2.59×10^{-5}	1.45×10^{-3}	5.63×10^{-2}	1.31×10^{-1}	9.48×10^{-2}	5.92×10^{-2}
Vase	9.97×10^{-3}	3.44×10^{-3}	1.05×10^{-1}	2.68×10^{-1}	2.48×10^{-2}	4.00×10^{-2}
Window	9.57×10^{-3}	3.51×10^{-3}	6.61×10^{-2}	1.81×10^{-1}	1.14×10^{-1}	1.88×10^{-1}
Wineglass	1.40×10^{-2}	3.12×10^{-3}	3.86×10^{-2}	1.04×10^{-1}	5.39×10^{-2}	6.99×10^{-2}
All (Std.)	2.92×10^{-3}	2.49×10^{-2}	7.23×10^{-2}	1.25×10^{-1}	1.94×10^{-1}	1.92×10^{-1}

Initialization from Part-to-Code Checkpoint. The second ablation study evaluates whether initializing the object-to-code model with the pre-trained checkpoint of the part-to-code inference model yields performance improvements. Table 6 demonstrates that such initialization leads to noticeably better results. This improvement may be attributed to the part-to-code model’s ability to learn robust 3D geometric representations and syntactic grammar structures from the diverse part-level dataset. These learned features likely provide transferable knowledge that facilitates generalization during the object-level inference process, thereby improving the effectiveness of the model.

Using Learnable Tokens as Queries. The third ablation study explores the role of learnable tokens when used as queries for the LLM. In our default setup, learnable tokens are used as input queries to the LLM, whereas in the alternative setup, the triplane-encoded features are passed through an MLP and directly fed into the LLM. As reported in Table 7, the learnable-token-as-query strategy achieves superior performance. We hypothesize that this advantage arises for two reasons: (1) the learnable tokens are capable of aggregating global information across the entire input, unlike the direct feature approach where each patch predominantly captures localized information, and (2) the learnable tokens can adaptively organize the input representation in a layout that is more aligned with the LLM’s internal understanding and processing patterns.

Table 4: Quantitative evaluation of the *part-to-code* inference model across different part categories. CD is reported in 10^{-2} , and IoU is reported in percentage.

Category	CD ($\times 10^{-2}$)	IoU (%)
Primitive	0.18	94.81
Boolean	0.03	96.13
Array	0.70	78.90
Bridge Loop	0.14	89.16
Translation	0.17	83.45

Table 5: Ablation study on triplane resolution and the number of learnable tokens in MeshCoder. We report L2 Chamfer Distance ($\times 10^{-4}$) and IoU (%).

Triplane Resolution	Token Number	L2 CD ($\times 10^{-4}$)	IoU (%)
256	128	6.32	86.75
256	64	7.72	85.11
256	32	7.05	85.02
128	128	6.55	86.62
64	128	6.51	86.05
32	128	6.88	84.56

Ablate triplane. To investigate the effectiveness of different point cloud encoding methods, we conducted an ablation study focusing on the triplane encoding strategy. Specifically, we compared it with the point cloud encoding approach from Hunyuan3D [37]: each input point is treated as an independent token, and its information is directly transferred to learnable tokens via cross-attention, bypassing the triplane intermediate representation. The experimental results are presented in the table, which demonstrate that the tri-plane encoding method achieves superior performance on the target task.

A.5 Complete experiment result of Shape Editing

We additionally present two examples of shape editing along with their complete code implementations. In Figure 27, we modify the thickness of the chair legs and armrests by adjusting the `scale` parameter. In Figure 28, we change the mesh resolution of a plate by modifying the `resolution` parameter.

A.6 Complete experiment result of Shape Understanding

When presented with a 3D point cloud of an object as input, MeshCoder can infer the corresponding code for the object. Upon execution of this code in Blender, the geometry of the object can be obtained. Notably, the comments within the code encompass a variety of semantically rich cues, such as the object’s identity and the specifics of each component. The primary aim of this experiment is to highlight that our model can assist existing large language models, like GPT-4, in understanding the structure of 3D objects. We provide the inferred code to GPT-4 with the implementation details and functional descriptions of the functions used in the reconstruction code. Then, we input the object’s reconstruction code itself and then inquire about the geometry or structure of the object, as showed in Figure 19, Figure 20 and Figure 21. GPT - 4 is able to generate relevant responses based on the code. More importantly, we found that this code-based representation has unique advantages over visual inputs like multi-view images. For instance, code enables the LLM to understand complex internal structures and provide precise dimensional measurements, which are very difficult to ascertain from images alone. Our further experiments demonstrate that code and multi-view images are complementary; providing both to the LLM leads to a higher accuracy than using either input in isolation. This demonstrates that our model possesses capabilities in understanding the geometry and structure of 3D objects and can aid large - scale models such as GPT in addressing such questions. However, our model does have limitations. Currently, the code inferred by our model solely contains geometric information of the object and does not include the color or texture information that images provide. As a result, it is unable to answer questions pertaining to color.

Table 6: Ablation study on whether to initialize object-to-code model from the part-to-code checkpoint.

Initialization Strategy	L2 CD ($\times 10^{-4}$)	IoU (%)
From Scratch	8.62	85.16
From Part-to-Code Checkpoint	6.32	86.75

Table 7: Ablation study on whether to use learnable tokens as queries in the transformer.

Query Type	L2 CD ($\times 10^{-4}$)	IoU (%)
MLP Projection Only	9.88	84.12
Learnable Tokens (Ours)	6.32	86.75

A.7 More comparative experiments

To demonstrate our model’s ability to generalize beyond the Infinigen dataset, we have conducted reconstruction experiments using non-Infinigen data. Specifically, we leveraged Trellis[38], a currently popular generative model that takes single images as input and outputs corresponding meshes, to create our experimental data. We collected images of three categories (chairs, shelves, and bottles) from online sources, with 10 images collected for each category. These images were input into Trellis to generate meshes, from which we sampled point clouds to serve as input for MeshCoder. To further demonstrate that our model does not merely memorize the training data, we selected examples from our training dataset that have the smallest CD loss with the point clouds generated by Trellis. All metric results are presented in Table 9.

This experiment shows that our model maintains satisfactory performance on non-Infinigen data, demonstrating its generalization capability. Additionally, we argue that the model does not merely memorize Infinigen-specific structures. If it did, the metrics of our model’s predictions are unlikely to outperform those of the dataset matches, which represent the closest possible examples from the training set. However, we found that the reconstruction results are suboptimal when attempting to reconstruct objects from categories not present in the training dataset. To further strengthen generalization in future work, we plan to expand to more extensive and diverse datasets such as Objaverse[39], and incorporate reinforcement learning techniques (DPO[40], GRPO[41]) into our training pipeline, like DeepMesh[42], to boost the model’s ability to handle broader data distributions.

Moreover, to investigate how varying point cloud densities affect reconstruction quality, we designed experiments with three different point cloud quantities, where the number of input points for our trained MeshCoder was set to 4096, 8192, and 16384 respectively to observe its reconstruction metrics across four representative categories: bottles, cell shelves, chairs, and sofas. The metric results are shown in Table 10. Notably, we observe a clear trend: as the number of input points increases, the reconstruction quality consistently improves.

Table 8: Ablation study on triplane

Encoding approach Strategy	L2 CD ($\times 10^{-4}$)	IoU (%)
Treat point-cloud as token	8.96	85.17
Triplane	6.32	86.75

Table 9: Metrics of different methods on non-Infinigen data.

Method	Chairs		Shelves		Bottles	
	CD Loss (\downarrow)	IoU (\uparrow)	CD Loss (\downarrow)	IoU (\uparrow)	CD Loss (\downarrow)	IoU (\uparrow)
PLAD	0.09	0.13	0.12	0.16	0.21	0.22
Shape2Prog	0.03	0.46	0.06	0.06	0.25	0.40
Dataset Matches	0.25	0.11	0.08	0.17	0.08	0.41
MeshCoder	0.005	0.52	0.011	0.61	0.0015	0.78

Note: "Dataset Matches" represent examples from the training set with minimal CD loss to Trellis-generated point clouds.

Table 10: Reconstruction metrics under different point cloud quantities

Number of Points	CD Loss (\downarrow)	IoU (\uparrow)
4096	0.000404	0.8707
8192	0.000396	0.8718
16384	0.000325	0.8907

```

import bpy
from math import radians, pi
from bpy_lib import *

delete_all()

# object name: dishwasher
# part_1: body
create_primitive(...)

# part_2: body
create_primitive(...)

# part_3: body
create_primitive(...)

# part_4: door
create_primitive(...)

# part_5: body
create_primitive(...)

# part_6: quadrilateral rack
create_circle(...)
create_rectangle(...)
array_1d(...)

# part_7: quadrilateral rack
create_circle(...)
create_rectangle(...)
array_1d(...)

# part_8: rack
create_primitive(...)
array_2d(...)

# part_9: rack
create_primitive(...)
array_2d(...)

# part_10: rack
create_primitive(...)
array_3d(...)

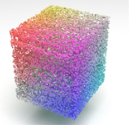
# part_11: handle
create_primitive(name='handle_11',
                 primitive_type='cube',
                 locations=[0.46, 0.39, -0.34],
                 scale=[0.02, 0.02, 0.02],
                 rotations=[0.0, -0.0, -0.71, 0.71])

# part_12: handle
create_primitive(name='handle_12',
                 primitive_type='cube',
                 locations=[0.49, 0.39, -0.0],
                 scale=[0.36, 0.02, 0.01],
                 rotations=[0.0, 0.7, -0.0, 0.71])

# part_13: handle
create_primitive(name='handle_13',
                 primitive_type='cube',
                 locations=[0.46, 0.39, 0.34],
                 scale=[0.02, 0.02, 0.02],
                 rotations=[0.0, -0.0, -0.71, 0.71])

# part_14: heater
create_primitive(...)

```



What is this object?

The object according to the comment at the beginning of the code, is a dishwasher

Does it have a rack inside?

Yes, the object has racks inside. The code contains multiple parts explicitly labeled as racks, including quadrilateral racks (frame_6, frame_7) and several cylinder racks (cylinder_8, cylinder_9, cylinder_10).

What is the shape of the handle?

The handle is composed of three cubes ('handle_11', 'handle_12', 'handle_13') forming a bar-like structure. 'handle_12' is longer and thinner compared to 'handle_11' and 'handle_13', suggesting a shape similar to a horizontal bar with two support ends.

Can you summarize the shape and component parts of this object?

The object is a dishwasher, consisting of a main body formed by several cube primitives ('body_1', 'body_2', 'body_3', 'body_5', and 'door_4'). The dish racks inside are created using rectangles and cylinders, arranged in a grid using array modifiers for creating multiple rack elements. These racks include quadrilateral frames (frame_6, frame_7) and cylinder rods (cylinder_8, cylinder_9, cylinder_10) which typically hold the dishes. The handle is composed of three cubes forming a bar, and there is also a heater ('heater_14') modeled similarly as part of the main body. Overall, the object is structurally coherent with typical dishwasher components, such as a door, racks for dishes, a handle for opening, and a heater for drying.

Figure 19: Experiments on how GPT-4o can understand shape through given dishwasher code.

```

import bpy
from math import radians, pi
from bpy_lib import *

delete_all()

# object name: office chair
# part_1: wheel
create_primitive(name='wheel_1',
primitive_type='cylinder',
location=...)
bevel(name='wheel_1', ...)

# part_2: wheel
create_primitive(...)

# part_3: wheel
create_primitive(...)
bevel(...)

# part_4: wheel
create_primitive(...)

# part_5: wheel cap
create_rectangle(...)
create_arc_by_3Dpoints(...)

# part_6: wheel axle
create_primitive(...)

# part_7: wheel cap
create_rectangle(...)
create_arc_by_3Dpoints(...)

# part_8: wheel axle
create_primitive(...)

# part_9: wheel axle
create_primitive(...)

# part_10: wheel cap
create_rectangle(...)
create_arc_by_3Dpoints(...)

# part_11: wheel cap
create_rectangle(...)
create_arc_by_3Dpoints(...)

# part_12: wheel axle
create_primitive(...)

# part_13: chair base
create_curve(...)
bevel(...)
create_curve(...)

# part_14: chair base
create_curve(name='curve_14',
control_points=[[0.0, 0.0, 0.0],
[0.0, -0.025, 0.0], [0.044, -
0.025, 0.0], [0.044, 0.0, 0.0]]
handle_type=[1, 1, 1, 1, 1,
1, 1], closed=True)
bevel(...)
create_curve(...)

# part_15: chair base
create_curve(name='curve_15',
control_points=[[0.0, 0.0, 0.0],
[-0.0, 0.025, 0.0], [-0.044,
0.025, 0.0], [-0.044, -0.0,
0.0]], handle_type=[1, 1, 1, 1,
1, 1, 1], closed=True)
bevel(...)
create_curve(...)

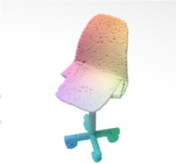
# part_16: chair base
create_curve(name='curve_16',
control_points=[[0.0, 0.0, 0.0],
[0.0, 0.025, 0.0], [-0.044,
0.025, 0.0], [-0.044, 0.0,
0.0]], handle_type=[1, 1, 1, 1,
1, 1, 1], closed=True)
bevel(...)
create_curve(...)

# part_17: Leg
create_primitive(...)

# part_18: Leg
create_primitive(...)

# part_19: seat
create_circle(name=['Circle_0_1
9', 'Circle_1_19',
'Circle_2_19', 'Circle_3_19',
'Circle_4_19', 'Circle_5_19',
'Circle_6_19', 'Circle_7_19',
'Circle_8_19'], location=[...])
bridge_edge_loops(name=...)
join_obj(name='seat_19',
seq_name=...,
weld_threshold=...)
add_simple_deform_modifier(name
='seat_19', angle=...,
origin=..., rotation=...)
create_curve(name='curve_19',
control_points=...,
points_radius=...,
handle_type=...)
add_curve_modifier_to_object(na
me='seat_19',
curve_name='curve_19',
origin=..., rotation=...,
axis='POS_Z')

```



What's the shape of the leg?

The leg parts ('part_17' and 'part_18') are created using the 'create_primitive' function with the primitive type 'cylinder'. This means the legs are cylindrical.

Does this chair have wheels? If so, how many?

The chair has wheels. They are labeled as 'part_1', 'part_2', 'part_3', and 'part_4'. Each wheel is created using the 'create_primitive' function with the primitive type 'cylinder'. There are **4 wheels**. **Answer: Yes**

Are there any stretchers between the legs?

Stretchers are structural components that connect parts of furniture (like legs of a chair). In the code, there is no mention of stretchers explicitly connecting the legs. Although parts such as 'chair base_13', 'chair base_14', 'chair base_15', and 'chair base_16' could serve as a connecting structure near the base of the chair, there is no direct indication that these are stretchers. Thus, we cannot explicitly state that stretchers are present. **Answer: No**

Figure 20: Experiments on how GPT-4o can understand shape through given office chair code.

```

import bpy
from math import radians, pi
from bpy.lib import *

delete_all()

# object name: cell shelf
# part_1: side board
create_primitive(name='side
board_1', primitive_type='cube',
location=[0.0, 0.0, -0.48],
scale=[0.46, 0.22, 0.02],
rotation=[0.0, -0.71, -0.71, -
0.0])

# part_2: bottom board
create_primitive(name='bottom
board_2', primitive_type='cube',
location=[0.0, -0.48, -0.0],
scale=[0.5, 0.22, 0.02],
rotation=[0.5, 0.5, -0.5, 0.5])

# part_3: vertical division
board
create_primitive(name='vertical
division board_3',
primitive_type='cube',
location=[0.0, 0.0, 0.0],
scale=[0.46, 0.22, 0.005],
rotation=[0.0, -0.71, 0.71,
0.0])

# part_4: side board
create_primitive(name='side
board_4', primitive_type='cube',
location=[0.0, 0.0, 0.48],
scale=[0.46, 0.22, 0.02],
rotation=[0.0, -0.71, -0.71, -
0.0])

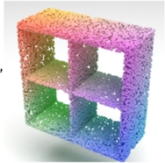
# part_5: horizontal division
board
create_primitive(name='horizont
al division board_5',
primitive_type='cube',
location=[0.0, 0.0, 0.0],
scale=[0.46, 0.22, 0.005],
rotation=[0.5, -0.5, -0.5, -
0.5])

# part_6: attach gadget
create_primitive(name='attach
gadget_6',
primitive_type='cube',
location=[-0.22, 0.43, -0.45],
scale=[0.01, 0.01, 0.001],
rotation=[0.5, 0.5, 0.5, 0.5])

# part_7: attach gadget
create_primitive(name='attach
gadget_7',
primitive_type='cube',
location=[-0.22, 0.43, 0.45],
scale=[0.01, 0.01, 0.001],
rotation=[0.5, 0.5, 0.5, 0.5])

# part_8: top board
create_primitive(name='top
board_8', primitive_type='cube',
location=[0.0, 0.48, -0.0],
scale=[0.5, 0.22, 0.02],
rotation=[0.5, -0.5, 0.5, 0.5])

```



What is this object?

The object is described as a **cell shelf**, which is likely a compact shelving unit designed to store items in individual compartments. The "cell" terminology suggests the presence of multiple internal divisions to create smaller storage sections.

What is the shape of the cell shelf?

The shape of the cell shelf is a rectangular cabinet with multiple compartments. These compartments are created by the following components from the code: Two **side boards** (part 1 and part 4) form the left and right walls of the shelf. **Bottom board** (part 2) is the lower horizontal section. **Top board** (part 8) is the upper horizontal section. **Vertical division board** (part 3) divides the shelf vertically into multiple compartments. **Horizontal division board** (part 5) adds horizontal subdivisions within the interior. **Attach gadgets** (part 6 and part 7) seem to be small elements for connecting or securing the boards together. The overall structure of the shelf creates compartments (or "cells") arranged horizontally and vertically.

What is its functionality?

The **cell shelf** has the functionality of providing organized storage space. Its features include: **Multi-compartment storage**: Created by vertical and horizontal division boards (parts 3 and 5), enabling organized storage within individual cells. **Attachment stability**: The **attach gadgets** (part 6 and part 7) appear to contribute to the structural integrity or assembly of the cell shelf, ensuring the divisions and boards remain securely connected. **Compact Design**: The size and scale of the components suggest a compact shelf suitable for small objects, potentially useful in retail displays, personal storage systems, or as a modular organizer. The targeted functionality of the shelf is likely for **organizing and storing small items** efficiently in a structured system of compartments.

Figure 21: Experiments on how GPT-4o can understand shape through given cell shelf code.


```

"""
Create primitive object
"""
def create_primitive(name, primitive_type="cube", location=None, scale=None, rotation=None, rotation_mode='QUATERNION', apply=False,
x_subdivisions=None, y_subdivisions=None, use_minimum_face=USE_MINIMUM_FACE, average_edge_length=AVERAGE_EDGE_LENGTH, resolution=None):

    if primitive_type=="uv_sphere":
        if not use_minimum_face:
            if average_edge_length !=None:
                res=int(2*pi//average_edge_length)
                segments, ring_count=res, res
            else:
                segments, ring_count=resolution[0], resolution[1]
            getattr(bpy.ops.mesh, f"primitive_{primitive_type}_add")(segments=segments,ring_count=ring_count)
            primitive = bpy.context.object
            primitive.name = name
        else:
            getattr(bpy.ops.mesh, f"primitive_{primitive_type}_add")()
            primitive = bpy.context.object
            primitive.name = name

    if primitive_type in ["cylinder","cone"]:
        if not use_minimum_face:
            if average_edge_length !=None:
                res_1=int(2*pi//average_edge_length)
                vertices=res_1
                res_2=int(2//average_edge_length)
            else:
                vertices,res_2=resolution[0], resolution[1]
            getattr(bpy.ops.mesh, f"primitive_{primitive_type}_add")(vertices=vertices)
            primitive = bpy.context.object
            primitive.name = name
            if primitive_type=="cylinder":
                primitive=subdivide_primitive(name,[res_2],['Z'])
            else:
                primitive=split_cone_z(name,res_2)
        else:
            getattr(bpy.ops.mesh, f"primitive_{primitive_type}_add")()
            primitive = bpy.context.object
            primitive.name = name

    if primitive_type=="torus":
        if not use_minimum_face:
            if average_edge_length !=None:
                res_1=int(2*pi//average_edge_length)
                res_2=int(0.5*pi//average_edge_length)
            else:
                res_1, res_2=resolution[0], resolution[1]
            getattr(bpy.ops.mesh, f"primitive_{primitive_type}_add")(major_segments=res_1,minor_segments=res_2)
            primitive = bpy.context.object
            primitive.name = name
        else:
            getattr(bpy.ops.mesh, f"primitive_{primitive_type}_add")()
            primitive = bpy.context.object
            primitive.name = name

    if primitive_type=="cube":
        if not use_minimum_face:
            if average_edge_length !=None:
                res=int(2//average_edge_length)
                if res>1:
                    resolution=[res,res,res]
            else:
                pass
            getattr(bpy.ops.mesh, f"primitive_{primitive_type}_add")()
            primitive = bpy.context.object
            primitive.name = name
            primitive=subdivide_primitive(name,resolution,['X','Y','Z'])
        else:
            getattr(bpy.ops.mesh, f"primitive_{primitive_type}_add")()
            primitive = bpy.context.object
            primitive.name = name

    if primitive_type=="grid":
        getattr(bpy.ops.mesh, f"primitive_{primitive_type}_add")(x_subdivisions=x_subdivisions,y_subdivisions=y_subdivisions)
        primitive = bpy.context.object
        primitive.name = name

    if location:
        primitive.location = location
    if scale:
        primitive.scale = scale
    if rotation:
        if rotation_mode=='XYZ':
            primitive.rotation_euler = [angle * pi for angle in rotation]
        elif rotation_mode=='QUATERNION':
            primitive.rotation_mode = 'QUATERNION'
            primitive.rotation_quaternion = rotation
        elif rotation_mode=='MATRIX':
            mat = np.eye(4)
            rotation = np.array(rotation).reshape([3,3])
            mat[:3,:3] = rotation
            bpy.context.view_layer.update()
            world_matrix = torch.tensor(bpy.data.objects[name].matrix_world)
            scale_now = world_matrix.norm(dim=0)[:3]
            scale_matrix = torch.eye(4)
            scale_matrix[0,0],scale_matrix[1,1],scale_matrix[2,2] = scale_now[0],scale_now[1],scale_now[2]
            scale_matrix_inv = scale_matrix.clone()
            for i in range(3):
                if scale_matrix_inv[i,i]>1e-10:
                    scale_matrix_inv[i,i]=1.0 / scale_matrix_inv[i,i]
            mat = scale_matrix_inv@torch.tensor(mat, dtype=torch.float32)@scale_matrix
            mat = mathutils.Matrix(np.array(mat))
            bpy.data.objects[name].matrix_world = bpy.data.objects[name].matrix_world@mat

    if apply:
        bpy.ops.object.transform_apply(location=True, rotation=True, scale=True)

    return primitive

```

Figure 22: Implementation of the function for creating primitives

```

"""
Creates a translational object of a line trajectory
"""
def create_curve(name, profile_name=None, control_points=[], points_radius=[], handle_type=[], closed=False, center="POINT", thickness=None,
fill_caps="None", flip_normals=False, bevel_width=None, bevel_segments=8, use_minimum_face=USE_MINIMUM_FACE, average_edge_length=AVERAGE_EDGE_LENGTH, resolution=24,
volumm_origin=True):
    if isinstance(name, str):
        type_dict={0:"AUTO", 1:"VECTOR", 2:"ALIGNED", 3:"FREE"}

        control_points = np.array(control_points).tolist()
        control_points_tmp = copy.deepcopy(control_points)
        num_handle_co = handle_type.count(3)
        num_control_points = len(control_points) - num_handle_co

        curveData = bpy.data.curves.new(name, type='CURVE')
        curveData.dimensions = '3D'

        bezierSpline = curveData.splines.new('BEZIER')
        bezierSpline.bezier_points.add(num_control_points - 1)
        bezierSpline.use_cyclic_u = closed

    for i in range(num_control_points):
        bezier_point = bezierSpline.bezier_points[i]
        bezier_point.handle_left_type = type_dict[handle_type[2*i]]
        if type_dict[handle_type[2*i]]=="FREE":
            bezier_point.handle_left = control_points.pop(0)
        bezier_point.co = control_points.pop(0)
        bezier_point.handle_right_type = type_dict[handle_type[2*i+1]]
        if type_dict[handle_type[2*i+1]]=="FREE":
            bezier_point.handle_right = control_points.pop(0)
        bezier_point.radius = points_radius[i] if len(points_radius)!=0 else 1.0

    assert len(control_points)==0, "cannot create curve"
    if use_minimum_face:
        use_resolution = 12
    elif not average_edge_length is None:
        for i in range(len(BezierSpline.bezier_points) - 1):
            p1 = BezierSpline.bezier_points[i].co
            p2 = BezierSpline.bezier_points[i + 1].co
            total_length += (p2 - p1).length
        use_resolution = total_length/average_edge_length

    if resolution:
        use_resolution = resolution
    curveData.resolution_u = use_resolution

    curveOB = bpy.data.objects.new(name, curveData)

    if profile_name != None:
        curveData.bevel_mode = "OBJECT"
        curveData.splines[0].use_smooth = False
        scn = bpy.context.scene.collection
        scn.objects.link(curveOB)

        if bevel_width!=None:
            bevel(name=name, width=bevel_width, segments=bevel_segments)
            curveData = bpy.data.objects[name].data
            curveData.bevel_mode = 'OBJECT'

        curveData.bevel_object = bpy.data.objects[profile_name]
        if fill_caps=="both":
            curveData.use_fill_caps = True
        else:
            curveData.use_fill_caps = False

        if use_minimum_face:
            use_resolution = 24
        elif not average_edge_length is None:
            for i in range(len(BezierSpline.bezier_points) - 1):
                p1 = BezierSpline.bezier_points[i].co
                p2 = BezierSpline.bezier_points[i + 1].co
                total_length += (p2 - p1).length
            use_resolution = total_length/average_edge_length

        if resolution:
            use_resolution = resolution

        curveData.resolution_u = use_resolution

        bpy.context.view_layer.objects.active = bpy.data.objects[name]
        bpy.ops.object.mode_set(mode = 'OBJECT')
        bpy.data.objects[name].select_set(True)
        bpy.ops.object.convert(target='MESH')
        bpy.data.objects.remove(bpy.data.objects[profile_name], do_unlink=True)
        if volumm_origin:
            bpy.ops.object.origin_set(type='ORIGIN_CENTER_OF_VOLUME', center='MEDIAN')

        if fill_caps in ["start", "end"]:
            make_caps(name, fill_caps)

        if flip_normals:
            recalculate_normals(name, inside=True)
        else:
            recalculate_normals(name, inside=False)

        if thickness>1e-10:
            solidify(name, thickness)

        weld(name, 1e-5)

        return curveOB
    else:
        scn = bpy.context.scene.collection
        scn.objects.link(curveOB)
        if center=="MEDIAN":
            bpy.data.objects[name].select_set(True)
            bpy.ops.object.origin_set(type='ORIGIN_GEOMETRY', center='MEDIAN')
            points=np.array(control_points_tmp)
            return {"name":name, "points":points, "handle_type":handle_type, "closed":closed, "center":center}

    elif profile_name==None:
        if isinstance(profile_name, str):
            profile_name = [profile_name]*len(name)
        if len(points_radius) != 0 and (isinstance(points_radius[0], float) or isinstance(points_radius[0], int)):
            points_radius = [points_radius]*len(name)
        elif len(points_radius) == 0:
            points_radius = [[]] * len(name)
        if isinstance(handle_type[0], int):
            handle_type = [handle_type]*len(name)
        if isinstance(closed, bool):
            closed = [closed]*len(name)
        if isinstance(center, str):
            center = [center]*len(name)
        for i in range(len(name)):
            create_curve(name=name[i], control_points=control_points[i], points_radius=points_radius[i], handle_type=handle_type[i], closed=closed[i],
            center=center[i], thickness=thickness, fill_caps=fill_caps, flip_normals=flip_normals, resolution=resolution)
        points = np.array(copy.deepcopy(control_points))
        return {"name":name, "points":points, "handle_type":handle_type, "closed":closed, "center":center}

```

Figure 23: Implementation of the function for creating curves



Figure 24: An example of sofa. The input is a point cloud of a sofa, and the figure shows the code inferred by the object-to-code inference model, as well as the resulting mesh generated by executing the inferred code.



Figure 25: An example of bathtub. The input is a point cloud of a bathtub, and the figure shows the code inferred by the object-to-code inference model, as well as the resulting mesh generated by executing the inferred code.



```
import bpy
from math import radians, pi
from bpy_lib import *

delete_all()

# object name: chair
# part.1: leg
create_primitive(name='leg_1', primitive_type='cube', location=[-0.44, -0.46, 0.37], scale=[0.53, 0.05, 0.05], rotation=[0.5, 0.51, -0.51, -0.49])
bevel(name='leg_1', width=0.12, segments=8)

# part.2: leg
create_primitive(name='leg_2', primitive_type='cube', location=[-0.31, -0.46, -0.46], scale=[0.53, 0.05, 0.05], rotation=[0.51, -0.5, -0.49, 0.51])

# part.3: leg
create_primitive(name='leg_3', primitive_type='cube', location=[0.31, -0.46, -0.46], scale=[0.53, 0.05, 0.05], rotation=[0.51, -0.5, -0.49, 0.51])

# part.4: leg
create_primitive(name='leg_4', primitive_type='cube', location=[0.44, -0.46, 0.37], scale=[0.53, 0.05, 0.05], rotation=[0.5, 0.51, -0.51, -0.49])

# part.5: leg decoration
create_primitive(name='leg decoration_5', primitive_type='cube', location=[-0.37, -0.35, -0.05], scale=[0.42, 0.05, 0.05], rotation=[0.76, 0.01, 0.65, -0.01])
bevel(name='leg decoration_5', width=0.13, segments=1)

# part.6: leg decoration
create_primitive(name='leg decoration_6', primitive_type='cube', location=[0.37, -0.35, -0.05], scale=[0.42, 0.05, 0.05], rotation=[0.65, -0.01, 0.76, 0.01])
bevel(name='leg decoration_6', width=0.13, segments=8)

# part.7: seat
create_curve(name='seat_7', control_points=[[0, 0, 0.03, -0.51], [-0.35, 0.03, -0.51], [-0.47, 0.05, 0.21], [-0.49, 0.03, 0.41], [0, 0, 0.03, 0.51], [0.49, 0.03, 0.41], [0.47, 0.05, 0.21], [0.35, 0.03, -0.51], [0, 0, 0.03, -0.51]], handle_type=[0, 0, 0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 1.0, 0.0, 0.0])
fill_grid(name='seat_7', thickness=0.1042)
bevel(name='seat_7', width=0.05, segments=1)

# part.8: arm
create_circle(name='circle_8', radius=0.08, center='MEDIAN')
create_curve(name='arm_8', profile_name='circle_8', control_points=[[0.33, 0.64, -0.46], [0.39, 0.64, -0.21], [-0.46, 1.01, -0.06], [-0.46, 0.64, 0.21], [-0.46, 0.1, 0.31]], points_radius=[1.0, 1.0, 1.0], handle_type=[0, 3, 3, 1, 1, 0], thickness=0.003, fill_caps='both')

# part.9: back
create_primitive(name='back_9', primitive_type='cube', location=[-0.31, 0.54, -0.46], scale=[0.45, 0.04, 0.04], rotation=[0.5, 0.51, -0.51, -0.49])

# part.10: back
create_primitive(name='back_10', primitive_type='cube', location=[0.31, 0.54, -0.46], scale=[0.45, 0.04, 0.04], rotation=[0.5, 0.51, 0.51, 0.49])

# part.11: arm
create_circle(name='circle_11', radius=0.08, center='MEDIAN')
create_curve(name='arm_11', profile_name='circle_11', control_points=[[0.33, 0.64, -0.46], [0.39, 0.64, -0.21], [0.46, 1.01, -0.06], [0.47, 0.62, 0.23], [0.47, 0.09, 0.32]], points_radius=[1.0, 1.0, 1.0], handle_type=[0, 3, 3, 1, 1, 0], thickness=0.003, fill_caps='both')

# part.12: back decoration
create_curve(name='back decoration_12', control_points=[[-0.35, 0.8, -0.51], [-0.35, 0.62, -0.51], [0.01, 0.62, -0.51], [0.35, 0.62, -0.51], [0.35, 0.8, -0.51], [0.35, 0.99, -0.51], [0.01, 0.99, -0.51], [-0.35, 0.99, -0.51], [-0.35, 0.8, -0.51]], handle_type=[0, 0, 0, 1.0, 1.0, 0.0, 0.0, 1.0, 1.0, 0.0, 0.0, 1.0, 1.0, 0.0, 0.0, 1.0, 1.0, 0.0, 0.0, 0.0])
fill_grid(name='back decoration_12', thickness=0.086)
bevel(name='back decoration_12', width=0.04, segments=10)
```



```
import bpy
from math import radians, pi
from bpy_lib import *

delete_all()

# object name: chair
# part.1: leg
create_primitive(name='leg_1', primitive_type='cube', location=[-0.44, -0.46, 0.37], scale=[0.53, 0.05, 0.05], rotation=[0.5, 0.51, -0.51, -0.49])
bevel(name='leg_1', width=0.12, segments=8)

# part.2: leg
create_primitive(name='leg_2', primitive_type='cube', location=[-0.31, -0.46, -0.46], scale=[0.53, 0.05, 0.05], rotation=[0.51, -0.5, -0.49, 0.51])

# part.3: leg
create_primitive(name='leg_3', primitive_type='cube', location=[0.31, -0.46, -0.46], scale=[0.53, 0.05, 0.05], rotation=[0.51, -0.5, -0.49, 0.51])

# part.4: leg
create_primitive(name='leg_4', primitive_type='cube', location=[0.44, -0.46, 0.37], scale=[0.53, 0.05, 0.05], rotation=[0.5, 0.51, -0.51, -0.49])

# part.5: leg decoration
create_primitive(name='leg decoration_5', primitive_type='cube', location=[-0.37, -0.35, -0.05], scale=[0.42, 0.05, 0.05], rotation=[0.76, 0.01, 0.65, -0.01])
bevel(name='leg decoration_5', width=0.13, segments=1)

# part.6: leg decoration
create_primitive(name='leg decoration_6', primitive_type='cube', location=[0.37, -0.35, -0.05], scale=[0.42, 0.05, 0.05], rotation=[0.65, -0.01, 0.76, 0.01])
bevel(name='leg decoration_6', width=0.13, segments=8)

# part.7: seat
create_curve(name='seat_7', control_points=[[0, 0, 0.03, -0.51], [-0.35, 0.03, -0.51], [-0.47, 0.05, 0.21], [-0.49, 0.03, 0.41], [0, 0, 0.03, 0.51], [0.49, 0.03, 0.41], [0.47, 0.05, 0.21], [0.35, 0.03, -0.51], [0, 0, 0.03, -0.51]], handle_type=[0, 0, 0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 1.0, 0.0, 0.0])
fill_grid(name='seat_7', thickness=0.1042)
bevel(name='seat_7', width=0.05, segments=1)

# part.8: arm
create_circle(name='circle_8', radius=0.08, center='MEDIAN')
create_curve(name='arm_8', profile_name='circle_8', control_points=[[0.33, 0.64, -0.46], [0.39, 0.64, -0.21], [-0.46, 1.01, -0.06], [-0.46, 0.64, 0.21], [-0.46, 0.1, 0.31]], points_radius=[1.0, 1.0, 1.0], handle_type=[0, 3, 3, 1, 1, 0], thickness=0.003, fill_caps='both')

# part.9: back
create_primitive(name='back_9', primitive_type='cube', location=[-0.31, 0.54, -0.46], scale=[0.45, 0.04, 0.04], rotation=[0.5, 0.51, -0.51, -0.49])

# part.10: back
create_primitive(name='back_10', primitive_type='cube', location=[0.31, 0.54, -0.46], scale=[0.45, 0.04, 0.04], rotation=[0.5, 0.51, 0.51, 0.49])

# part.11: arm
create_circle(name='circle_11', radius=0.08, center='MEDIAN')
create_curve(name='arm_11', profile_name='circle_11', control_points=[[0.33, 0.64, -0.46], [0.39, 0.64, -0.21], [0.46, 1.01, -0.06], [0.47, 0.62, 0.23], [0.47, 0.09, 0.32]], points_radius=[1.0, 1.0, 1.0], handle_type=[0, 3, 3, 1, 1, 0], thickness=0.003, fill_caps='both')

# part.12: back decoration
create_curve(name='back decoration_12', control_points=[[-0.35, 0.8, -0.51], [-0.35, 0.62, -0.51], [0.01, 0.62, -0.51], [0.35, 0.62, -0.51], [0.35, 0.8, -0.51], [0.35, 0.99, -0.51], [0.01, 0.99, -0.51], [-0.35, 0.99, -0.51], [-0.35, 0.8, -0.51]], handle_type=[0, 0, 0, 1.0, 1.0, 0.0, 0.0, 1.0, 1.0, 0.0, 0.0, 1.0, 1.0, 0.0, 0.0, 1.0, 1.0, 0.0, 0.0, 0.0])
fill_grid(name='back decoration_12', thickness=0.086)
bevel(name='back decoration_12', width=0.04, segments=10)
```

Figure 27: By modifying the scale parameters of the leg and arm parts, we adjust their thickness. The highlighted sections indicate the changes made.

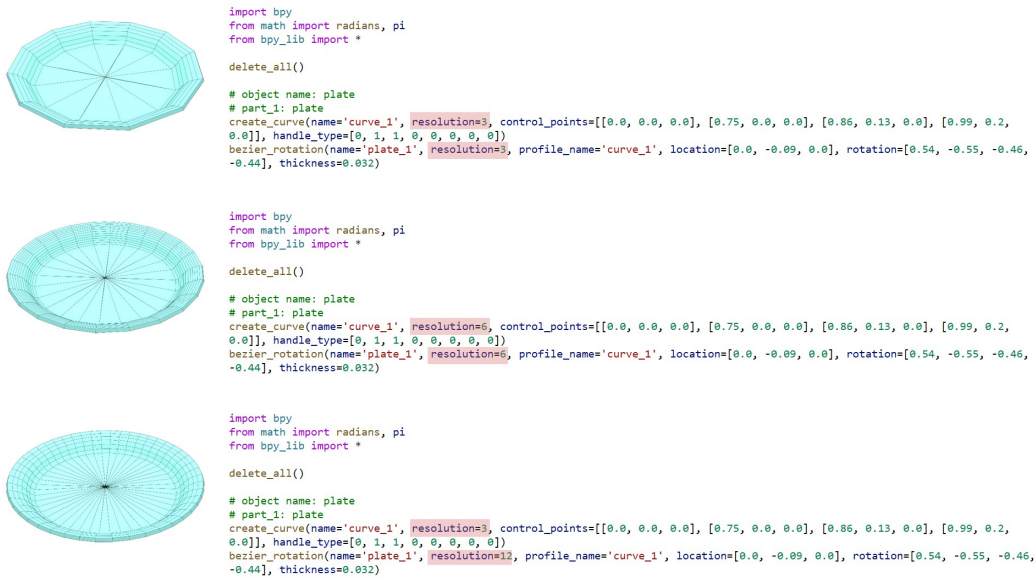


Figure 28: By modifying the resolution parameter, we change its resolution. The highlighted sections indicate the changes made.