

Mamba-CAD: State Space Model For 3D Computer-Aided Design Generative Modeling

Xueyang Li, Yunzhong Lou, Yu Song, Xiangdong Zhou✉

School of Computer Science, Fudan University, Shanghai, China

xueyangli21@m.fudan.edu.cn, yzlou20@fudan.edu.cn, songy23@m.fudan.edu.cn, xdzhou@fudan.edu.cn

Abstract

Computer-Aided Design (CAD) generative modeling has a strong and long-term application in the industry. Recently, the parametric CAD sequence as the design logic of an object has been widely mined by sequence models. However, the industrial CAD models, especially in component objects, are fine-grained and complex, requiring a longer parametric CAD sequence to define. To address the problem, we introduce Mamba-CAD, a self-supervised generative modeling for complex CAD models in the industry, which can model on a longer parametric CAD sequence. Specifically, we first design an encoder-decoder framework based on a Mamba architecture and pair it with a CAD reconstruction task for pre-training to model the latent representation of CAD models; and then we utilize the learned representation to guide a generative adversarial network to produce the fake representation of CAD models, which would be finally recovered into parametric CAD sequences via the decoder of Mamba-CAD. To train Mamba-CAD, we further create a new dataset consisting of 77,078 CAD models with longer parametric CAD sequences. Comprehensive experiments are conducted to demonstrate the effectiveness of our model under various evaluation metrics, especially in the generation length of valid parametric CAD sequences.

Introduction

Computer-Aided Design (CAD) generative modeling has garnered widespread attention in the industry, particularly in fields such as manufacturing, vehicle engineering, and architectural design. The design logic of each object in the world can be expressed by a sequence of parametric commands. For example, constructing a cylinder is defined with the radius to draw a complete circle which would be further extruded with the determined height. Typically, the complex design logic requires a longer sequence of parametric commands while a simpler object only requires a shorter parametric CAD sequence to define. In CAD fields, this kind of representation is called parametric CAD sequences, allowing designers to use CAD tools (e.g., AutoCAD and SolidWorks) to quickly edit the model by modifying their commands and parameters. It is undeniable that prototypes of various objects in the world almost starts with CAD designs.

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

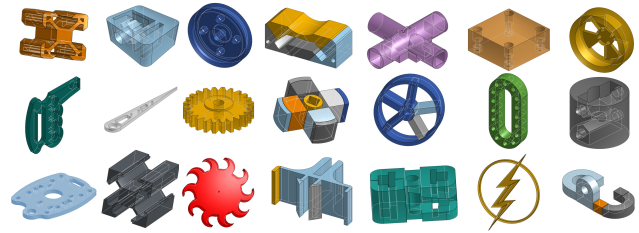


Figure 1: The gallery of CAD models created by Mamba-CAD, including visual illustrations of the created CAD models from CAD sequence reconstruction, completion, and random generation.

Therefore, it attracts much attention of generative modeling on CAD models recently (Ren et al. 2022; Jones et al. 2023a,b). The parametric CAD sequence, as a key point used in CAD tools to create 3D shapes, can be seen as a special discrete language consisting of commands and their parameters. Hence, some interesting efforts (Wu, Xiao, and Zheng 2021; Xu et al. 2022, 2023) leverage Transformer-based architecture (Vaswani et al. 2017) to learn the representation of parametric CAD sequences and achieve impressive generation results. However, the parametric CAD sequences generated by them are not long enough to satisfy the construction of complex 3D shapes. In practical applications, especially in the industrial manufacturing, the design logic of objects is relatively complex, which necessitates longer parametric CAD sequences to define. This motivates us to make a step forward, trying to design a generative model with the capability of producing a longer parametric CAD sequence. Inspired by recent Mamba (Gu and Dao 2023), a State Space Model (SSM) performing well in modeling long sequences, we further explore its capability for the parametric CAD sequence modeling. Note that we mainly discuss the CAD models defined with Construct Solid Geometry (CSG) representation as it is conducive to be parameterized, while the other Boundary Representation (BRep) for CAD models is beyond the scope of this paper as it lacks sufficient workflow construction to recover how the designs are created by CAD commands.

To overcome the challenge of generative modeling of CAD models, we propose Mamba-CAD, a self-supervised

framework for a longer parametric CAD sequence generation, which is easy to implement and straightforward to train. Specifically, we first design an auto encoder-decoder architecture based on a Mamba backbone to process the parametric CAD sequence and pair it with a CAD reconstruction task for pre-training. When the network is well-trained, we further select the learned latent representation as a knowledgeable distribution to train a latent generative adversarial network (Chen and Zhang 2019; Achlioptas et al. 2018) to produce the fake representation of CAD models, which is randomly sampled from a standard gaussian distribution. Finally, these fake representations would be fed into the decoder of Mamba-CAD to generate parametric CAD sequences. Given that the benchmark DeepCAD dataset (Wu, Xiao, and Zheng 2021) only contains CAD models with the maximum 60 commands, it is hard to show the capability of Mamba-CAD in handling a longer parametric CAD sequence. Hence, we again filter the DeepCAD dataset by discarding some CAD models with a total number of commands less than 10, at the same time we use a feature script from (Onshape 2019) to parse CAD models from the ABC dataset (Koch et al. 2019) to retain samples with the total number of commands from 61 to 128. Finally, total 77,078 CAD models described as parametric CAD sequences are grouped into a new dataset that will be released publicly to promote future research on the parametric CAD modeling.

In summary, our main contribution is Mamba-CAD, a simple yet effective self-supervised framework for generative CAD modeling, which can handle a longer parametric CAD sequence to achieve more complex CAD models. To train Mamba-CAD, we further construct a new dataset of parametric CAD sequences that is more complex than the existing benchmark dataset of the same type and will make it publicly available in the future. Comprehensive experiments show that Mamba-CAD outperforms other competitive methods in tasks such as CAD sequence reconstruction, completion, and random generation, especially in the generation length of valid parametric CAD sequences.

Related Work

Parametric CAD modeling

Boundary Representation (BRep) and Construct Solid Geometry (CSG) are two widely used representations in parametric CAD fields. For BRep representation, BRepBERT (Lou et al. 2023) applies Graph Neural Network (Scarselli et al. 2008) as a tokenizer to process BRep representation and further enhance its feature representation via conducting the masked entity modeling. ComplexGen (Guo et al. 2022) views the reconstruction process of CAD models as a chain complex in the BRep representation, achieving the task of CAD reconstruction from point cloud. Solidgen (Jayaraman et al. 2023) employs Transformer (Vaswani et al. 2017) and Pointer network (Vinyals, Fortunato, and Jaitly 2015) to facilitate the generation of complex B-rep structures. (Zhou, Tang, and Zhou 2023) design a CADParser, a learning approach of sequence modeling for converting BRep data to CSG representation. For CSG representation, CSG-Stump (Ren et al. 2021) sim-

plifies the CSG representation into a three-layer structure consisting of union, intersection, and complement operation. SkexGen (Xu et al. 2022) utilizes two individual and distinct Transformer-based branches to learn representations of “sketch” and “extrude” operation separately. HNC-CAD (Xu et al. 2023) proposes a code tree representation learned by a two-stage cascaded auto-regressive networks. Besides, some concurrent works also explore reconstructing CAD models from point clouds, voxels, meshes, and images (Yu et al. 2022; Uy et al. 2022; Li et al. 2022, 2023; Yu et al. 2024; Khan et al. 2024; Li et al. 2024b). Unlike these efforts, our method mainly focuses on addressing the generation problem of a longer parametric CAD sequence based on CSG like representation. The most related to our work is DeepCAD (Wu, Xiao, and Zheng 2021), a Transformer-based architecture, which encodes the parametric CAD sequence into a latent space and then decodes it back to the constructed 3D shape. Different from DeepCAD our framework is designed for generate longer parametric CAD sequence to construct a more complex 3D shape.

Mamba-based models

Mamba, a recent selective structured state space model (Gu and Dao 2023) with excellent capabilities of processing long sequence data, has been widely explored in vision, language, and sequence modeling. For Mamba in the vision task, inspired by ViT (Liu et al. 2021), VMamba (Liu et al. 2024b) and ViMamba (Zhu et al. 2024) are two early representative Mamba models that view several image patches as sequences from the horizontal and vertical dimensions of the image and perform bi-directional scanning on these two directions. Gradually, many interesting studies delve into the exploration of conducting Mamba backbone in image segmentation (Ma, Li, and Wang 2024; Xing et al. 2024), video understanding (Chen et al. 2024; Li et al. 2024a), and 3D shape (Liu et al. 2024a; Zhou et al. 2024). For language tasks, JMamba (Lieber et al. 2024) consolidates Transformer layers, Mamba layers and a mixture-of-experts (MoE) (Riquelme et al. 2021; Akbari et al. 2024) into a hybrid architecture. Similarly, several notable works make efforts in integrating or modifying components into Mamba backbone, achieving competitive performance across various language applications (Du, Liu, and Chua 2024; Jiang, Han, and Mesgarani 2024; He et al. 2024). For sequence modeling, many interesting works leverages Mamba to capture long-term dependencies and patterns in time series data (Ahamed and Cheng 2024; Patro and Agneeswaran 2024; Behrouz, Santacatterina, and Zabih 2024), tackling the challenge of attention complexity in Transformer-based methods. Given Mamba is a sequence model with powerful capabilities of long sequence modeling, at the same time the parametric CAD model can be seen as a sequence consisting of series of commands and their parameters, which motivates us make a step forward to conduct Mamba in the task of generative CAD modeling.

Method

In this section, we now introduce our Mamba-CAD model. To learn Mamba-CAD capable of encoding and decoding

CAD models, we adopt a well defined data structure namely parametric CAD sequences and make them discretized into friendly network representations, which will be fed forward into Mamba-CAD. To train our approach, we further create a new collection of 77,078 CAD models, in which the maximize length of parametric CAD sequences within the dataset is up to 128 and the average length of them is significantly longer than existing benchmark datasets of the same type, and we plan to make it publicly available shortly.

CAD sequence representation

For each CAD model, its workflow construction can be expressed by a parametric CAD sequence as each object is constructed by iteratively adding features such as line, arc, extrusion, or other CAD operations. The parametric CAD sequence S is kind of a discrete language consisting of a group of CAD commands $c_i \in C$ and their parameters $p_i \in P$, where $S = \{(c_i, p_i) \in (C, P) | G[(c_i, p_i)_1, (c_i, p_i)_2, \dots, (c_i, p_i)_N]\}$. The function of $G[*]$ is to gather (c_i, p_i) into a group in sequential order, thereby forming a complete parametric CAD sequence S . In the practical design, orderly executing the CAD commands C defined in a parametric CAD sequence S can construct the shape of each object. For the CSG representation discussed in this paper, the commonly used parametric CAD commands can be found in Figure 2. Intuitively, parametric CAD sequence is kind of complex as its different commands are coupled with various parameters that are not in the same scale. For example, the command E owns 12 parameters while the command L only owns 2 parameters. Besides, the command E is composed of discrete and continuous parameters simultaneously. These attributes increase the difficulty of using CAD commands directly for network learning. To make it friendly representation for deep learning networks, we fix the length of parametric CAD sequences up to $N = 128$ and pad those short parametric CAD sequences to achieve the length of 128 via adopting the command $\langle \text{EOS} \rangle$, the empty symbol for representing the end of a parametric CAD sequence. As seen in Figure 2, the number of parameters varies for different types of CAD commands c_i . Again, for each CAD command, we fix the number of parameters into 16 including all parameters listed in Figure 2 and leverage scalar -1 to pad unused parameters. Each CAD command is finally stacked into a 16×1 vector. To unify discrete and continuous values in parameters p_i , we refer DeepCAD (Wu, Xiao, and Zheng 2021) to quantize continuous values into a 256-dimensional vector using 8-bit integers after normalizing each CAD model into a $2 \times 2 \times 2$ cube. So far, the representation of each raw CAD model has been described in a series of discrete values.

Mamba-CAD

In this section, we first introduce the rational logic of our designed Mamba-CAD and then present details of each component in Mamba-CAD. The whole architecture of Mamba-CAD can be overviewed in Figure 3. Our objective is to generate parametric CAD sequences from randomly sampling the noise from a standard gaussian distribution.

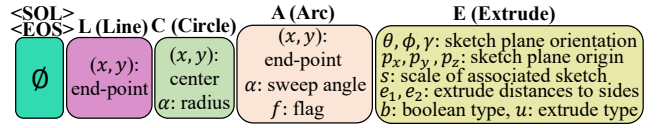


Figure 2: Common CAD commands discussed in this paper. $\langle \text{SOL} \rangle$ and $\langle \text{EOS} \rangle$ are two symbols for representing the start of one loop and the end of a whole parametric CAD sequence, respectively. f is a counter-clockwise flag.

Preliminary. The CAD models can be represented by parametric CAD sequences which are similar to discrete languages, inspiring us to leverage the sequential model to process CAD data. Mamba (Gu and Dao 2023), one of the discrete versions of recent State Space Model (SSM) shows the powerful capability of modeling on long sequences, which is very suitable for CAD models in the industry as they are relatively complex to require longer parametric CAD sequences to define. Hence, we make a step forward to explore the generative capability of Mamba on handling the complex CAD models requiring longer parametric CAD sequences. For a quick review, The origin of Mamba starts from SSM, which is conceptualized based on continuous systems with the ability to map a 1-D function or sequence $X(t) \in R \rightarrow Y(t) \in R$ via a hidden state $H(t) \in R^N$. Formally, the SSM model can be defined with Ordinary Differential Equation (ODE): $H'(t) = A \cdot H(t) + B \cdot X(t), Y(t) = C \cdot H(t)$. Next, the continuous parameters A, B in this ODE would further be discretized to parameters \mathbb{A}, \mathbb{B} via employing the Zero-Order Hold (ZOH) method with a timescale parameter Δ , as defined with:

$$\mathbb{A} = \exp(\Delta A), \quad (1)$$

$$\mathbb{B} = (\Delta A)^{-1}(\exp(\Delta A) - I) \cdot \Delta B, \quad (2)$$

$$H(t) = \mathbb{A} \cdot H(t-1) + \mathbb{B} \cdot X(t), \quad (3)$$

$$Y(t) = C \cdot H(t), \quad (4)$$

Based on the original SSM model, Mamba further introduces a Selective Scan Mechanism as its core SSM operator. For more details please refer to (Gu and Dao 2023).

Embedding. Given the parametric CAD sequence can be seen as a special discrete “language”, it inspires us to project CAD commands into an embedding space. As each CAD command is comprised by two parts: types of commands c and their parameters p , we follow (Wu, Xiao, and Zheng 2021) to separately map them into a two highly continuous space and then integrate them into a fusion embedding $E(C)$ as defined with:

$$E(C_i) = e_c^{(c_i)} + e_p^{(p_i)} + e_{pos}^i, \quad (5)$$

where $e_c^{(c_i)}$ accounts for types of CAD commands c_i , calculated through $e_c^{(c_i)} = w_{c_i} \delta_i^c$. Here, $w_{c_i} \in R^{d_E \times k}$ is a learnable matrix. $\delta_i^c \in R^k$ denotes c_i within the k command types. $e_p^{(p_i)}$ is the embedding of command parameters p_i , given by $e_p^{(p_i)} = w_{p_i}^a f(w_{p_i}^b \delta_i^p)$. $f(*)$ flattens the matrix to a vector. Each CAD command is composed of

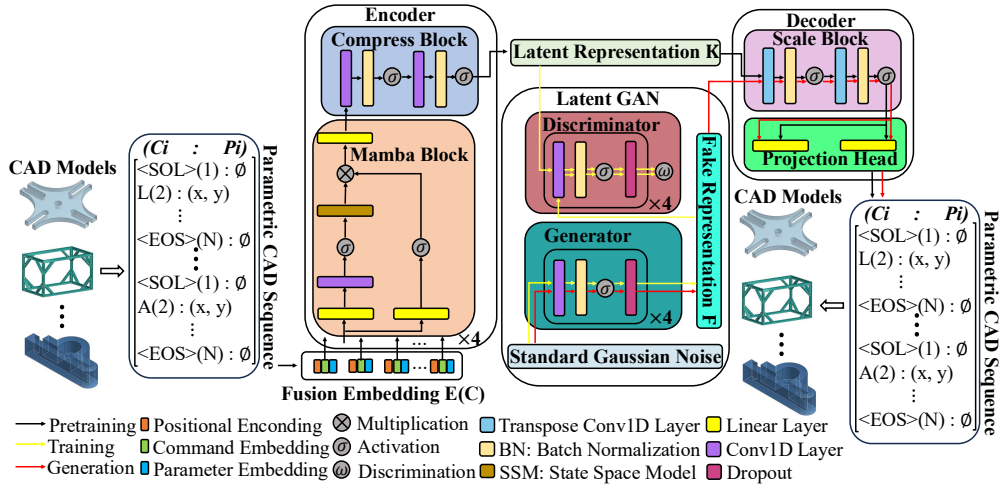


Figure 3: The overview of our Mamba-CAD architecture. The whole pipeline can be divided into three parts which marked in three different colored line-arrows. **(Pre-training stage \rightarrow)**: We leverage a task of CAD sequence reconstruction to pre-train the encoder and decoder of Mamba-CAD. Given a mini-batch of CAD-models, their corresponding parametric CAD sequences would be first discretized into friendly network representations (Sec. CAD sequence representation). Next, these representations keep going forward Fusion Embedding, Encoder, and Decoder of Mamba-CAD (Sec. Mamba-CAD) to recover predicted parametric CAD sequences. **(Training stage \rightarrow)**: When the pre-training stage is complete, the encoder parameters of Mamba-CAD would be frozen to produce the latent representation \mathbb{K} that is further used to guide a 1-D Latent-GAN training. **(Generation stage \rightarrow)**: Once the 1-D Latent GAN is well-trained, it enables to randomly sample the noise from a standard gaussian distribution and further feed it into Generator to produce the fake latent representation \mathbb{F} , which would further undergo the decoder of Mamba-CAD to recover the parametric CAD sequence.

16 parameters and each of them can be quantized into an 8-bit integer. Including one additional dimension for those unused CAD commands, the integer is converted into a one-hot vector with dimension of $2^8 + 1 = 257$, thereby stacking these 16 one-hot vectors into a matrix $\delta_i^p \in R^{257 \times 16}$. $w_{p_i}^a \in R^{d_E \times 16d_E}$ and $w_{p_i}^b \in R^{d_E \times 257}$ are learnable matrices. e_{pos}^i is the positional encoding the same as in Transformer (Vaswani et al. 2017).

Encoder. With the feed-forward strategy, a fusion embedding $E(C)$ as the input for encoder, which is composed of four Mamba Blocks, each with vocabulary size of 256 to match the dimension of a fusion embedding ($d_E = 256$). After passing forward four Mamba Blocks, each $E(C)$ would be mapped into a vector $l \in R^{B \times 256 \times L}$, where the sequence length $L = 128$ and batch size $B = 32$ are set up practically. Considering the random generation as one of our designed tasks, it needs generate a fake latent representation from a standard gaussian noise and make it close enough to l . To make l easy to learn during the generation stage, we further employ a Compress Block consisting of two 1-D Convolutional layers to reduce the feature dimension of l from 256 to 64, namely the latent representation $\mathbb{K} \in R^{B \times 64 \times L}$. According to Equation 4, the feed-forward of encoder in Mamba-CAD can be factorized as:

$$\mathbb{K} = F[\mathbb{C} \cdot H(E(C_i)) | S(6)], \quad (6)$$

where $\mathbb{C} \in R^{N \times 1}$ is a matrix projection and the function $F[*]$ denotes Compress Block. $S(6)$ is a selective mechanism of SSM model adopted in Mamba.

Decoder. In order to restore the scale of the latent repre-

sentation \mathbb{K} , we first employ a Scale Block consisting of two 1-D Transpose-Convolutional layers with Batch Normalization (Bjorck et al. 2018) to turn it back. Next, we conduct two linear layers to decode types of commands c_i^* and their parameters p_i^* separately from \mathbb{K} , and make them gathered into a final predicted parametric CAD sequence $S^* = G[(c_i^*, p_i^*)_1, (c_i^*, p_i^*)_2, \dots, (c_i^*, p_i^*)_N]$. We consider the whole Mamba-CAD as a feed-forward fashion and the output of our model can be formally defined with:

$$D(S^* | \mathbb{K}, \Psi) = \prod_{j=1}^N G[D((c_i^*, p_i^*)_j | \mathbb{K}, \Psi)], \quad (7)$$

where Ψ denotes the decoder parameters of Mamba-CAD.

Pre-training. As usual settings for pre-training an auto encoder-decoder network, we leverage a self-supervised task of CAD sequence reconstruction to train our Mamba-CAD. Specifically, the input is the parametric CAD sequence S and would be fed into Mamba-CAD and undergo encoding and decoding to recover the predicted parametric CAD sequence S^* . The training objective is to make S^* and S close enough to each other, ensuring that Mamba-CAD learns well to model the parametric CAD sequence. This is important for learning the latent representation \mathbb{K} that can transfer the knowledge of CAD models to subsequent random generation. In practical, we adopt the standard Cross-Entropy loss to minimize the distance from S^* to S . Formally, the constraint of Mamba-CAD in pre-training is defined within one mini-batch N_m :

$$\mathcal{L}_{CE} = - \sum_{j=1}^{N_m} (c_i, p_i)_j \log(c_i^*, p_i^*)_j. \quad (8)$$

The pre-training is conducted on one NVIDIA RTX 3090 GPU with the batch size of 32 under 10 epochs in about 5 hours. The initial learning rate is set to 0.001 with warm-up (He et al. 2016) and gradient clipping of 1.0 is applied in back-propagation.

Training and Generation. Once the Pre-training task is complete, we can leverage the latent representation $\mathbb{K} \in \mathbb{R}^{B \times D \times L}$ produced by Compress Block as the well-learned knowledge of CAD models to further train a 1-D Latent GAN (Chen and Zhang 2019) with 200,000 epochs and batch size of 256 in about 3 hours. To adapt to the main architecture of Mamba-CAD, we also use the same convolutional operations for Generator \mathbb{G} and Discriminator \mathbb{D} . Practically, we use both four 1-D convolutional blocks consisting of Batch Normalization and Dropout (0.25) (Baldi and Sadowski 2013) operations for \mathbb{G} and \mathbb{D} (Figure 3). Besides, we also adopt the same penalty mechanism as WGAN (Gulrajani et al. 2017; Wu, Xiao, and Zheng 2021) to stabilize training. The training objective is to enable a gaussian noise to approach the latent representation \mathbb{K} after passing forward Generator \mathbb{G} . When the network is well-trained, it enables to randomly sample the noise from a standard gaussian distribution $\mu \sim \mathcal{N}(0, 1)$ and feed it into Generator \mathbb{G} to achieve a fake latent representation \mathbb{F} , which is highly similar to \mathbb{K} . Finally, \mathbb{F} would pass forward the decoder of Mamba-CAD to recover the parametric CAD sequence, as factorized:

$$D(S^*|\mathbb{F}, \Psi) = \prod_{j=1}^N G[D((c_i^*, p_i^*)_j|\mathbb{F}, \Psi)], \quad (9)$$

where the parameters Ψ of decoder is frozen in the random generation stage. The pseudo code of Mamba-CAD can be found in the uploaded ‘‘Technical Appendix’’.

Dataset collection

Although several datasets of CAD models are publicly available, none of them can be directly used for training Mamba-CAD. Specifically, the ABC dataset (Koch et al. 2019) collected by (Onshape 2019) includes over 1 million CAD models defined with BRep representation, which lacks workflow constructions to make it difficult to be parameterized. DeepCAD (Wu, Xiao, and Zheng 2021) is a large-scale CAD dataset defined with CSG representation, consisting of 178,238 CAD models and their parametric CAD sequences. However, the maximize length of parametric CAD sequences within DeepCAD dataset is up to 60 and the average length of them is not long enough. Another Fusion360 Gallery (Willis et al. 2021) provides a sub-reconstruction dataset consisting of about 7,000 CAD models as represented in parametric CAD sequences, which is also not enough to train a generative CAD model. To break through the limitation of parametric CAD sequences within the existing benchmark datasets, we further construct a new CAD dataset as represented in parametric CAD sequences. To this end, two steps are adopted to filter and collect CAD models. First, we again filter the DeepCAD dataset by discarding CAD models in which the length of their parametric CAD sequences is less than 10. Second, we further use the feature script from (Onshape 2019) to parse CAD models within ABC dataset and translate them into parametric CAD sequences by only retaining the CAD commands used

Dataset	TN	AL	[1-10]	[11-25]	[26-40]	[41-60]	[60-128]
DeepCAD	178,238	15	44.58%	38.73%	11.38%	5.31%	-
Mamba-CAD	77,078	38	-	41.25%	26.60%	16.13%	16.02%

Table 1: The statistic comparison of between DeepCAD dataset and Our Mamba-CAD dataset, including total number (TN), average length (AL) of parametric CAD sequences, and the length distribution.

in CSG representation. Similar to DeepCAD dataset, we further remove those CAD models which use commands beyond sketch and extrusion. Next, we set the length of parametric CAD sequences up to 128 to again filter these CAD models. Finally, total 77,078 CAD models are collected as represented in parametric CAD sequences, in which 61,662 CAD models for training, 7,707 models for validation and 7,709 models for testing. This is a CAD dataset with longer parametric CAD sequences than those in the DeepCAD dataset (Wu, Xiao, and Zheng 2021). The parametric CAD commands contained in our dataset are listed in Figure 2. We also statistically analyze the length of parametric CAD sequences in our new dataset and compare it with the benchmark DeepCAD dataset, including the average length and the length distribution, as shown in Table 1. It can be found that our dataset is significantly more complex than DeepCAD dataset in both the average length and length distribution of parametric CAD sequences.

Experiments

In this section, we mainly evaluate our Mamba-CAD in two aspects: (1) modeling on parametric CAD sequences; (2) random generation of parametric CAD sequences. Our goal here is to explore the capability of Mamba-CAD as a pure unimodal framework on handling the longer parametric CAD sequence, which is an open problem for previous efforts based on the Transformer network that only use parametric CAD sequences as inputs.

Evaluation Metrics and Comparison Methods

Evaluation metrics. For evaluating the performance of modeling on parametric CAD sequences. Our model encodes each parametric CAD sequence S into a latent space, and then decode them back to predicted CAD sequence S^* , A_c and A_p are used to evaluate the difference between S and S^* as defined below:

$$A_c = \frac{1}{N} \sum_{i=1}^N \nabla [c_t^i = c_t^{i*}], \quad (10)$$

$$A_p = \frac{1}{T} \sum_{i=1}^N \sum_{j=1}^K \nabla [c_t^i = c_t^{i*}] \nabla [|c_p^{i,j} - c_p^{i,j*}| < \eta], \quad (11)$$

where $\nabla[*]$ is a boolean function with scalar 0 or 1. T is the total number of parameters in all correctly predicted commands. We set $\eta = 3$ in the experiments. The predicted sequence would probably be invalid to construct 3D shape. Therefore, we also calculate Invalid Ratio (IR) and Median Chamfer Distance (MCD) to better evaluate its rationality

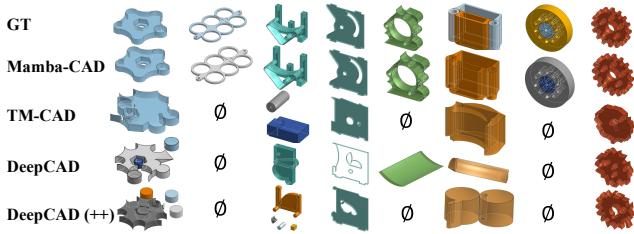


Figure 4: Visual illustrations of CAD sequence reconstruction. GT represents Ground Truth. \emptyset denotes the CAD sequence is invalid, leading to the failure of constructing a 3D shape. Please enlarge 200% for a clear view.

of shape construction. For its capability on random generation, we adopt three frequently used 3D metrics of Coverage (COV), Jensen-Shannon Divergence (JSD) and Minimum matching distance (MMD). We also report metrics of “Unique” and “Novel”. The “Unique” score is the percentage of data appears once in the generated set and the “Novel” score is the percentage of the generated data does not appear in the training set. To evaluate Mamba-CAD in CAD sequence modeling and generation comprehensively, we further calculate the average length of generated or recovered parametric CAD sequences, which are valid to construct into final 3D shapes. Considering the STEP format is widely used in CAD tools (e.g., AutoCAD, SolidWorks), we also give a metric named Step Ratio (SR), which indicates the proportion of the parametric CAD sequences that can be successfully converted to STEP format.

Comparison methods. Our primary goal is to explore the modeling capabilities of Mamba-CAD for parametric CAD sequences, rather than to prove the effectiveness of improving the modeling capability via adding additional modalities. Hence, the discussion here is based on unimodal approaches, modeling parametric CAD sequences. To make fair comparisons, all comparable models are re-trained on the Mamba-CAD dataset. In practical, we compare Mamba-CAD with following methods under several tasks to justify the effectiveness of Mamba-CAD in handling the longer parametric CAD sequence. (1) **DeepCAD** (Wu, Xiao, and Zheng 2021) is an auto-encoder framework based on Transformer network, which encodes parametric CAD sequences into a latent space and then decodes them back to construct 3D shapes. (2) **DeepCAD (++)** is an ultra version of DeepCAD, which is trained longer and composed of more Transformer layers and the larger dimension compared to DeepCAD. (3) **TM-CAD** is our another simple attempt at consolidating Transformer and Mamba into a fusion framework to process parametric CAD sequences. Practically, we leverage four Mamba Blocks to bridge the encoder and decoder of the framework based on Transformer network, where the settings of encoder and decoder are the same as those in DeepCAD. (4) **SkexGen** (Xu et al. 2022) is also a unimodal Transformer-based network but processes the “sketch” and “extrude” operations separately. (5) **HNC-CAD** (Xu et al. 2023) is a current state-of-the-art unimodal generative CAD model enabled by a two-stage

Method	$A_c \uparrow$	$A_p \uparrow$	MCD \downarrow	IR \downarrow	AL \uparrow	$L \geq 60 \uparrow$	SR \uparrow
DeepCAD	96.49	87.38	2.97	14.83%	32.52	51.82%	86.58%
DeepCAD (++)	96.92	87.68	2.81	14.05%	32.49	50.85%	87.49%
TM-CAD	95.73	85.48	4.82	15.85%	31.21	50.31%	85.61%
Mamba-CAD	99.96	99.93	0.64	5.41%	38.91	85.18%	95.43%

Table 2: The comparison of Mamba-CAD and variants in CAD sequence reconstruction. MCD is multiplied by 10^3 . A_c , A_p , and IR are multiplied by 100%. $L \geq 60$: the ratio of valid reconstructed CAD sequences (Length ≥ 60) in total CAD sequences (Length ≥ 60) within test set.

Method	$A_c \uparrow$	$A_p \uparrow$	MCD \downarrow	IR \downarrow	AL \uparrow	$L \geq 60 \uparrow$	SR \uparrow
DeepCAD	83.75	69.77	79.08	42.36%	28.72	19.11%	59.19%
DeepCAD (++)	85.95	71.27	69.68	40.19%	29.78	20.74%	61.12%
TM-CAD	80.75	67.71	82.34	44.34%	28.26	17.79%	57.23%
HNC-CAD	-	-	-	-	30.51	42.89%	74.36%
Mamba-CAD	99.85	98.91	1.03	8.03%	35.43	79.46%	92.84%

Table 3: Quantitative results of CAD sequence completion.

cascaded auto-regressive Transformer-based network. Similar to Skexgen, HNC-CAD leverages a code tree to decompose one CAD model in there parts including “loop”, “profile”, and “solid”. Among these methods, both Skexgen and HNC-CAD decompose each CAD model into a specific representation, which is the core to fit their proposed frameworks. Hence, we follow the same pre-processing strategy as adopted in Skexgen and HNC-CAD to decompose CAD models within our Mamba-CAD dataset to ensure the compatibility with their frameworks training.

Modeling on parametric CAD sequences

CAD sequence reconstruction. The ability to model the parametric CAD sequences is important for generative models, as it reflects whether the generative model can effectively represent the target distribution. Hence, we leverage a pre-training task of CAD sequence reconstruction to observe the performance of Mamba-CAD and comparable methods. Specifically, when the pre-training task is completed, 7,709 CAD models in test set would go forward networks to achieve reconstructed parametric CAD sequences. From Table 2, we can see that our Mamba-CAD surpasses other comparable methods on all metrics, especially in A_p , MCD, IR, SR, and $L \geq 60$. This demonstrates that Mamba-CAD has superior modeling capabilities for long parametric CAD sequences compared to Transformer-based networks. Figure 4 gives more visual illustrations of the recovered CAD models from reconstructed parametric CAD sequences. As shown in Figure 4, the CAD models recovered by Mamba-CAD is closer to the ground truth than other comparable methods, which again demonstrates the effectiveness of Mamba-CAD in handling longer parametric CAD sequences.

CAD sequence completion. To further evaluate the effectiveness of Mamba-CAD in learning CAD representations in the latent space, we randomly mask 40% of each parametric CAD sequence with scalar 0 in the test set and make it go forward networks to achieve complete parametric CAD sequences. Table 3 indicates that even with incomplete parametric CAD sequences as inputs, Mamba-CAD

Method	$A_c \uparrow$	$A_p \uparrow$	MCD \downarrow	IR \downarrow	SR \uparrow
DeepCAD	95.23	85.58	10.97	10.78%	90.34%
DeepCAD (++)	96.79	87.72	8.93	9.35%	91.22%
TM-CAD	90.94	83.85	12.46	12.37%	87.61%
Mamba-CAD	99.21	98.21	1.31	3.72%	96.51%

Table 4: The quantitative results of generalization ability in Fusion360 reconstruction dataset

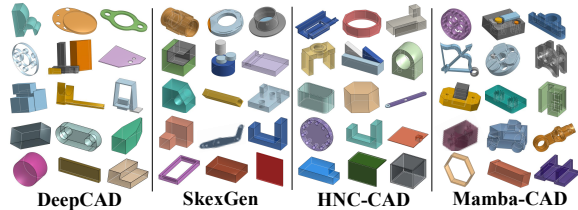


Figure 5: Visualizations of random generation of parametric CAD sequences.

still achieves precise reconstruction results, again validating the effectiveness of our method in processing longer parametric CAD sequences. Besides, The advantages of Mamba-CAD are amplified compared to other models (e.g., the metrics of IR and MCD widen the gap). This completion task again demonstrates the superior ability of Mamba-CAD in context modeling for longer parametric CAD sequences, enabling it to accurately fill in the missing parts.

Generalization ability. To further evaluate the generalization capability of Mamba-CAD, we make it first well-trained on the Mamba-CAD dataset and then directly test it on the Fusion360 reconstruction dataset that consists of about 7,000 CAD models as represented in parametric CAD sequences. Although the CAD models in this dataset is simpler (e.g., CAD sequence length ≤ 60) than those in Mamba-CAD dataset, which narrows the gap of Step Ratio between Mamba-CAD and other Transformer-based networks, Mamba-CAD still achieves the best scores in all metrics. The quantitative results can be found in Table 4.

Random generation of parametric CAD sequences

When our framework is well-trained, we randomly sample total 10,000 noise vectors from a standard gaussian distribution as fake representations of CAD models and feed them into the decoder of Mamba-CAD to generate 10,000 parametric CAD sequences. For other comparable methods 10,000 samples are random generated to compare with Mamba-CAD. To evaluate Mamba-CAD and comparable methods in random generation quantitatively, we further convert generated CAD models into point clouds and calculate the related metrics. The results can be found in Table 5. Figure 5 shows that the created CAD models from Mamba-CAD are more complex than those created by Transformer-based networks. It again proves Mamba-CAD outperforms other comparable methods in processing longer parametric CAD sequences to achieve more complex CAD models. Besides, from the showcases at the bottom of Figure 5, it shows Mamba-CAD also has the ability to generate simple CAD models with short parametric CAD sequences.

Method	MMD \downarrow	JSD \downarrow	COV \uparrow	Unique \uparrow	Novel \uparrow	AL \uparrow	SR \uparrow
DeepCAD	3.51	8.93	63.77%	83.41	51.94	27.78	62.76%
SkexGen	2.96	7.39	65.89%	93.33	60.47	29.18	63.49%
HNC-CAD	2.65	6.07	70.43%	92.18	55.32	31.36	68.84%
Mamba-CAD	1.94	4.28	75.61%	94.73	63.25	36.75	74.13%

Table 5: The quantitative results of random generation of parametric CAD sequences. MMD, JSD, Unique, and Novel are all multiplied by 10^2 .

Method	$A_c \uparrow$	$A_p \uparrow$	MCD \downarrow	IR \downarrow	AL \uparrow	$L \geq 60 \uparrow$	SR \uparrow
A	93.67	83.41	7.89	20.79%	26.93	43.57%	79.36%
B	94.23	83.95	5.67	18.04%	30.48	50.26%	82.77%
C	99.91	99.57	0.69	6.29%	38.16	84.53%	93.91%
Mamba-CAD	99.96	99.93	0.64	5.41%	38.91	85.18%	95.43%

Table 6: Ablation studies of Mamba-CAD.

Ablation study

To better understanding the mechanism of Mamba-CAD, we design three more ablation studies based on the CAD sequence reconstruction task: (A) *replace Mamba block with Transformer block*, (B) *replace CNN layers with MLPs in the Compress Block and Scale Block*, (C) *add bottleneck layer between Encoder and Decoder*. From Table 6, it shows Mamba-CAD surpasses other comparable versions, which indicates the design components in Mamba-CAD are effective and reasonable. Especially compared to (A), Mamba block brings significant improvement to the CAD sequence reconstruction task, which again proves its capability of processing longer parametric CAD sequences.

Discussion and Conclusion

Although Mamba-CAD shows a significant capability of modeling on parametric CAD sequences, there are still some limitations in our method. First, the types of commonly used CAD commands in practical designs are more than CAD commands discussed in our paper. For example, one specific curve of Bézier has not been parameterized and defined in our dataset, which imposes limitations on the topological structures of CAD models generated by our approach. Second, Mamba-CAD cannot parse CAD commands such as “fillet” and “chamfer” from BRep representation. To combine CAD commands from CSG and BRep representation for generative CAD modeling would be an innovative touch in the future work. Third, Mamba-CAD mainly focuses on the unconditioned generative modeling of longer parametric CAD sequence. However, other modalities such as texts, images, and point clouds can be integrated as an additional condition to further make the generation controllable. This is also an interesting task left for the future work.

In summary, we introduce Mamba-CAD, a deep generative model based on a Mamba backbone for CAD designs, which can handle a longer parametric CAD sequence to construct complex 3D shapes of CAD models. To achieve this, we also create a new CAD dataset consisting of 77,078 CAD models represented in parametric CAD sequences. The average length of CAD sequences in the new dataset is longer than that in benchmark datasets of the same type.

References

- Achlioptas, P.; Diamanti, O.; Mitliagkas, I.; and Guibas, L. 2018. Learning representations and generative models for 3d point clouds. In *International conference on machine learning*, 40–49. PMLR.
- Ahamed, M. A.; and Cheng, Q. 2024. Timemachine: A time series is worth 4 mambas for long-term forecasting. *arXiv preprint arXiv:2403.09898*.
- Akbari, H.; Kondratyuk, D.; Cui, Y.; Hornung, R.; Wang, H.; and Adam, H. 2024. Alternating gradient descent and mixture-of-experts for integrated multimodal perception. *Advances in Neural Information Processing Systems*, 36.
- Baldi, P.; and Sadowski, P. J. 2013. Understanding dropout. *Advances in neural information processing systems*, 26.
- Behrouz, A.; Santacatterina, M.; and Zabih, R. 2024. Mambamixer: Efficient selective state space models with dual token and channel selection. *arXiv preprint arXiv:2403.19888*.
- Bjorck, N.; Gomes, C. P.; Selman, B.; and Weinberger, K. Q. 2018. Understanding batch normalization. *Advances in neural information processing systems*, 31.
- Chen, G.; Huang, Y.; Xu, J.; Pei, B.; Chen, Z.; Li, Z.; Wang, J.; Li, K.; Lu, T.; and Wang, L. 2024. Video mamba suite: State space model as a versatile alternative for video understanding. *arXiv preprint arXiv:2403.09626*.
- Chen, Z.; and Zhang, H. 2019. Learning implicit fields for generative shape modeling. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5939–5948.
- Du, Y.; Liu, X.; and Chua, Y. 2024. Spiking structured state space model for monaural speech enhancement. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 766–770. IEEE.
- Gu, A.; and Dao, T. 2023. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*.
- Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; and Courville, A. C. 2017. Improved training of wasserstein gans. *Advances in neural information processing systems*, 30.
- Guo, H.; Liu, S.; Pan, H.; Liu, Y.; Tong, X.; and Guo, B. 2022. Complexgen: Cad reconstruction by b-rep chain complex generation. *ACM Transactions on Graphics (TOG)*, 41(4): 1–18.
- He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.
- He, W.; Han, K.; Tang, Y.; Wang, C.; Yang, Y.; Guo, T.; and Wang, Y. 2024. Densemamba: State space models with dense hidden connection for efficient large language models. *arXiv preprint arXiv:2403.00818*.
- Jayaraman, P. K.; Lambourne, J. G.; Desai, N.; Willis, K. D. D.; Sanghi, A.; and Morris, N. J. W. 2023. SolidGen: An Autoregressive Model for Direct B-rep Synthesis. *Trans. Mach. Learn. Res.*, 2023.
- Jiang, X.; Han, C.; and Mesgarani, N. 2024. Dual-path mamba: Short and long-term bidirectional selective structured state space models for speech separation. *arXiv preprint arXiv:2403.18257*.
- Jones, B. T.; Hu, M.; Kodnongbua, M.; Kim, V. G.; and Schulz, A. 2023a. Self-supervised representation learning for cad. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 21327–21336.
- Jones, R. K.; Guerrero, P.; Mitra, N. J.; and Ritchie, D. 2023b. Shapecoder: Discovering abstractions for visual programs from unstructured primitives. *ACM Transactions on Graphics (TOG)*, 42(4): 1–17.
- Khan, M. S.; Dupont, E.; Ali, S. A.; Cherenkova, K.; Kacem, A.; and Aouada, D. 2024. 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)*, 4713–4722.
- Koch, S.; Matveev, A.; Jiang, Z.; Williams, F.; Artemov, A.; Burnaev, E.; Alexa, M.; Zorin, D.; and Panozzo, D. 2019. Abc: A big cad model dataset for geometric deep learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9601–9611.
- Li, C.; Pan, H.; Bousseau, A.; and Mitra, N. J. 2022. Free2cad: Parsing freehand drawings into cad commands. *ACM Transactions on Graphics (TOG)*, 41(4): 1–16.
- Li, K.; Li, X.; Wang, Y.; He, Y.; Wang, Y.; Wang, L.; and Qiao, Y. 2024a. Videomamba: State space model for efficient video understanding. *arXiv preprint arXiv:2403.06977*.
- Li, P.; Guo, J.; Li, H.; Benes, B.; and Yan, D.-M. 2024b. SfmCAD: Unsupervised CAD Reconstruction by Learning Sketch-based Feature Modeling Operations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4671–4680.
- Li, P.; Guo, J.; Zhang, X.; and Yan, D.-M. 2023. Secadnet: Self-supervised cad reconstruction by learning sketch-extrude operations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16816–16826.
- Lieber, O.; Lenz, B.; Bata, H.; Cohen, G.; Osin, J.; Dalmedigos, I.; Safahi, E.; Meiom, S.; Belinkov, Y.; Shalev-Shwartz, S.; et al. 2024. Jamba: A hybrid transformer-mamba language model. *arXiv preprint arXiv:2403.19887*.
- Liu, J.; Yu, R.; Wang, Y.; Zheng, Y.; Deng, T.; Ye, W.; and Wang, H. 2024a. Point mamba: A novel point cloud backbone based on state space model with octree-based ordering strategy. *arXiv preprint arXiv:2403.06467*.
- Liu, Y.; Tian, Y.; Zhao, Y.; Yu, H.; Xie, L.; Wang, Y.; Ye, Q.; and Liu, Y. 2024b. Vmamba: Visual state space model. *arXiv preprint arXiv:2401.10166*.
- Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; and Guo, B. 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of*

- the *IEEE/CVF international conference on computer vision*, 10012–10022.
- Lou, Y.; Li, X.; Chen, H.; and Zhou, X. 2023. BRep-BERT: Pre-training Boundary Representation BERT with Sub-graph Node Contrastive Learning. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 1657–1666.
- Ma, J.; Li, F.; and Wang, B. 2024. U-mamba: Enhancing long-range dependency for biomedical image segmentation. *arXiv preprint arXiv:2401.04722*.
- Onshape. 2019. Onshape featurescript. <https://cad.onshape.com/FsDoc/>.
- Patro, B. N.; and Agneeswaran, V. S. 2024. Simba: Simplified mamba-based architecture for vision and multivariate time series. *arXiv preprint arXiv:2403.15360*.
- Ren, D.; Zheng, J.; Cai, J.; Li, J.; Jiang, H.; Cai, Z.; Zhang, J.; Pan, L.; Zhang, M.; Zhao, H.; et al. 2021. Csg-stump: A learning friendly csg-like representation for interpretable shape parsing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 12478–12487.
- Ren, D.; Zheng, J.; Cai, J.; Li, J.; and Zhang, J. 2022. Extrudenet: Unsupervised inverse sketch-and-extrude for shape parsing. In *European Conference on Computer Vision*, 482–498. Springer.
- Riquelme, C.; Puigcerver, J.; Mustafa, B.; Neumann, M.; Jenatton, R.; Susano Pinto, A.; Keysers, D.; and Hounsby, N. 2021. Scaling vision with sparse mixture of experts. *Advances in Neural Information Processing Systems*, 34: 8583–8595.
- Scarselli, F.; Gori, M.; Tsoi, A. C.; Hagenbuchner, M.; and Monfardini, G. 2008. The graph neural network model. *IEEE transactions on neural networks*, 20(1): 61–80.
- Uy, M. A.; Chang, Y.-Y.; Sung, M.; Goel, P.; Lambourne, J. G.; Birdal, T.; and Guibas, L. J. 2022. Point2cyl: Reverse engineering 3d objects from point clouds to extrusion cylinders. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11850–11860.
- Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.
- Vinyals, O.; Fortunato, M.; and Jaitly, N. 2015. Pointer networks. *Advances in neural information processing systems*, 28.
- Willis, K. D.; Pu, Y.; Luo, J.; Chu, H.; Du, T.; Lambourne, J. G.; Solar-Lezama, A.; and Matusik, W. 2021. Fusion 360 gallery: A dataset and environment for programmatic cad construction from human design sequences. *ACM Transactions on Graphics (TOG)*, 40(4): 1–24.
- Wu, R.; Xiao, C.; and Zheng, C. 2021. DeepCAD: A Deep Generative Network for Computer-Aided Design Models. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 6752–6762.
- Xing, Z.; Ye, T.; Yang, Y.; Liu, G.; and Zhu, L. 2024. Segmamba: Long-range sequential modeling mamba for 3d medical image segmentation. *arXiv preprint arXiv:2401.13560*.
- Xu, X.; Jayaraman, P. K.; Lambourne, J. G.; Willis, K. D.; and Furukawa, Y. 2023. Hierarchical Neural Coding for Controllable CAD Model Generation. In *International Conference on Machine Learning*, 38443–38461. PMLR.
- Xu, X.; Willis, K. D.; Lambourne, J. G.; Cheng, C.-Y.; Jayaraman, P. K.; and Furukawa, Y. 2022. SkexGen: Autoregressive Generation of CAD Construction Sequences with Disentangled Codebooks. In *International Conference on Machine Learning*.
- Yu, F.; Chen, Q.; Tanveer, M.; Mahdavi Amiri, A.; and Zhang, H. 2024. D²CSG: Unsupervised Learning of Compact CSG Trees with Dual Complements and Dropouts. *Advances in Neural Information Processing Systems*, 36.
- Yu, F.; Chen, Z.; Li, M.; Sanghi, A.; Shayani, H.; Mahdavi-Amiri, A.; and Zhang, H. 2022. CAPRI-Net: Learning Compact CAD Shapes With Adaptive Primitive Assembly. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11768–11778.
- Zhou, Q.; Yang, W.; Fei, B.; Xu, J.; Zhang, R.; Liu, K.; Luo, Y.; and He, Y. 2024. 3dmambaipf: A state space model for iterative point cloud filtering via differentiable rendering. *arXiv preprint arXiv:2404.05522*.
- Zhou, S.; Tang, T.; and Zhou, B. 2023. Cadparser: a learning approach of sequence modeling for b-rep cad. In *International Joint Conference on Artificial Intelligence*.
- Zhu, L.; Liao, B.; Zhang, Q.; Wang, X.; Liu, W.; and Wang, X. 2024. Vision mamba: Efficient visual representation learning with bidirectional state space model. *arXiv preprint arXiv:2401.09417*.