

Combinatorial 3D Shape Generation via Sequential Assembly

Jungtaek Kim, Hyunsoo Chung, Minsu Cho, and Jaesik Park

POSTECH, Pohang 37673, Republic of Korea
`{jtkim,hschung2,mscho,jaesik.park}@postech.ac.kr`

Abstract. 3D shape generation has drawn attention in computer vision and machine learning since it opens an inspiring way to designing or creating new objects. Existing methods, however, do not reflect an important aspect of human generation processes in real life – we often create a 3D shape by sequentially assembling geometric primitives into a combinatorial configuration. In this work, we propose a new 3D shape generation algorithm that aims to create such a combinatorial configuration from a set of volumetric primitives. To tackle the exponential growth of feasible combinations in terms of the number of primitives, we adopt sequential model-based optimization. Our method sequentially assembles primitives by exploiting and exploring adequate regions that are constrained by the current primitive placements. The evaluation function conveys global structure guidance for the assembling process to follow. Experimental results demonstrate that our method successfully generates combinatorial objects and simulates more realistic generation processes. We also introduce a new dataset for combinatorial 3D shape generation.

Keywords: 3D shape generation, Combinatorial generation, Unit primitive combination, Sequential assembly, Combinatorial 3D shape dataset

1 Introduction

Generating a 3D shape gains popularity since it can relieve a significant amount of human effort in designing new volumetric objects, such as cars, furniture, or buildings, for real and/or virtual worlds. For 3D shape generation, generative models such as variational auto-encoders [24] and generative adversarial networks [13] have been adopted to construct a new shape using a set of primitives. Previous methods typically use a fixed number of points [1], a deformable mesh [12,14], or a voxel grid [40,41]. These generation schemes, however, do not reflect an important aspect of human generation processes in real life – we often create a 3D shape by sequentially assembling geometric primitives into a combinatorial configuration.

In this work, we thus attempt to solve the problem with a new framework of *combinatorial 3D shape* generation that aims to create a 3D object by assembling a set of *geometric primitives*. The main challenge of combinatorial generation lies in combinatorial explosion as the number of primitives increasing. For instance,

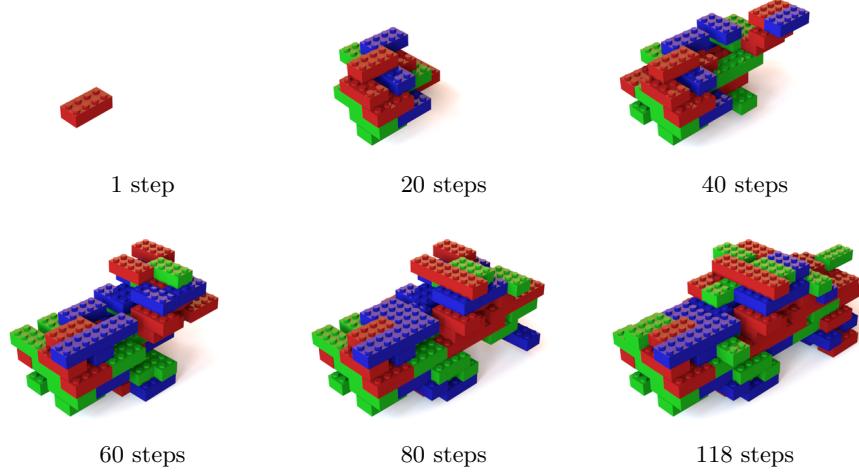


Fig. 1. Car object generated with unit geometric primitives using our method. Our generation process creates a new assembling sequence to build a 3D shape with given primitive budget.

if we assemble 2×4 LEGO bricks in a free 3D space, the most naïve way is to take all combinations into account and pick the most probable one for the purpose. However, with only six 2×4 LEGO bricks, the number of candidates amounts to 915 million combinations [8].¹ Due to the prohibitive number of possible combinations, a brute-force approach to combinatorial generation is to find a needle in a haystack. To tackle this challenge, we propose a sequential assembling method that iteratively evaluates the next possible primitive combinations in a sample-efficient manner.

Difficulties in realizing the combinatorial 3D shape assembly are two-fold: (i) how to evaluate a combinatorial 3D shape in the middle of the assembling procedure; (ii) how to determine the position to place a new primitive. To resolve the former issue, we present an evaluation method that compares the generated assemble with global shape guidance. For solving the latter issue, we use Bayesian optimization [6,29], which can significantly reduce the number of samples we need to observe. From the perspective that our task is a score maximization problem, Bayesian optimization guarantees to determine the optimal location using our predefined evaluation functions. With the proposed pipeline, our approach can assemble the primitives on the open space by gathering the combinations of any size, given a desired class of the generated shape.

¹ Based on the assembling conditions suggested in this paper, there exist 46 combinations for two bricks, 3,566 combinations for three bricks, 405,716 combinations for four bricks, and 59,814,648 combinations for five bricks.

In addition to the proposed pipeline, we introduce a new combinatorial 3D shape dataset that consists of 14 classes and 406 instances generated by 12 human subjects. Due to the nature of the combinatorial object, the dataset can be readily augmented by manipulating assembling orders. We hope the new dataset opens a new benchmark for shape combination.

Our contributions are summarized as follows:

- (i) Introducing the first approach that generates an assembling sequence for a new 3D shape;
- (ii) Proposing a Bayesian optimization-based method to decide the next primitive position to assemble efficiently;
- (iii) Creating a new dataset that can open a new benchmark for the task of combinatorial shape generation.

2 Related Work

Layout Optimization. Finding an optimal layout using geometric primitives from known 3D models under some constraints is a traditional research topic in various fields, such as computer vision, machine learning, and topology optimization [3,5,26,37,38]. The approach by Testuz *et al.* [37] identifies a suitable primitive set for a given mesh and applies a greedy method to repair weak connections. Lee *et al.* [26] propose to optimize a primitive layout using a genetic algorithm. Luo *et al.* [28] consider the physical stability of a constructed model, which helps to create realistic and realizable assembly accomplishments. Hengel *et al.* [38] create a collection of primitives from real-world images and silhouettes. Kozaki *et al.* [25] construct a combination from multiple 2D images using stochastic global optimization.

Topology Optimization. Topology optimization [3] is to find an optimal layout where predefined configurations and constraints are provided. It has widely been used in shape design, prototyping, and manufacturing. Eschenauer *et al.* [9] introduce a method that inserts holes into a component with iterative positioning. Borrvall *et al.* [4] use a topology optimization technique in fluid mechanics to solve applications such as pipe and airfoil designs. Kharmanda *et al.* [21] suggest a method to find reliable and efficient structures with the reliability index. Brackett *et al.* [5] utilize topology optimization in additive manufacturing for producing end-use parts.

Generative Models for 3D Objects. Generative models can be categorized into three classes according to the underlying 3D representations: (i) 3D points, (ii) a deformable mesh, and (iii) an occupancy grid. Fan *et al.* [10] propose an approach to generating a set of 3D points given a single 2D image. Achlioptas *et al.* [1] learn representations from point clouds via autoencoder (AE). Their approach employs either raw point sets or learned representations, to train a generative adversarial network (GAN). For the deformable mesh generation, Groueix *et al.* [14] suggest a method for transforming 2D texture map atlases into 3D surface. Gao *et al.* [12] generate structured deformable meshes. The

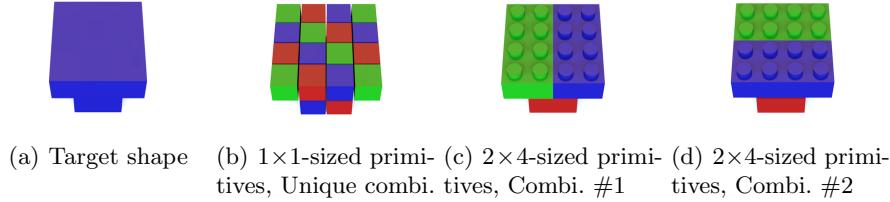


Fig. 2. Results using 1×1-sized primitives and 2×4-sized primitives. When we assemble a target shape, with 1×1 ones, a single unique combination can be assembled, but with 2×4-sized primitives two combinations can be assembled.

network is composed of part-level and structural-level variational AE. On the other hand, convolutional deep belief networks [41] and GANs [40] are used to generate an occupancy grid. An approach by Henzler *et al.* [15] generates voxels from a 2D image. Rendering layers preceded by a generator provide images of the generated voxels. The training performed in an adversarial manner, so the discriminator assesses the generated image. Nash *et al.* [30] predict mesh vertices and faces using Transformer-based neural networks [39].

The aforementioned techniques optimize the layout with a reference shape, but it is not intended to generate a new shape. Our goal is to create combinatorial 3D objects by using minimal information such as desired class labels. To the best of our knowledge, this is the first attempt to build a new shape via a sequential and combinatorial approach. Our algorithm seeks a possible primitive combination effectively using Bayesian optimization, in order to reduce the number of observations required.

3 Sequential Assembly with Unit Primitives

In this section, we provide detailed configurations and assumptions, which are used to propose our combinatorial 3D shape generation method. In this work, we focus on volumetric representation, since it is the most straightforward representation to determine contacts between primitives. Such representation is also a reasonable choice for progressive shape assembling with a group of geometric primitives.

The choice of unit geometric primitives introduces interesting and challenging problems. Assume that we have a target shape (*i.e.*, Fig. 2(a)) to assemble. If we use 1×1-sized voxel primitives, there exists only *one combination* composed of 24 primitives, as shown in Fig. 2(b). However, if we use the primitives identical to 2×4 LEGO bricks, we can assemble *two combinations* (Fig. 2(c) and Fig. 2(d)). These observations indicate that smaller and simpler primitives tend to create more fine-grained shapes but have less combinatorial sequences. On the contrary, larger and more complicated primitives tend to create coarser

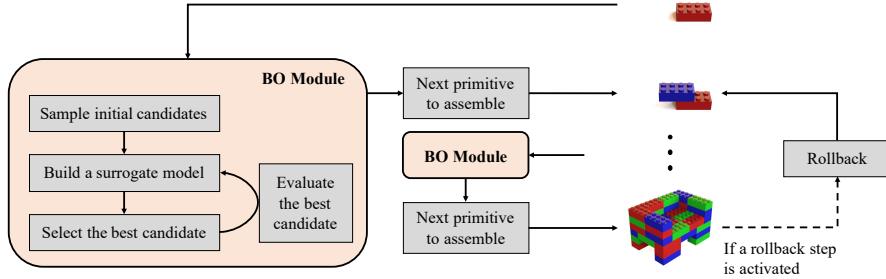


Fig. 3. Overview of our combinatorial 3D shape generation pipeline. BO stands for Bayesian optimization. A rollback step, which is activated if Eq. (6) is satisfied, moves back few steps to get rid of wrong placement. See Section 5 for the details.

shapes but have more interesting combinatorial sequences. These facts also relates to the real-world problem when we realize shapes by assembling multiple parts. As a result, we restrict the primitive we employ as a single type, which can only be connected with another primitive in fixed configurations² and does not allow overlap between primitives. In this work, we select 2×4 -LEGO brick as a unit primitive, since a 2×4 -sized primitive with studs and cavities is one of the representative basic LEGO bricks.³ However, the proposed approach is not limited to geometric primitive types.

4 Occupiability Grid

To tackle an assembly problem, we start by defining a search region $\mathcal{S} \subseteq \mathbb{R}^3$, which is a space to construct a 3D object. We employ *occupiability grid*, which is the opposite concept of probabilistic volumetric models [31]. The occupiability grid is a grid of which the unit cell (*i.e.*, a voxel) possesses the possibility of being occupied in the future.

Given the number of partitions for each of three axes m_1 , m_2 , and m_3 , a voxel \mathbf{v}_{ijk} can be represented as (i, j, k) where $i \in [m_1]$, $j \in [m_2]$, and $k \in [m_3]$,⁴ and a collection of entire voxels is $\{\mathbf{v}_{ijk}\}_{i \in [m_1], j \in [m_2], k \in [m_3]}$. For a generic voxel grid, the occupancy of a voxel is expressed as one of two alternatives:

$$q(\mathbf{v}_{ijk}) = \begin{cases} 1 & \text{if it is occupied,} \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

² If two 2×4 bricks are connected, there exist 46 types where one of them is a lower brick, and the other is an upper brick. See the supplementary material for the details.

³ Historically, as introduced in [16], the choice of unit primitives to 2×4 LEGO bricks is also significant, because this primitive is the most commonly used part for building a 3D statue.

⁴ $[m_i]$ denotes $\{1, \dots, m_i\}$.

On the other hand, the occupiability of voxel \mathbf{v}_{ijk} can be expressed as

$$o(\mathbf{v}_{ijk}) = \begin{cases} p & \text{if it is occupiable,} \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

where $p \in [0, 1]$ is the confidence of whether \mathbf{v}_{ijk} is occupiable in the future. In contrast with deterministic voxel definition, our concept of occupiability is to capture the likelihood of an empty voxel would be occupied in the near future. Thus, the voxel that is already occupied or prohibited due to specific constraints (see our formulation about these constraints in Section 3) will be regarded as 0 occupiability. We identify occupiable regions and determine p for all voxels as introduced in the subsequent sections.

Since a volumetric 2×4 brick as a unit primitive is placed in \mathcal{S} , it should transform into a single covariate to compare to other primitives. We thus denote the coordinate of each primitive as a 3D vector $\mathbf{x} = [x_1, x_2, x_3] \in \mathcal{S}$, where (x_1, x_2) is the center over the first two axes and x_3 is the bottom of the primitive, and the direction of each primitive as a scalar d . With this representation, suppose that every primitive is placed over the plane that $x_3 = 0$, and d of each primitive is placed either lengthwise (*i.e.*, denoted as $d = 0$) or breadthwise (*i.e.*, denoted as $d = 1$). It means every primitive can turn either $0, \pi/2, \pi$, or $3\pi/2$ radians along the third axis of \mathcal{S} . To sum up, each primitive is defined as a tuple (\mathbf{x}, d) (henceforth, denoted as $\mathbf{p} \in \mathcal{P}$ where $\mathcal{P} = \mathcal{S} \times \{0, 1\}$) where $x_3 \geq 0$ and $d \in \{0, 1\}$, and a n -primitive combination is expressed as a set $\{\mathbf{p}_i\}_{i=1}^n = \{(\mathbf{x}_i, d_i)\}_{i=1}^n$.

5 Combinatorial 3D Shape Generation

An assembling sequence is generated where each step in the series is suggested by one of the sequential model-based optimization methods, Bayesian optimization [6]. The Bayesian optimization strategy efficiently samples the position of the next primitive to assemble.

Evaluating Primitive Combinations. To determine the position of the next primitive guided by Bayesian optimization, we need to define an evaluation function f over $(n + 1)$ -primitive combinations:

$$s = f(\mathbf{p}_{n+1}; \{\mathbf{p}_i\}_{i=1}^n) + \epsilon, \quad (3)$$

which is considered as a true function, where ϵ is an observation noise. The critical point of choosing an evaluation function is that f should exhibit how every primitive in a combination follows the guidance of shape proposals. Our final (hard-to-achieve) goal is to find a global maximizer \mathbf{p}_{n+1}^\dagger of f with given n -primitive combination $\{\mathbf{p}_i\}_{i=1}^n$.

We design f to guide to follow a global shape context without providing exact probable positions of the primitive we would like to place. Because we assume objects in the same category share common parts such as the back of chairs and the leg of tables, this context guidance enables our method to configure an assembling sequence and generate a 3D object, given target class information.

Suppose that a voxel \mathbf{v}_{ijk} is not occupied and is able to be occupied where $\{\mathbf{p}_l\}_{l=1}^n$ is the current primitive combination. Given training examples $\{\mathbf{p}_l\}_{l=1}^{n_m}$ for $m \in [M]$ or a desired example $\{\mathbf{p}_l\}_{l=1}^{n_1}$, the occupiability p in Eq. (2) is set by the confidence of the volumes overlapped with the training examples or the desired example:

$$o(\mathbf{v}_{ijk}; \{\mathbf{p}_l\}_{l=1}^n) = \frac{1}{M} \sum_{m=1}^M q(\mathbf{v}_{ijk}; \{\mathbf{p}_l\}_{l=1}^{n_m}), \quad (4)$$

where M is the number of training examples of the desired class. With Eq. (3) and Eq. (4), the evaluation function over \mathbf{p}_{n+1} is

$$f(\mathbf{p}_{n+1}; \{\mathbf{p}_l\}_{l=1}^n) = \sum_{\mathbf{v}_{ijk} \in \mathbf{p}_{n+1}} o(\mathbf{v}_{ijk}; \{\mathbf{p}_l\}_{l=1}^n). \quad (5)$$

Determining Next Primitives to Assemble. Using an evaluation function Eq. (3) over possible primitive combinations, we can efficiently determine the position of the primitive to assemble. Bayesian optimization, which is a sample-efficient global optimization method for black-box functions, sequentially finds the next primitive candidate that maximizes an acquisition function. Since it optimizes an acquisition function, we do not need to assume differentiability, continuity, or any other specific functional form of the original function [6,11,34]. The detailed explanations for Bayesian optimization are introduced in the supplementary material to satisfy the page limit.

As shown in (5), the evaluation function that defines where we should assemble cannot be optimized using generic optimization strategies due to the unknown of functional forms. Due to this property, Bayesian optimization can be used to decide, where primitives should be assembled automatically. Moreover, determining a primitive position to assemble is taken into account as a process to reveal where we assemble the next position among huge possible primitives, which is a sequential combinatorial procedure to assembling primitives with Bayesian optimization. In this intuition, Algorithm 1 and Algorithm 2 are introduced.

Sampling a specific number of primitives is the main difference between our method and common Bayesian optimization strategies (see the supplementary material for the elaborate explanations). Well-known techniques for optimizing an acquisition function (*e.g.*, DIRECT [19] and L-BFGS-B [27]) are poorly worked, because our search space contains the complicated constraints that are determined by occupiabilities. Moreover, the combinatorial approach [2] is difficult to apply due to the curse of dimensionality, which is derived from the combinatorial explosion of inputs. We thus sample a feasible set from a primitive set, each element of which can assemble. This technique is used in the automated machine learning community [17,23].

As shown in Algorithm 1, after choosing v random primitives and evaluating those primitives with $\{\mathbf{p}_i\}_{i=1}^m$, where m is the cardinality of given primitive combination, a primitive candidate is queried, and new observation is accumulated,

Algorithm 1 Select Next Primitive Position to Assemble

Input: Initial primitive combination $\{\mathbf{p}_i\}_{i=1}^m$, the number of initial primitives v , the number of primitive candidates $q > v$.

Output: The next primitive to assemble \mathbf{p}_{m+1} .

- 1: Sample v primitives to be able to assemble randomly, $\{\mathbf{p}_i\}_{i=1}^v$.
- 2: Evaluate the primitive combinations each of which is composed of initial combination $\{\mathbf{p}_i\}_{i=1}^m$ and one of v primitives, $\{(\mathbf{p}_i, s_i)\}_{i=1}^v$.
- 3: **for** $j = v + 1, \dots, q$ **do**
- 4: Query the next primitive candidate to assemble \mathbf{p}_j .
- 5: Evaluate the primitive combination composed of $\{\mathbf{p}_i\}_{i=1}^m$ and \mathbf{p}_j .
- 6: Update the candidate set, $\{(\mathbf{p}_i, s_i)\}_{i=1}^j$.
- 7: **end for**
- 8: Select a next primitive \mathbf{p}_{m+1} , which has achieved the best score from $\{(\mathbf{p}_i, s_i)\}_{i=1}^q$.
- 9: **Return** \mathbf{p}_{m+1}

Algorithm 2 Find a Meaningful Primitive Combination

Input: An initial primitive combination $\{\mathbf{p}_i\}_{i=1}^m$, the number of primitives to assemble T , the number of rollback steps τ , threshold for rolling back α .

Output: A primitive combination $\{\mathbf{p}_j\}_{j=1}^{m+T}$.

- 1: Initialize $t \leftarrow 0$ and compute s_0 with $\{\mathbf{p}_t\}_{i=1}^m$.
- 2: **while** $t < T$ **do**
- 3: Select a next primitive to assemble \mathbf{p}_{t+1} , using Algorithm 1.
- 4: Assemble \mathbf{p}_{t+1} to the combination, $\{\mathbf{p}_t\}_{i=1}^{m+t}$.
- 5: Update $t \leftarrow t + 1$.
- 6: Compute s_t with $\{\mathbf{p}_t\}_{i=1}^{m+t}$.
- 7: **if** $t \geq \tau$ and $\sum_{k=0}^{\tau-1} (\max_{\mathbf{p}_{m+t-k}} s_{t-k}) - s_{t-\tau} < \alpha$ **then**
- 8: Roll back $\{\mathbf{p}_t\}_{i=1}^{m+t}$ to $\{\mathbf{p}_t\}_{i=1}^{m+t'}$ where $t' = t - \tau$.
- 9: Update $t \leftarrow t - \tau$.
- 10: **end if**
- 11: **end while**
- 12: **Return** $\{\mathbf{p}_j\}_{j=1}^{m+T}$

until q primitives are observed. Finally, the next primitive that has achieved the best score is returned. Bayesian optimization iterates these two steps up to a time budget: (i) building a surrogate function (in this paper Gaussian process regression) with historical observations; (ii) determining and observing a primitive to assemble, which maximizes an acquisition function. See the supplementary material for the details. Algorithm 2 describes how a meaningful primitive combination is assembled with the evaluation function that can guide a stacking process. Consequently, we obtain a $(m + T)$ -primitive combination.

Rolling Back Primitives Previously Assembled. Our method mentioned above might be able to provide a sub-optimal sequence, because the Bayesian optimization module, which always guarantees to find a global solution [6,33,36], accumulates local (but possibly global) optima in a row. Hence, our method

includes a rollback step, as shown in Line 7 to Line 10 in Algorithm 2. Given the number of rollback steps τ and a threshold for rolling back α , if the condition written below is satisfied:

$$t \geq \tau \quad \text{and} \quad \sum_{k=0}^{\tau-1} \left(\max_{\mathbf{P}_{m+t-k}} s_{t-k} \right) - s_{t-k} < \alpha, \quad (6)$$

t -primitive combination is rolled back to $t - \tau$ -primitive combination. To avoid rolling back in the same combination infinitely, we skip if the same assembly step is repeated for five times. Since our BO module attempts to find a global optimum with a candidate sampling, this heuristic helps to improve a generation quality as shown in Section 6.2.

6 Experimental Results

In this section, we demonstrate the assembly results for solving a combinatorial 3D shape generation task. Before introducing the results, we first describe the experimental setups in detail.

Gaussian process regression with Matérn 5/2 kernels [32] and Gaussian process upper confidence bound (GP-UCB) [36] are used as a surrogate model and an acquisition function, respectively. The hyperparameters of kernels in the regression models are optimized by marginal likelihood maximization with BFGS algorithm. Similar to the setting in [36], the trade-off hyperparameter of GP-UCB is monotonically increasing over iterations. Rather than setting a specific number of samples for Line 4 of Algorithm 1, we sample as many candidates of maximizer as possible within the given time budget. We set the time budget as 1 second in most of the cases. Unless otherwise specified, v and q in Algorithm 1 are set to 10 and 20, respectively. In all the experiments, an initial primitive combination in Algorithm 2 is given as $\{([0, 0, 0], 0)\}$, to assemble from scratch. Bayesian optimization modules are implemented with [22].

Open3D framework [43] and Mitsuba renderer [18] are employed to deal with 3D objects and visualize the primitive assembly results. For attractive visualization, we randomly pick the color of primitives from red, blue, and green.

6.1 Combinatorial 3D Shape Generation via Sequential Assembly

We sequentially generate a combinatorial 3D shape with one of two schemes: (i) optimizing an evaluation function f , given the occupiability computed with multiple examples; (ii) optimizing f , given the occupiability with a single example, as described in Section 5. By Eq. (5), the former and latter provide continuous and binary evaluation scores, respectively. As it will be shown in this section, the generation process with binary evaluation scores can easily create sophisticated and decent objects, while the process with continuous evaluation scores enables the local segments to be captured and shared across training examples. From now, we demonstrate the diverse generation examples guided by our novel

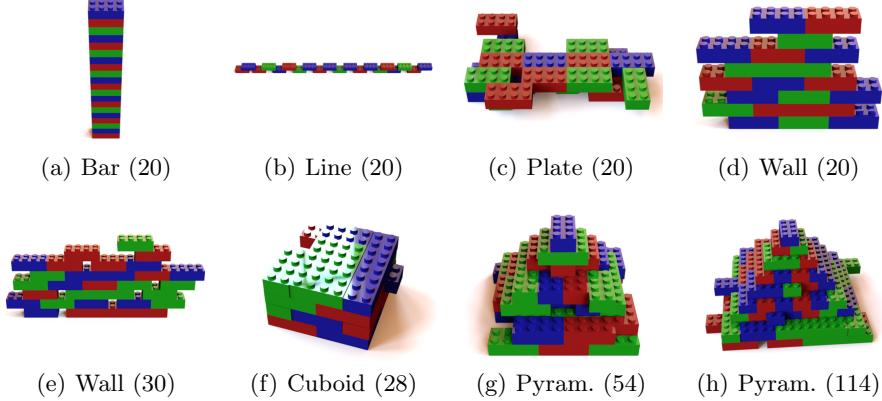


Fig. 4. Results on assembling primitive combinations using Algorithm 2. Pyram. stands for square pyramid, and the number of primitives are described in parentheses.

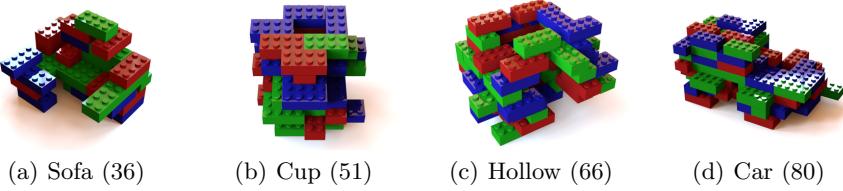


Fig. 5. Results on low-quality primitive combinations generated by our method. The number of primitives are described in parentheses.

combinatorial 3D shape dataset and explicitly show the assembly sequences that can create a 3D object in a human assembling procedure.

The assembly results with the continuous occupiability are displayed in Fig. 4. It successfully generates unseen 3D shapes, which belong to group B of our dataset. For the categories in group C of our dataset, Fig. 1 and Fig. 6 represent the assembly results of combining our method with the binary occupiability. In addition, we provide low-quality generation results in Fig. 5. There may exist multiple reasons why it fails to create the 3D shapes of which the category is intended. However, due to the combinatorial properties defined in Section 3 and Section 4, the main reason is prominent circumstances of which the probable positions for the placement of next primitive cannot be readily found along with the primitive assembly.

6.2 Ablation Study

In this section, we discuss the ablation studies when (i) we use explicit functions for the Bayesian optimization module, instead of Eq. (5) and (ii) we assemble a combinatorial shape without rollbacks.

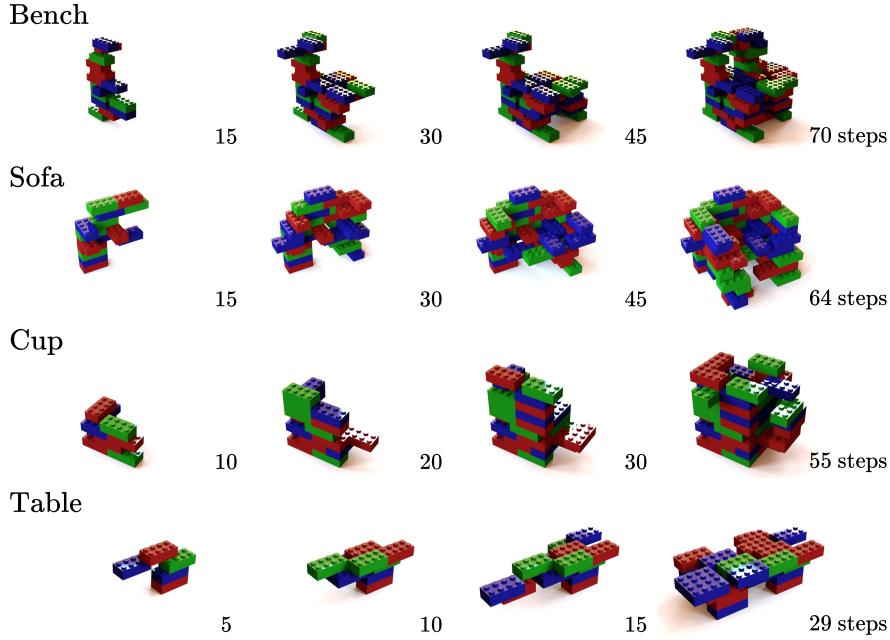


Fig. 6. Results on sequential primitive assemblies. The last column shows the final primitive combinations, and the first three columns show the intermediate steps to create the final primitive combinations. The step of each panel is specified in the bottom right corner of each panel.

Explicit Evaluation Functions for Bayesian Optimization Module. To confirm the validity of our Bayesian optimization module where an evident evaluation function is given. For example, the height of primitive combination can express the current status of the given combination as an evaluation score:

$$f(\mathbf{p}_{n+1}; \{\mathbf{p}_i\}_{i=1}^n) = \max_{(x_1, x_2, x_3, d) \in \{\mathbf{p}_i\}_{i=1}^{n+1}} x_3 + 1. \quad (7)$$

Similar to Eq. (7), we define three more specific functions: width, depth, and the number of connected studs. Likewise, these BO modules attempt to maximize the evaluation functions we define. For these experiments, we use three baselines:

- (i) Oracle: It is the best achievable result;
- (ii) Random: This baseline is a fully-randomized result. One of the primitives possible to assemble is uniformly selected at every assembling step;
- (iii) Random with evaluations: It is a randomized result, but the best primitive is chosen at every step after evaluating the primitives uniformly sampled. The number of the sampled primitives is set to q to compare with our method fairly.

As shown in Fig. 7, our method outperforms other methods and achieves the results close to the oracles of four experiments.

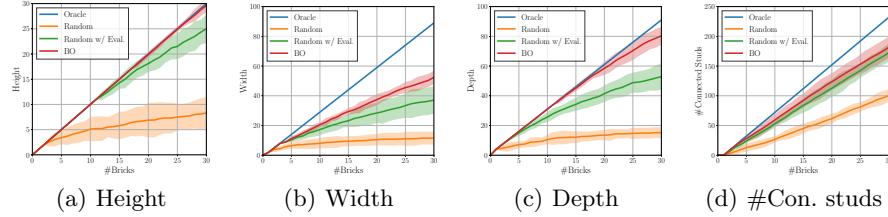


Fig. 7. Quantitative results on maximizing four explicit evaluation functions (*i.e.*, height, width, depth, and the number of connected studs) with our method. BO indicates our sequential assembly method. All the experiments are repeated ten times, and their means and the 1.96 standard deviations are plotted. #Con. studs stands for the number of connected studs.

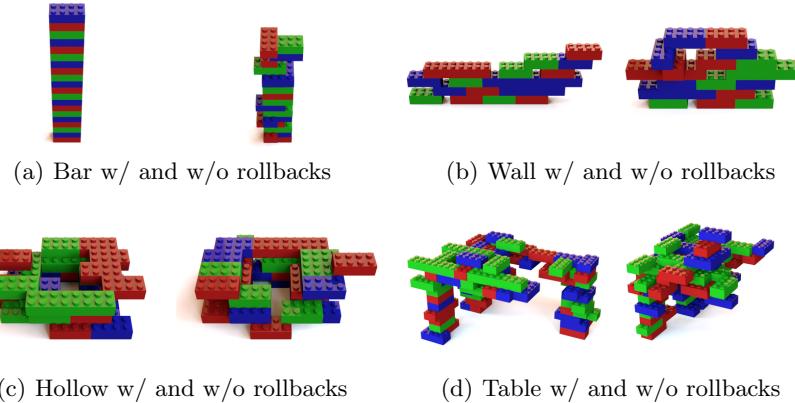


Fig. 8. Results on sequential assemblies with and without rollbacks. The left and right panels of each figure are the objects generated with and without rollbacks, respectively.

Without Rollbacks. We test a combinatorial 3D shape generation process without rollbacks and compare it to the generation process with rollbacks to empirically demonstrate its ability to enhance the overall quality of generated objects. As discussed before, it does not always guarantee to improve generation quality, but it is helpful in many cases, such as the examples in Fig. 8. For the cases where the next primitive should be assembled in the distinct position of limited probable regions (*e.g.*, Fig. 8(a) and Fig. 8(b)), the effect of rollbacks is clearly helpful. Rolling back and placing several times more until we obtain a reasonable score can fix the probable placement of primitives. On the contrary, the shapes are shown in Fig. 8(c) and Fig. 8(d) demonstrate appropriate rollback conditions – in general, it enhances the quality of the assembled results considering the global context of the class information. The object shown in Fig. 8(c) presents a 3D shape of generated hollow without rollbacks that is

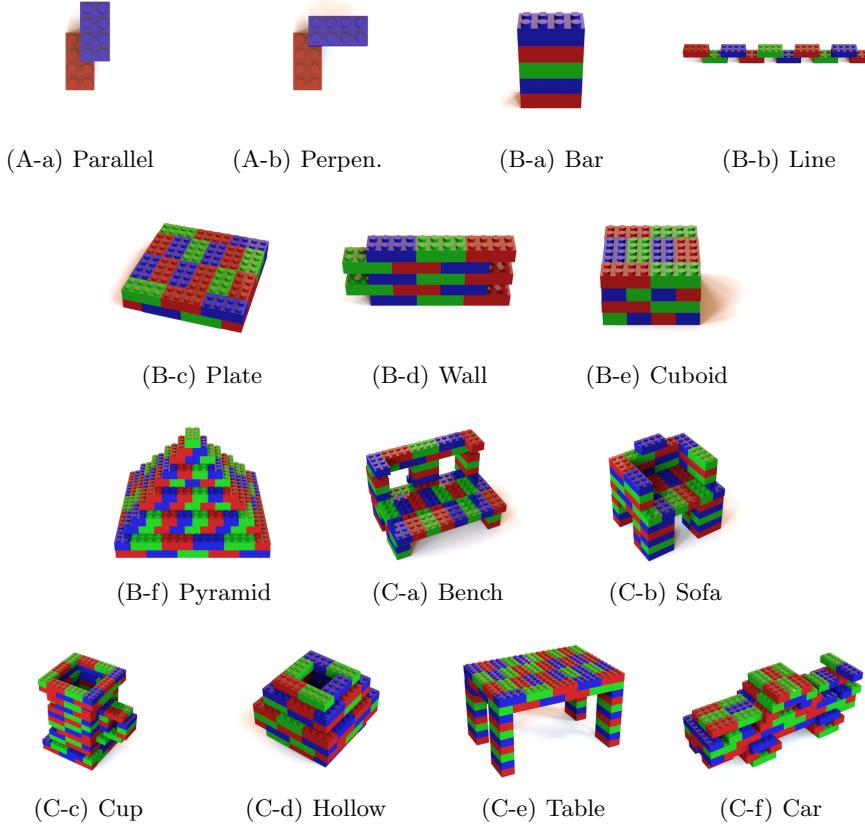


Fig. 9. Selected examples from our combinatorial 3D shape dataset. One example per class is visualized.

loosely connected rather than the one with rollbacks. Moreover, Fig. 8(d) exhibits the generation with rollbacks follows better than the one without rollbacks.

7 Combinatorial 3D Shape Dataset

Twelve human subjects constructed a combinatorial 3D shape dataset, composed of 406 instances of 14 classes. Specifically, each object in our dataset is considered equivalent to a sequence of primitive placement. For this reason, compared to other 3D object datasets [7,42], our proposed dataset contains an assembling sequence of unit primitives. It implies that we can quickly obtain a sequential generation process that is a human assembling mechanism. Furthermore, we can sample valid random sequences from a given combinatorial shape after validating the tested combination. To sum up, the characteristics of our combinatorial 3D shape dataset are

- (i) combinatorial: Duplicates of unit primitive is repeatedly connected;
- (ii) sequential: Because there are allowable connections between primitives, they are sequentially added;
- (iii) decomposable: Compared to other common datasets, because of the combinatorial property, parts of combination can be sampled if they are validated by the contact and overlap conditions defined in Section 3;
- (iv) manipulable: New primitive are addable or existing primitives are removable.

Our 3D shape dataset is implemented to satisfy the above properties, supporting sequential assembly, combination validation, possible position listing, and part sampling.

As shown in Fig. 9, we select 14 classes: parallel, perpendicular, bar, line, plate, wall, cuboid, square pyramid, bench, sofa, cup, hollow, table, and car. Parallel that implies the directions of two primitives are same, and perpendicular that implies the directions of two primitives are different classes are own connection types of 2×4 -sized primitives with studs and cavities (denoted as group A). Bar, plate, cuboid, wall, and square pyramid classes are taken into account as default components to assemble sophisticated shapes (denoted as group B). The other classes are abstractly thought of as the combination of those default components (denoted as group C). More diverse examples and the statistics of our dataset can be found in the supplementary material. We will release our dataset and its associated codes as an open-source benchmark.

8 Conclusion

We propose a sequential assembly method for a combinatorial 3D generation problem. It can generate a combinatorial object, considering sample efficiency that is guided by Bayesian optimization. The evaluation function based on global shape guidance demonstrates that our method generates 3D objects composed of unit primitives, as shown in Section 6. Besides, we create a new dataset for combinatorial 3D models. This dataset allows us to generate 3D objects sequentially. To develop our combinatorial 3D generation further, we could apply a reinforcement learning-based approach to plan a long-term policy for sequential decision making. This enables a more sophisticated policy to be involved within the process, which then allows us to look ahead for a future primitive combination. Consequently, the position of the next primitive can be provided based on the evaluation of these feasible future configurations.

Acknowledgement

We wish to thank Jinhwi Lee for the constructive discussions and suggestions and also for the help in developing this research.

References

1. Achlioptas, P., Diamanti, O., Mitliagkas, I., Guibas, L.: Learning representations and generative models for 3D point clouds. In: Proceedings of the International Conference on Machine Learning (ICML). pp. 40–49 (2018)
2. Baptista, R., Poloczek, M.: Bayesian optimization of combinatorial structures. In: Proceedings of the International Conference on Machine Learning (ICML). pp. 462–471 (2018)
3. Bendsøe, M.P., Kikuchi, N.: Generating optimal topologies in structural design using a homogenization method. Computer Methods in Applied Mechanics and Engineering **71**(2), 197–224 (1988)
4. Borrvall, T., Petersson, J.: Topology optimization of fluids in Stokes flow. International Journal for Numerical Methods in Fluids **41**(1), 77–107 (2003)
5. Brackett, D., Ashcroft, I., Hague, R.: Topology optimization for additive manufacturing. In: Proceedings of the Solid Freeform Fabrication Symposium. pp. 348–362 (2011)
6. Brochu, E., Cora, V.M., de Freitas, N.: A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. arXiv preprint arXiv:1012.2599 (2010)
7. Chang, A.X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., Xiao, J., Yi, L., Yu, F.: ShapeNet: An information-rich 3D model repository. arXiv preprint arXiv:1512.03012 (2015)
8. Eilers, S.: The LEGO counting problem. The American Mathematical Monthly **123**(5), 415–426 (2016)
9. Eschenauer, H.A., Kobelev, V.V., Schumacher, A.: Bubble method for topology and shape optimization of structures. Structural Optimization **8**(1), 42–51 (1994)
10. Fan, H., Su, H., Guibas, L.J.: A point set generation network for 3D object reconstruction from a single image. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR). pp. 605–613 (2017)
11. Frazier, P.I.: A tutorial on Bayesian optimization. arXiv preprint arXiv:1807.02811 (2018)
12. Gao, L., Yang, J., Wu, T., Yuan, Y., Fu, H., Lai, Y., Zhang, H.: SDM-NET: Deep generative network for structured deformable mesh. arXiv preprint arXiv:1908.04520 (2019)
13. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in Neural Information Processing Systems (NeurIPS). vol. 27, pp. 2672–2680 (2014)
14. Groueix, T., Fisher, M., Kim, V.G., Russell, B.C., Aubry, M.: AtlasNet: A Papier-Mâché approach to learning 3D surface generation. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR). pp. 216–224 (2018)
15. Henzler, P., Mitra, N., Ritschel, T.: Escaping Plato’s Cave using adversarial training: 3D shape from unstructured 2D image collections. In: Proceedings of the International Conference on Computer Vision (ICCV). pp. 9984–9993 (2019)
16. Herman, S.: Building a History: The LEGO Group. Grub Street Publishers (2012)
17. Hutter, F., Hoos, H.H., Leyton-Brown, K.: Sequential model-based optimization for general algorithm configuration. In: Proceedings of the International Conference on Learning and Intelligent Optimization (LION). pp. 507–523 (2011)
18. Jakob, W.: Mitsuba renderer. <http://www.mitsuba-renderer.org> (2010)

19. Jones, D.R., Perttunen, C.D., Stuckman, B.E.: Lipschitzian optimization without the Lipschitz constant. *Journal of Optimization Theory and Applications* **79**(1), 157–181 (1993)
20. Jones, D.R., Schonlau, M., Welch, W.J.: Efficient global optimization of expensive black-box functions. *Journal of Global Optimization* **13**, 455–492 (1998)
21. Kharmanda, G., Olhoff, N., Mohamed, A., Lemaire, M.: Reliability-based topology optimization. *Structural and Multidisciplinary Optimization* **26**(5), 295–307 (2004)
22. Kim, J., Choi, S.: bayeso: A Bayesian optimization framework in Python. <http://bayeso.org> (2017)
23. Kim, J., Jeong, J., Choi, S.: AutoML Challenge: AutoML framework using random space partitioning optimizer. In: International Conference on Machine Learning Workshop on Automatic Machine Learning (AutoML) (2016)
24. Kingma, D.P., Welling, M.: Auto-encoding variational Bayes. In: Proceedings of the International Conference on Learning Representations (ICLR) (2014)
25. Kozaki, T., Tedenuma, H., Maekawa, T.: Automatic generation of LEGO building instructions from multiple photographic images of real objects. *Computer-Aided Design* **70**, 13–22 (2016)
26. Lee, S., Kim, J., Kim, J.W., Moon, B.: Finding an optimal LEGO® brick layout of voxelized 3D object using a genetic algorithm. In: Proceedings of the Annual Conference on Genetic and Evolutionary Computation (GECCO). pp. 1215–1222 (2015)
27. Liu, D.C., Nocedal, J.: On the limited memory BFGS method for large scale optimization. *Mathematical Programming* **45**(3), 503–528 (1989)
28. Luo, S., Yue, Y., Huang, C., Chung, Y., Imai, S., Nishita, T., Chen, B.: Legolization: Optimizing LEGO designs. *ACM Transactions on Graphics* **34**(6), 222:1–222:12 (2015)
29. Moćkus, J., Tiesis, V., Žilinskas, A.: The application of Bayesian methods for seeking the extremum. *Towards Global Optimization* **2**, 117–129 (1978)
30. Nash, C., Ganin, Y., Eslami, S.M.A., Battaglia, P.W.: PolyGen: An autoregressive generative model of 3D meshes. arXiv preprint arXiv:2002.10880 (2020)
31. Pollard, T., Mundy, J.L.: Change detection in a 3-d world. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1–6 (2007)
32. Rasmussen, C.E., Williams, C.K.I.: Gaussian Processes for Machine Learning. MIT Press (2006)
33. Scarlett, J.: Tight regret bounds for Bayesian optimization in one dimension. In: Proceedings of the International Conference on Machine Learning (ICML). pp. 4500–4508 (2018)
34. Shahriari, B., Swersky, K., Wang, Z., Adams, R.P., de Freitas, N.: Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE* **104**(1), 148–175 (2016)
35. Snoek, J., Larochelle, H., Adams, R.P.: Practical Bayesian optimization of machine learning algorithms. In: Advances in Neural Information Processing Systems (NeurIPS). vol. 25, pp. 2951–2959 (2012)
36. Srinivas, N., Krause, A., Kakade, S., Seeger, M.: Gaussian process optimization in the bandit setting: No regret and experimental design. In: Proceedings of the International Conference on Machine Learning (ICML). pp. 1015–1022 (2010)
37. Testuz, R., Schwartzburg, Y., Pauly, M.: Automatic generation of constructable brick sculptures. In: Proceedings of the Annual Conference of the European Association for Computer Graphics (Eurographics). pp. 81–84 (2013), short Paper

38. van den Hengel, A., Russell, C., Dick, A., Bastian, J., Pooley, D., Fleming, L., Agapito, L.: Part-based modelling of compound scenes from images. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR). pp. 878–886 (2015)
39. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: Advances in Neural Information Processing Systems (NeurIPS). pp. 5998–6008 (2017)
40. Wu, J., Zhang, C., Xue, T., Freeman, B., Tenenbaum, J.: Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling. In: Advances in Neural Information Processing Systems (NeurIPS). pp. 82–90 (2016)
41. Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., Xiao, J.: 3D ShapeNets: A deep representation for volumetric shapes. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1912–1920 (2015)
42. Xiang, Y., Kim, W., Chen, W., Ji, J., Choy, C., Su, H., Mottaghi, R., Guibas, L., Savarese, S.: ObjectNet3D: A large scale database for 3D object recognition. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 160–176 (2016)
43. Zhou, Q., Park, J., Koltun, V.: Open3D: A modern library for 3D data processing. arXiv preprint arXiv:1801.09847 (2018)

Supplementary Material

In this article, we describe the parts which are omitted from the main article to accommodate a page limit. First, we show the connection types between two 2×4 -sized primitives. Then, we explain the details of Bayesian optimization module, and visualize some examples of our combinatorial 3D shape dataset.

S.1 Connection Types between Two 2×4 -sized Primitives

There exist 46 connection types between two 2×4 LEGO brick-shaped primitives where upper and lower primitives are fixed, as shown in Fig. s.1. They comprise group A of our combinatorial dataset.

S.2 Bayesian Optimization Module

In this section, we describe the elaborate explanations of Bayesian optimization module, which is about how to query the next primitive candidate.

First of all, similar to common Bayesian optimization [20,29], a surrogate function over primitives is estimated, given r historical observations $\{(\mathbf{p}_j, s_j)\}_{j=1}^r$. In general, Gaussian process regression [32] is used as a surrogate function in the Bayesian optimization community [6,35], because it can express any function in the reproducing kernel Hilbert space. Note that each primitive \mathbf{p} is regarded as a four-dimensional vector, composed of a 3D vector of primitive \mathbf{x} and a direction of primitive d . By the Gaussian process regression, given r four-dimensional inputs

$$\mathbf{P} = \begin{bmatrix} \mathbf{p}_1 \\ \vdots \\ \mathbf{p}_r \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 & d_1 \\ \vdots & \vdots \\ \mathbf{x}_r & d_r \end{bmatrix} \in \mathbb{R}^{r \times 4}, \quad (\text{s.1})$$

and their associated outputs $\mathbf{s} = [s_1 \cdots s_r] \in \mathbb{R}^r$, a function value and its uncertainty are represented by posterior mean and variance functions:

$$\mu(\mathbf{p}) = \mathbf{k}(\mathbf{p}, \mathbf{P}) (\mathbf{K}(\mathbf{P}, \mathbf{P}) + \sigma_n^2 \mathbf{I})^{-1} \mathbf{s}, \quad (\text{s.2})$$

$$\sigma^2(\mathbf{p}) = k(\mathbf{p}, \mathbf{p}) - \mathbf{k}(\mathbf{p}, \mathbf{P}) (\mathbf{K}(\mathbf{P}, \mathbf{P}) + \sigma_n^2 \mathbf{I})^{-1} \mathbf{k}(\mathbf{P}, \mathbf{p}), \quad (\text{s.3})$$

where kernels are defined as $k : \mathbb{R}^4 \times \mathbb{R}^4 \rightarrow \mathbb{R}$, $\mathbf{k} : \mathbb{R}^4 \times \mathbb{R}^{r \times 4} \rightarrow \mathbb{R}^r$, and $\mathbf{K} : \mathbb{R}^{r_1 \times 4} \times \mathbb{R}^{r_2 \times 4} \rightarrow \mathbb{R}^{r_1 \times r_2}$. In addition, σ_n^2 is a noise scale and \mathbf{I} is an identity matrix.

With the surrogate function represented by (s.2) and (s.3), we compute the acquisition function values for ν primitives possible to assemble, and find a maximizer among the ν primitives. A maximizer \mathbf{p}^* of the acquisition function over \mathbf{p} , $a(\mathbf{p}; \mu(\mathbf{p}), \sigma^2(\mathbf{p}))$ for balancing exploitation and exploration is found to observe the true function $f(\mathbf{p}^*)$:

$$\mathbf{p}^* = \arg \max_{\mathbf{p} \in \mathcal{P}} a(\mathbf{p}; \mu(\mathbf{p}), \sigma^2(\mathbf{p})), \quad (\text{s.4})$$

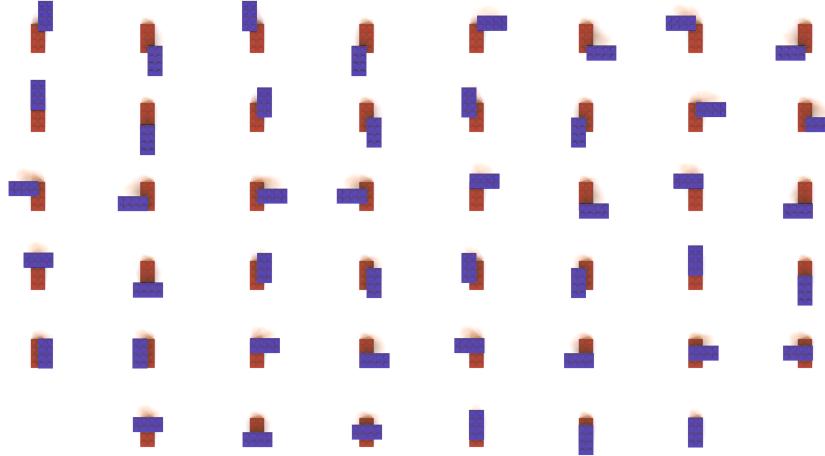


Fig. s.1. 46 connection types between two 2×4 -sized primitives.

Algorithm s.1 Query a Candidate of Next Primitive

Input: A n -primitive combination $\{\mathbf{p}_i\}_{i=1}^n$, r possible primitives observed and their evaluation scores with the n -primitive combination $\{(\mathbf{p}_j, s_j)\}_{j=1}^r$, the number of samples for acquisition function optimization ν .

Output: A candidate of next primitive \mathbf{p}_{r+1}^* .

1: Estimate a surrogate function

$$\hat{f}(\mathbf{p}; \{\mathbf{p}_i\}_{i=1}^n, \{(\mathbf{p}_j, s_j)\}_{j=1}^r),$$

using the primitive combination and the historical observations.

2: Sample ν primitives possible to assemble.

3: Find a maximizer \mathbf{p}_{r+1}^* , i.e., one of ν primitives, of the acquisition function computed by \hat{f} .

4: **Return** \mathbf{p}_{r+1}^*

where \mathcal{P} is a compact set. Expected improvement [29] and Gaussian process upper confidence bound (GP-UCB) [36] are widely used as an acquisition function. In this paper, we use GP-UCB as an acquisition function:

$$a_{\text{UCB}}(\mathbf{p}; \mathbf{P}, \mathbf{s}) = \mu(\mathbf{p}) + \gamma\sigma(\mathbf{p}), \quad (\text{s.5})$$

where γ is a trade-off hyperparameter for function values and its uncertainties. These are presented in Algorithm s.1.

S.3 Combinatorial 3D Shape Dataset

We demonstrate a part of our combinatorial 3D shape dataset, as shown in Fig. s.1, and Fig. s.2 to Fig. s.13. In addition, the statistics on our dataset is specified in Table s.1.



Fig. s.2. Examples of *Bar* class.



Fig. s.3. Examples of *Line* class.

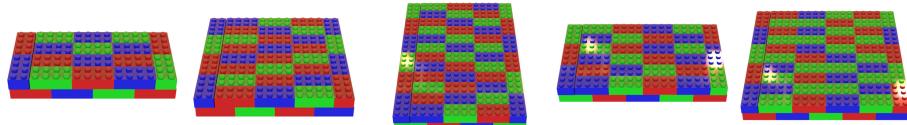


Fig. s.4. Examples of *Plate* class.

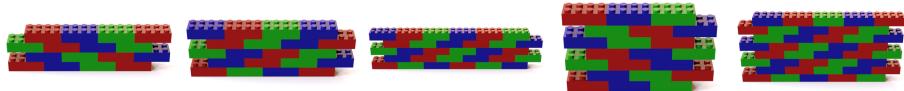


Fig. s.5. Examples of *Wall* class.

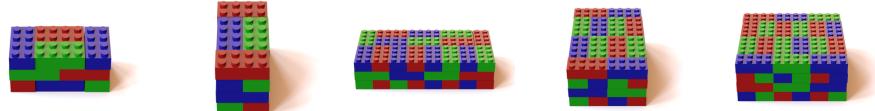


Fig. s.6. Examples of *Cuboid* class.

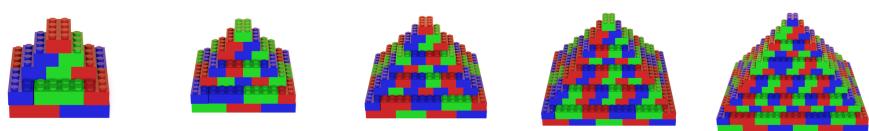


Fig. s.7. Examples of *Square Pyramid* class.

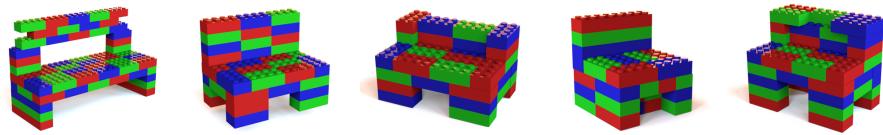


Fig. s.8. Examples of *Bench* class.



Fig. s.9. Examples of *Sofa* class.

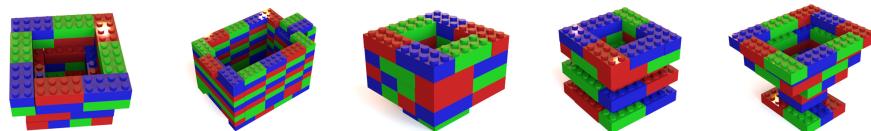


Fig. s.10. Examples of *Cup* class.

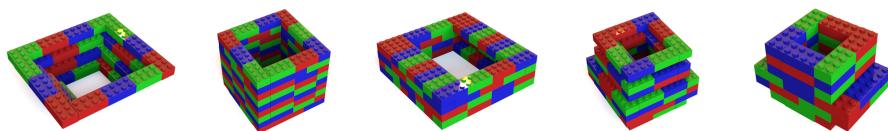


Fig. s.11. Examples of *Hollow* class.



Fig. s.12. Examples of *Table* class.

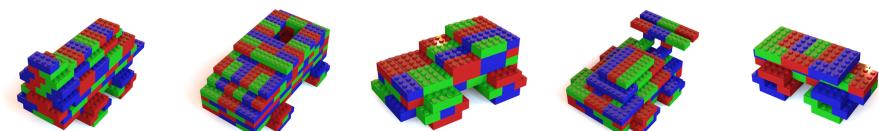


Fig. s.13. Examples of *Car* class.

Table s.1. Statistics on combinatorial 3D shape dataset. Std stands for standard deviation.

Group A	Parallel	Perpen.				
#instances	21	25				
Mean of #primitives per instance	2.0	2.0				
Std of #primitives per instance	0.0	0.0				
Group B	Bar	Line	Plate	Wall	Cuboid	Sq. Pyramid
#instances	30	30	30	30	30	30
Mean of #primitives per instance	11.9	32.5	56.0	27.9	26.4	164.0
Std of #primitives per instance	6.6	42.0	35.1	14.6	17.6	129.2
Group C	Bench	Sofa	Cup	Hollow	Table	Car
#instances	30	30	30	30	30	30
Mean of #primitives per instance	55.4	59.6	49.7	46.3	36.9	83.6
Std of #primitives per instance	28.0	30.5	31.2	31.8	19.2	41.0