

SECAD-Net: Self-Supervised CAD Reconstruction by Learning Sketch-Extrude Operations

Pu Li^{1,2} Jianwei Guo^{1,2*} Xiaopeng Zhang^{1,2} Dong-Ming Yan^{1,2}

¹MAIS, Institute of Automation, Chinese Academy of Sciences

²School of Artificial Intelligence, University of Chinese Academy of Sciences

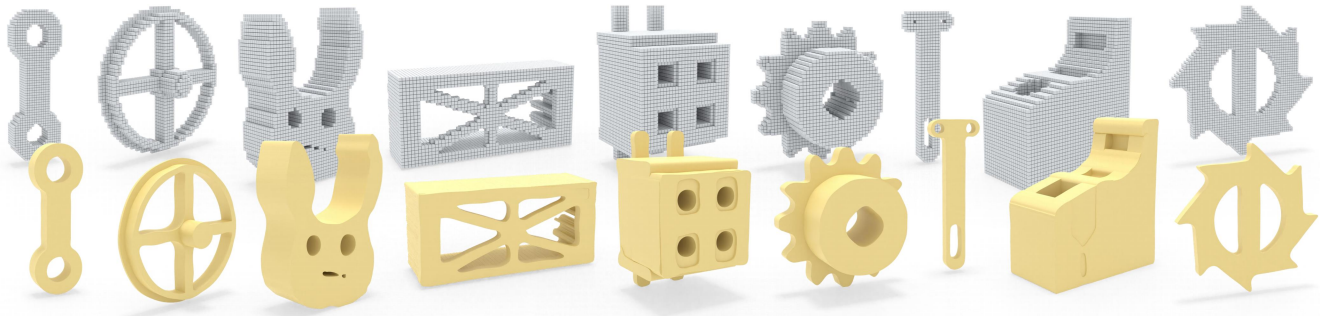


Figure 1. **SECAD-Net for CAD reconstruction.** Starting from a voxel grid (top), SECAD-Net learns *Sketch-Extrude* operations to reconstruct CAD models (bottom), without any supervision of part segmentation and sketch labels.

Abstract

Reverse engineering CAD models from raw geometry is a classic but strenuous research problem. Previous learning-based methods rely heavily on labels due to the supervised design patterns or reconstruct CAD shapes that are not easily editable. In this work, we introduce SECAD-Net, an end-to-end neural network aimed at reconstructing compact and easy-to-edit CAD models in a self-supervised manner. Drawing inspiration from the modeling language that is most commonly used in modern CAD software, we propose to learn 2D sketches and 3D extrusion parameters from raw shapes, from which a set of extrusion cylinders can be generated by extruding each sketch from a 2D plane into a 3D body. By incorporating the Boolean operation (i.e., union), these cylinders can be combined to closely approximate the target geometry. We advocate the use of implicit fields for sketch representation, which allows for creating CAD variations by interpolating latent codes in the sketch latent space. Extensive experiments on both ABC and Fusion 360 datasets demonstrate the effectiveness of our method, and show superiority over state-of-the-art alternatives including the closely related method for supervised CAD reconstruction. We further apply our ap-

proach to CAD editing and single-view CAD reconstruction. Code will be released at <https://github.com/BunnySoCrazy/SECAD-Net>.

1. Introduction

CAD reconstruction is one of the most sought-after geometric modeling technologies, which plays a substantial role in reverse engineering in case of the original design document is missing or the CAD model of a real object is not available. It empowers users to reproduce CAD models from other representations and supports the designer to create new variations to facilitate various engineering and manufacturing applications.

The advance in 3D scanning technologies has promoted the paradigm shift from time-consuming and laborious manual dimensions to automatic CAD reconstruction. A typical line of works [3, 6, 35, 47] first reconstructs a polygon mesh from the scanned point cloud, then followed by mesh segmentation and primitive extraction to obtain a boundary representation (B-rep). Finally, a CAD shape parser is applied to convert the B-rep into a sequence of modeling operations. Recently, inspired by the substantial success of point set learning [1, 32, 49] and deep 3D representations [9, 28, 30], a number of methods have exploited neural

*Corresponding author: jianwei.guo@nlpr.ia.ac.cn

networks to improve the above pipeline, *e.g.*, detecting and fitting primitives to raw point clouds directly [25, 27, 40]. A few works (*e.g.*, CSG-Net [39], UCSG-Net [19], and CSG-Stump [33]) further parse point cloud inputs into a constructive solid geometry (CSG) tree by predicting a set of primitives that are then combined with Boolean operations. Although achieving encouraging compact representation, they only output a set of simple primitives with limited types (*e.g.*, planes, cylinders, spheres), which restricts their representation capability for reconstructing complex and more general 3D shapes. CAPRI-Net [59] introduces quadric surface primitives and the difference operation based on BSP-Net [8] to produce complicated convex and concave shapes via a CSG tree. However, controlling the implicit equation and parameters of quadric primitives is difficult for designers to edit the reconstructed models. Thus, the editability of those methods is quite limited.

In this paper, we develop a novel and versatile deep neural framework, named SECAD-Net, to reconstruct high-quality and editable CAD models. Our approach is inspired by the observation that a CAD model is usually designed as a command sequence of operations [7, 38, 50, 51, 57], *i.e.*, a set of planar 2D sketches are first drawn then extruded into 3D solid shapes for Boolean operations to create the final model. At the heart of our approach is to learn the sketch and extrude modeling operations, rather than CSG with parametric primitives. To determine the position and axis of each sketch plane, SECAD-Net first learns multiple extrusion boxes to decompose the entire shape into multiple local regions. Afterward, for the local profile in each box, we utilize a fully connected network to learn the implicit representation of the sketch. An extrusion operator is then designed to calculate the implicit expression of the cylinders according to the predicted sketch and extrusion parameters. We finally apply a union operation to assemble all extrusion cylinders into the final CAD model.

Benefiting from our representation, our approach is flexible and efficient to construct a wide range of 3D shapes. As the predictions of our method are fully interpretable, it allows users to express their ideas to create variations or improve the design by operating on 2D sketches or 3D cylinders intuitively. To summarize, our work makes the following contributions:

- We present a novel deep neural network for reverse engineering CAD models with self-supervision, leading to faithful reconstructions that closely approximate the target geometry.
- SECAD-Net is capable of learning implicit sketches and differentiable extrusions from raw 3D shapes without the guidance of ground truth sketch labels.
- Extensive experiments demonstrate the superiority of SECAD-Net through comprehensive comparisons. We

also showcase its immediate applications to CAD interpolation, editing, and single-view reconstruction.

2. Related work

Neural implicit representation. 3D shapes can be represented either *explicitly* (*e.g.*, point sets, voxels, meshes) or *implicitly* (*e.g.*, signed-distance functions, indicator functions), each of them comes with its own advantages and drawbacks. Recently, there is an explosion of neural implicit representations [9, 28, 30] that allow for generating detail-rich 3D shapes by predicting the underlying signed distance fields. Thanks to the ability to learn priors over shapes, many deep implicit works have been proposed to solve various 3D tasks, such as shape representation and completion [2, 10, 42], image-based 3D reconstruction [45, 52, 58], shape abstraction [16, 44] and novel view synthesis [12, 29]. Theoretically, any of the above shape representations can be used to represent sketches. However, primitive-based methods usually suppress the ability cap of shape representation. In this work, we choose to fit an implicit sketch representation using a neural network, and show its superiority over other representations (*e.g.*, BSP [8]) in the ablation study, see Sec. 5.4.

Reverse engineering CAD reconstruction. Over the past decades, reverse engineering has been extensively studied; it aims at converting measured data (a surface mesh or a point cloud) into solid 3D models that can be further edited and manufactured by industries. Traditional approaches addressing this problem consist of the following tasks: (1) segmentation of the point clouds/meshes [5, 41, 60], (2) fitting of parametric primitives to segmented regions [11, 36, 55], (3) finishing operations for CAD modeling [4, 24]. Important drawbacks of these conventional methods are the time-consuming process and the requirement of a skilled operator to guide the reconstruction [6].

With the release of several large-scale CAD datasets (*e.g.*, ABC [21], Fusion 360 [50]), SketchGraphs [37]), numerous approaches have explored deep learning to address primitive segmentation/detection [25, 56], parametric curve or surface inference from point clouds [17, 27, 31, 40, 48] or B-rep models [18, 23]. However, by only outputting individual curves or surfaces, these methods lack the CAD modeling operations that are needed to build solid models. Focusing on CAD generation rather than reconstruction task as ours, some approaches propose deep generative models that predict sequences of CAD modeling operations to produce CAD designs [26, 50, 51, 53, 54]. Aiming at CAD reconstruction involving inverse CSG modeling [15], CSGNet [39] first develops a neural model that parses a shape into a sequence of CSG operations. More recent works follow the line of CSG parsing by advancing the inference without any supervision [19], or improving representation capability with a three-layer reformulation of the classic

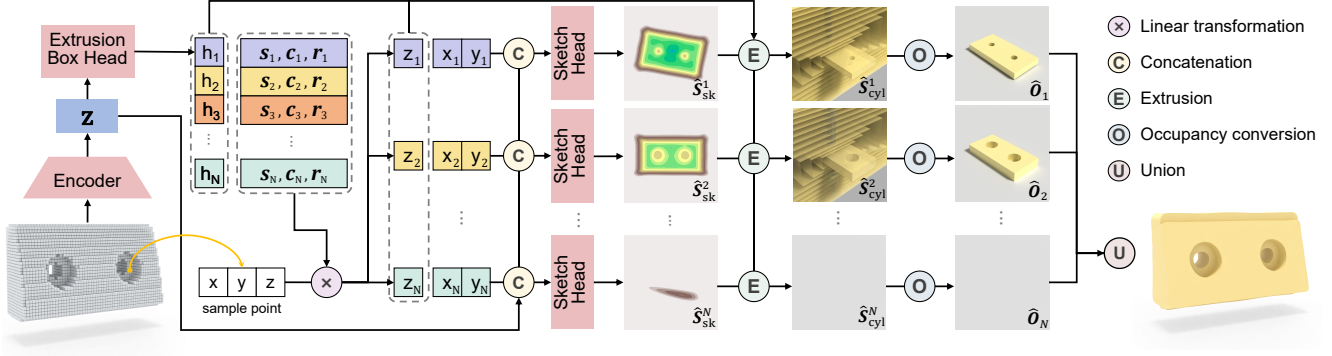


Figure 2. **Network architecture for SECAD-Net:** The embedding \mathbf{z} encoded from the voxel input is first fed to the extrusion box head to predict extrusion boxes. It is also sent to the sketch head network to calculate the sketch SDF \hat{S}_{sk}^i after concatenating with the linear transformed sampling point. \hat{S}_{cyl}^i stands for the SDF of the cylinder, which is acquired by extruding \hat{S}_{sk}^i with height h_i . Then we convert \hat{S}_{cyl}^i to occupancy of cylinder \hat{O}_i and finally obtain the complete shape by union all the occupancies.

CSG-tree [33], or handling richer geometric and topological variations by introducing quadric surface primitives [59]. While achieving high-quality reconstruction, CSG tends to combine a large number of shape primitives that are not as flexible as the extrusions of 2D sketches and are also not easily user edited to control the final geometry.

Motivated by modern design tools, supervised methods are proposed [22, 46] utilizing the sketch-extrude procedural models and learning 2D sketches that can be extruded to 3D shapes. In contrast to their reliance on 2D labels, SECAD-Net is trained in a self-supervised manner. Most closely related to our work is ExtrudeNet [34]. SECAD-Net distinguishes itself from ExtrudeNet in several significant aspects: i) Following the traditional reconstruction process, ExtrudeNet first predicts the parameters of Bézier curves and then converts them into SDFs. In contrast, we jumped out of this paradigm and directly used neural networks to predict the 2D implicit fields of the profiles. ii) ExtrudeNet adopts closed Bézier curves to avoid self-intersection in sketches. This makes ExtrudeNet can only predict star-shaped profiles, which limits the expressive power of their CAD shapes. Our method does not impose any restrictions on the shape of the profile, thus having greater flexibility in shape expression. iii) To pursue the reconstruction effect, ExtrudeNet relies on a larger number of primitives, while our method is able to predict more compact CAD shapes.

3. Problem Statement and Overview

In this section, we present an overview of the proposed approach. To precisely explain our techniques, we first provide the definition of several related terminologies (Fig. 3).

3.1. Preliminaries

Definition 1 (Loop, Profile and Sketch) *In CAD terminology, a sketch is represented by a collection of geometric*

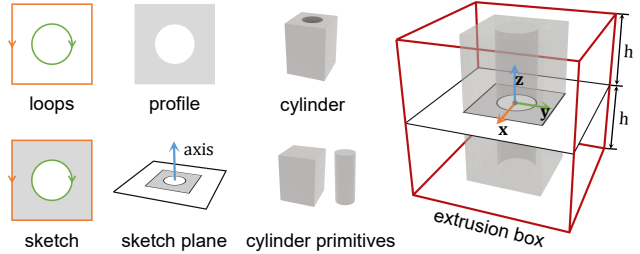


Figure 3. Definitions of CAD terminologies used in this paper. Note that the axis of the sketch plane in the figure is the same as the z-axis in the extrusion box.

primitives. By referring to a closed curve as a loop and an enclosed region composed of one or multiple inner/outer loops as a profile, we define a sketch as the collection of one profile and its loops.

Definition 2 (Sketch plane and Extrusion box) *A sketch plane is a finite plane with width w and length l , containing one or more sketches with the same extrusion height h . Then we define an extrusion box as a cuboid with the sketch plane as the base and $2h$ as the height.*

Definition 3 (Cylinder primitive and Cylinder) *In this work, a cylinder refers to the shape obtained by extruding a sketch, and a cylinder primitive is obtained by performing an extrude operation on a closed area formed by a loop. A cylinder may contain one cylinder primitive or be obtained from several cylinder primitives through the Difference operation used in CSG modeling.*

3.2. Overview

We formulate the problem of CAD reconstruction as sketch and extrude inference: taking an input 3D shape,

SECAD-Net aims to reconstruct the CAD model by predicting a set of geometric proxies that are decomposed to sketch-extrude operations. The overall pipeline of SECAD-Net is visualized in Fig. 2. Given a 3D voxel model, we first map it into a latent feature embedding \mathbf{z} by using an encoder based on a 3D convolutional network. An extrusion box head network is then applied to predict the parameters of the sketch planes from \mathbf{z} . We employ N sketch head network to independently learn N 2D signed distance fields (SDFs) as the implicit representation of a sketch. Next, we design a differentiable extrusion operator to calculate the SDF of the 3D cylinder primitives corresponding to the sketches. Finally, an occupancy transformation operation and a union operation transform the multiple SDFs into the full 3D reconstructed shape as the output of the network.

4. Method

4.1. Sketch-Extrude Inferring

We apply a standard 3D CNN encoder to extract a shape code \mathbf{z} with size 256 from the input voxel. The code is then passed to the proposed SECAD-Net to output the sketch and extrusion cylinder parameters. Below we introduce the main modules of SECAD-Net following the order of data transmission during prediction.

Extrusion box prediction. We first apply a fully connected layer to predict the parameters of the extrusion boxes. Taking the feature encoding \mathbf{z} as input, a decoder, referred to *sketch box head*, outputs a set of sketch boxes $\mathcal{B} = \{\mathbf{s}_i, \mathbf{c}_i, \mathbf{r}_i \mid i \in N\}$, where $\mathbf{s}_i \in \mathbb{R}^3$ describes the 2D size (i.e., length and width) of the box, $\mathbf{c}_i \in \mathbb{R}^3$ represents the predicted position of the box’s center, and $\mathbf{r}_i \in \mathbb{R}^4$ is the rotation quaternion. The positive z-axis of the extrusion box determines the axial direction \mathbf{e}_i of the sketch plane, and the height of the extrusion box is twice the height of the extrude operation (see Fig. 3).

2D sketch inference. The sketches in each sketch plane depict the shape contained within the sketch box. Inspired by the recent neural implicit shapes [2, 46], we encode the shape of each sketch into a sketch latent space. To this end, we first project the 3D sampling points *w.r.t* the corresponding occupancy value onto the sketch plane along the axis \mathbf{e}_i . A *sketch head network* (SK-head) then computes the signed distance from each sampling point to the sketch contour. The distance is negative for points in the sketch and positive for points outside. Each SK-head contains N_{lay} layers of fully connected layers, with softplus activation functions used between layers, and we clamp the output distance to $[-1, 1]$ in the last layer. Each 2D point is concatenated with the feature encoding \mathbf{z} as a global condition before being fed into the SK-head. Regarding the i -th SK-head as an implicit function f_i , then formally we get:

$$\hat{\mathcal{S}}_{\text{sk}}^i = f_i(\mathbf{x}_i^t, \mathbf{z}), \quad (1)$$

where \mathbf{x}_i^t is the result of a linear transformation of the sampling points contained in the i -th extrusion box, which can be expressed as $\mathbf{r}_i^{-1}(\mathbf{x}_i - \mathbf{c}_i)$. $\hat{\mathcal{S}}_{\text{sk}}^i$ represents the signed distance field of the i -th sketch plane.

Differentiable extrusion. Next, we calculate the SDF of a cylinder based on the 2D distance field and the extrusion height h . We denote Ω as the volume between two hyperplanes p^u and p^l , where p^u and p^l are the upper and lower surfaces on which the cylinder is located. Similarly, we define Ψ as the volume inside the infinite cylinder where the side of the cylinder is located. The implicit field of the i -th cylinder, $\hat{\mathcal{S}}_{\text{cyl}}^i$, is equal to one of the following cases: (1) the distance from a point \mathbf{x}_i to p^u or p^l , when $\mathbf{x}_i \in \Omega \cap \Psi^c$, where superscript c stands for complement; (2) $\hat{\mathcal{S}}_{\text{sk}}^i$, when $\mathbf{x}_i \in \Omega^c \cap \Psi$; (3) the distance from \mathbf{x}_i to the intersection curves of the cylinder and hyperplanes, when $\mathbf{x}_i \in \Omega \cap \Psi^c$; (4) the maximum distance between $\hat{\mathcal{S}}_{\text{sk}}^i$ and the point to p^u or p^l , when $\mathbf{x}_i \in \Omega^c \cap \Psi$. The sub-formulas for each case are as follows:

$$\hat{\mathcal{S}}_{\text{cyl}}^i = \begin{cases} \max(\hat{\mathcal{S}}_{\text{sk}}^i, |\mathbf{x}_{i_z}| - h_i) & , (\hat{\mathcal{S}}_{\text{sk}}^i \leq 0) \wedge (|\mathbf{x}_{i_z}| \leq h_i) \\ |\mathbf{x}_{i_z}| - h_i & , (\hat{\mathcal{S}}_{\text{sk}}^i \leq 0) \wedge (|\mathbf{x}_{i_z}| > h_i) \\ \hat{\mathcal{S}}_{\text{sk}}^i & , (\hat{\mathcal{S}}_{\text{sk}}^i > 0) \wedge (|\mathbf{x}_{i_z}| \leq h_i) \\ \left\| \hat{\mathcal{S}}_{\text{sk}}^i, (|\mathbf{x}_{i_z}| - h_i) \right\|_2 & , (\hat{\mathcal{S}}_{\text{sk}}^i > 0) \wedge (|\mathbf{x}_{i_z}| > h_i) \end{cases} \quad (2)$$

Combining the above four sub-formulas with the *max* and *min* operations, the following result is obtained:

$$\hat{\mathcal{S}}_{\text{cyl}}^i = \min(\max(\hat{\mathcal{S}}_{\text{sk}}^i, |\mathbf{x}_{i_z}| - h_i), 0) + \left\| \max(\hat{\mathcal{S}}_{\text{sk}}^i, 0), \max(|\mathbf{x}_{i_z}| - h_i, 0) \right\|_2 \quad (3)$$

Occupancy conversion and assembly. The occupancy function represents points inside the shape as 1 and points outside the shape as 0, which can be transformed by SDF. Following [13, 33], we use the Sigmoid function to perform differentiable transformation operations:

$$\hat{\mathcal{O}}_i = \text{Sigmoid}(-\eta \cdot \hat{\mathcal{S}}_{\text{cyl}}^i). \quad (4)$$

We finally assemble the occupancy $\hat{\mathcal{O}}_i$ of each cylinder to obtain the reconstructed shape. In order to express complex shapes, many works use intersection, union, and difference operations in CSG in the assembly stage [8, 19, 33, 59]. In contrast to them, we only use the union operation, because the extrusion cylinders can naturally represent concave shapes. This helps us avoid designing intricate loss functions or employing multi-stage training strategies without losing the flexibility of reconstructing shape representations. We adopt the Softmax to compute the union operation as it is shown to be effective in avoiding vanishing gradients [33]:

$$\hat{\mathcal{O}}_{\text{total}} = \sum_i^N \text{Softmax}(\varphi \cdot \hat{\mathcal{O}}_i) \cdot \hat{\mathcal{O}}_i, \quad (5)$$

where φ is the modulating coefficient and $\hat{\mathcal{O}}_{total}$ is the occupancy representation of the final reconstructed shape.

4.2. Loss Function

We train SECAD-Net in a self-supervised fashion through the minimization of the sum of two objective terms. The supervision signal is mainly quantified by the reconstruction loss, which measures the mean squared error between the predicted shape occupancy $\hat{\mathcal{O}}_{total}$ and the ground truth \mathcal{O}_{total}^* :

$$\mathcal{L}_{recon} = \mathbb{E}_{x \in \mathbf{X}} \left[(\hat{\mathcal{O}}_{total} - \mathcal{O}_{total}^*)^2 \right], \quad (6)$$

where x is a randomly sampled point in the shape volume.

However, we find that applying only \mathcal{L}_{recon} makes the network always learn fragmented cylinders. To tackle this problem, we design a 2D sketch loss to facilitate the network to learn the axis of the sketch plane and the complete profile. Specifically, each sketch plane cuts the voxel model to form an occupancy cross-section \mathcal{O}_{cs}^{i*} . We project the 3D sampling points inside the i -th extrusion box \mathcal{B}^i onto the sketch plane along the axial direction, and calculate the difference between the occupancy value of the projected points $\hat{\mathcal{O}}_{proj}$ and ground truth \mathcal{O}_{cs}^{i*} :

$$\mathcal{L}_{sketch} = \sum_{i=1}^N \mathbb{E}_{x \in \mathcal{B}^i} \left[(\hat{\mathcal{O}}_{proj}^i - \mathcal{O}_{cs}^{i*})^2 \right]. \quad (7)$$

The overall objective of SECAD-Net is defined as the combination of the above two terms:

$$\mathcal{L}_{total} = \mathcal{L}_{recon} + \lambda \mathcal{L}_{sketch}, \quad (8)$$

where λ is a balance factor.

4.3. CAD Reconstruction

The output of SECAD-Net during the training phase is an implicit occupancy function of the 3D shape. In the prediction stage, we reconstruct CAD models by using sketch-extrude operations instead of the marching cubes method.

Sketch and extrusion. To convert a 2D implicit field (Fig. 4 (a)) in the sketch latent space into an editable sketch, we input uniform 2D sampling points to the SK-head, and attach the implicit value to the position of the sampling point to obtain an explicit image-like 2D profile (Fig. 4 (b)). We then use the Teh-Chin chain approximation [43] to extract the contours of the profiles and the hierarchical relationships between them. We further apply Dierckx’s fitting [14] to convert the contours into closed B-splines (Fig. 4 (c)).

After extruding each sketch to get the cylinder primitives according to half the height of \mathcal{B}^i , we assemble cylinder primitives into cylinders by alternately performing union or difference operations according to the hierarchical relationship between contours (primitive at hierarchy 0 *difference*

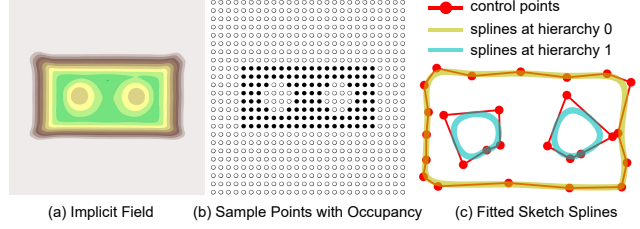


Figure 4. Illustration of converting 2D implicit sketches into the closed B-splines.

primitives at hierarchy 1 in the case of Fig. 4). Finally, we take the union of all cylinders to obtain the CAD model.

Post-processing. We take two post-processing operations to clean up overlapping and shredded shapes in the result. First, for any two cylinders, when their overlapping coefficient is greater than 0.95, the smaller of them is discarded. Second, we delete all cylinders whose height is less than 0.01 in the reconstruction result. We demonstrate our final reconstructions in Fig. 5 and Fig. 6.

4.4. Implementation Details

SECAD-Net is implemented in PyTorch and trained on a TITAN RTX GPU from NVIDIA®. We train our model using an Adam optimizer [20] with learning rate 1×10^{-4} and beta parameters (0.5, 0.99). We set both the number of MLP layers in the sketch head network and the number of output cylinders to 4. For hyper-parameters in Eq. 4, Eq. 5 and Eq. 8, we set $\eta = 150$, $\varphi = 25$ and $\lambda = 0.01$ in default, which generally works well in our experiments. Employing a similar training strategy to [59], we first pre-train SECAD-Net on the training datasets for 1,000 epochs using batch size 24, which takes about 8 hours, and fine-tuning on each test shape for 300 epochs, which takes about 3 minutes per shape.

5. Experimental Results

In this section, we examine the performance of SECAD-Net on the ABC dataset [21] and Fusion 360 Gallery [50]. Through extensive comparisons and ablation studies, we demonstrate the effectiveness of our approach and show its superiority over state-of-the-art reference approaches for CAD reconstruction.

5.1. Setup

Dataset preparation. For the ABC dataset, the voxel grids and sampling point data are provided by [59]. We use 5,000 groups of data for training and 1,000 for testing. For Fusion 360, which does not contain available voxels, we first randomly select 6,000 meshes, then discretize them into internally filled voxels. The train-test split is the same as ABC. We obtain sampling points with the corresponding occu-

Table 1. Quantitative comparison between reconstruction results on ABC dataset.

Methods	CD↓	ECD↓	NC↑	#P↓
UCSG-Net [19]	1.849	1.255	0.820	12.84
CSG-Stump [33]	4.031	0.754	0.828	17.18
ExtrudeNet [34]	0.471	0.914	0.852	14.46
Ours	0.330	0.724	0.863	4.30

Table 2. Quantitative comparison between reconstruction results on Fusion 360 dataset.

Methods	CD↓	ECD↓	NC↑	#P↓
UCSG-Net [19]	2.950	5.277	0.770	10.84
CSG-Stump [33]	2.781	4.590	0.744	12.08
Point2Cyl [46]	13.889	14.657	0.669	2.76
ExtrudeNet [34]	2.263	3.558	0.819	15.72
Ours	2.052	3.282	0.803	5.44

pancy value following [9]. The resolution of voxel shapes is 64^3 for both datasets, and the number of sampling points is 8,192. Considering that fine-tuning each method and generating high-accuracy meshes is time-consuming, we take 50 shapes from each dataset to form 100 shapes for quantitative evaluation.

Evaluation metrics. For quantitative evaluations, we follow the metrics that are commonly used in previous methods [33, 59], including symmetric Chamfer Distance (*CD*), Normal Consistency (*NC*), Edge Chamfer Distance (*ECD*). Details of computing these metrics are given in the supplemental materials. Additionally, we also report the number of generated primitives, *#p*, as a measure of how easy the output CAD results are to edit.

5.2. Comparison on CAD Reconstruction

We thoroughly compare our method with two types of primitive-based CAD reconstruction methods that output editable CAD models, including two CSG-like methods (*i.e.*, UCSG-Net [19], CSG-Stump [33]) and two cylinder decomposition counterpart (*i.e.*, point2Cyl [46]), ExtrudeNet [34]. For each method, we adopt the implementation provided by the corresponding authors, and use the same training strategy for training and fine-tuning. For CSG-Stump, we set the number of intersection nodes to 64, making it output a comparable number of primitives to other methods. Those methods provide a plethora of comparisons to other techniques and establish themselves as state-of-the-art. Note that for point2Cyl, we only report its results on Fusion 360 dataset, as the ABC dataset does not provide the labels needed to train point2Cyl.

Quantitative results on the ABC and Fusion 360 datasets are reported in Table 1 and Table 2, respectively. It can be

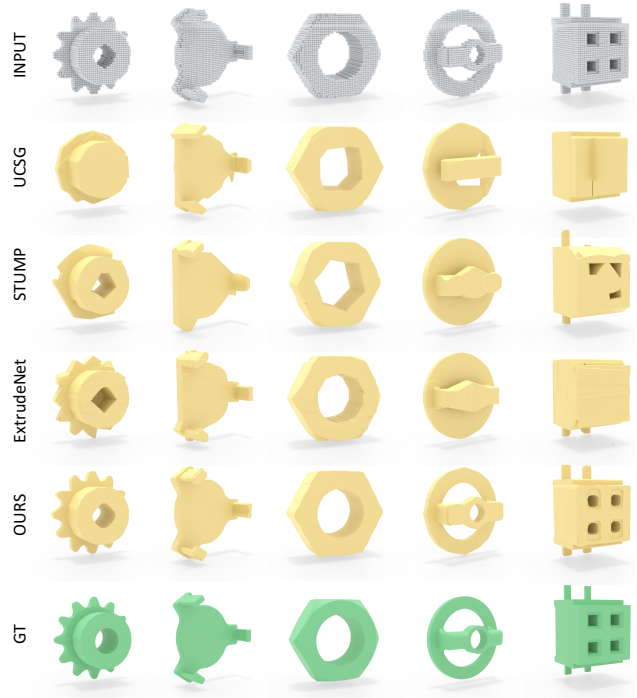


Figure 5. Visual comparison between reconstruction results on ABC dataset.

seen that the proposed SECAD-Net outperforms both kinds of methods on all evaluation metrics while still generating a relatively small number of primitives. Fig. 5 and Fig. 6 display several qualitative comparison results. For fairness, all the reconstructed CAD models are visualized using marching cubes (MC) with 256 resolution. Visually as shown in the figures, our method achieves much better geometry and topological fidelity with more accurate structures (*e.g.*, holes, junctions) and sharper features.

5.3. CAD Generation via Sketch Interpolation

Although without ground truth labels as guidance, SECAD-Net can learn plausible 2D sketches from raw 3D shapes. Thanks to the implicit sketch representation, we are able to generate different CAD variations when a pair of shapes is interpolated in the complete and continuous sketch latent space, as shown in Fig. 7. The results suggest that the generated sketch is gradually transformed even if the pair of shapes have significantly different structures, and we draw two further conclusions: (1) the predicted position of each extrusion box is relatively deterministic, although the input shape is different (see the left and right column sketches in the leftmost group); (2) when an extrusion box does not contain a shape, our SK-head does not generate a sketch, making the network output an adaptive number of cylinders (see the middle column sketches of the leftmost and the rightmost groups).

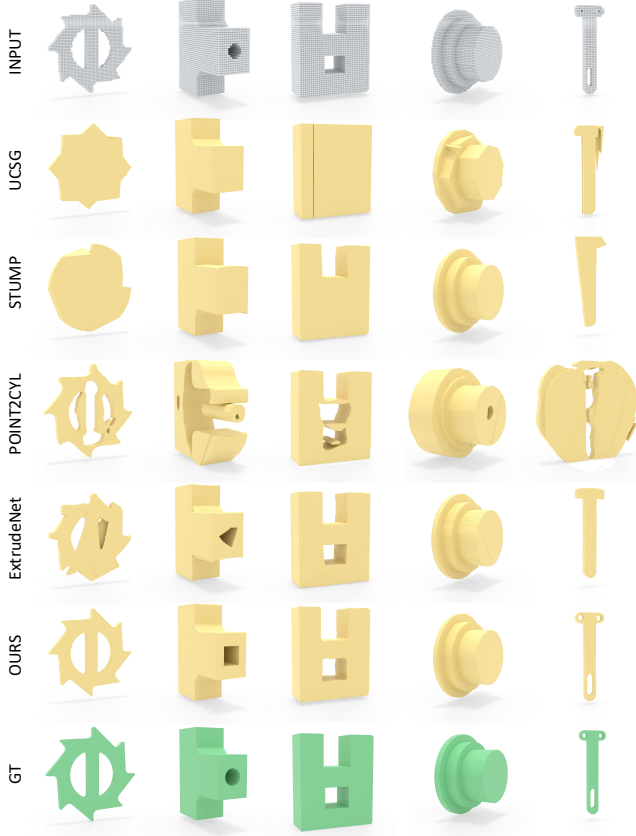


Figure 6. Visual comparison between reconstruction results on Fusion 360 dataset.

Table 3. Ablation study on network design and sketch loss. We adopted setting (e) in the final model.

Settings	(a)	(b)	(c)	(d)	(e)
N_{sh}	1	4	4	4	4
N_{lay}	2	2	2	4	4
N_{cyl}	4	4	8	4	4
\mathcal{L}_{sketch}	✓	✓	✓	✗	✓
CD↓	2.627	0.993	1.504	0.336	0.330
ECD↓	1.754	0.882	1.098	0.772	0.724
NC↑	0.713	0.835	0.761	0.863	0.863

5.4. Ablations

We perform ablation studies to carefully analyze the efficiency of major components of our designed model. All quantitative metrics are measured on the ABC dataset.

Effect of network design and sketch loss. We first examine the effect of the number of components/parameters in SECAD-Net, including the number of SK-heads (N_{sh}), the number of fully connected layers (N_{lay}) in each SK-head, and the number of output cylinders (N_{cyl}). Then we show the necessity of the sketch loss by deactivating it to train the

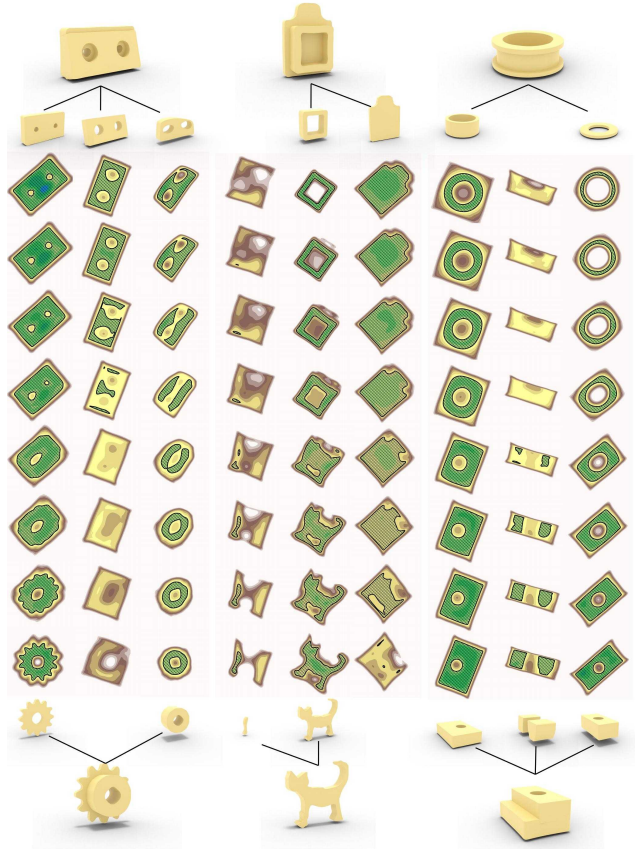


Figure 7. For each example, we encode the sketches of top and bottom shapes in latent vector space and then linearly interpolate the corresponding latent codes.

Table 4. Ablation study on sketch representation.

Representation	CD↓	ECD↓	NC↑	#P↓
Box primitives	0.523	0.982	0.825	5.38
BSP	0.612	0.838	0.852	5.84
SK-head (Ours)	0.330	0.724	0.863	4.30

network. The quantified results are presented in Table 3. Settings (a), (b), and (c) show that reducing the number of SK-heads or increasing the number of cylinder outputs will damage the model prediction accuracy. Settings (b), (d), and (e) show that increasing the number of MLP layers in SK-head or enabling \mathcal{L}_{sketch} will improve the prediction accuracy.

Effect of implicit sketch representation. To assess the efficiency of neural implicit sketch representation, we adopt two other classical shape representations, namely binary space partitioning (BSP [8]) and box-like primitives, to compare with our SK-head in SECAD-Net. For BSP, we set the number of output convex shapes to 8, each containing 12 partitions. The assembly method is consistent with [8] to

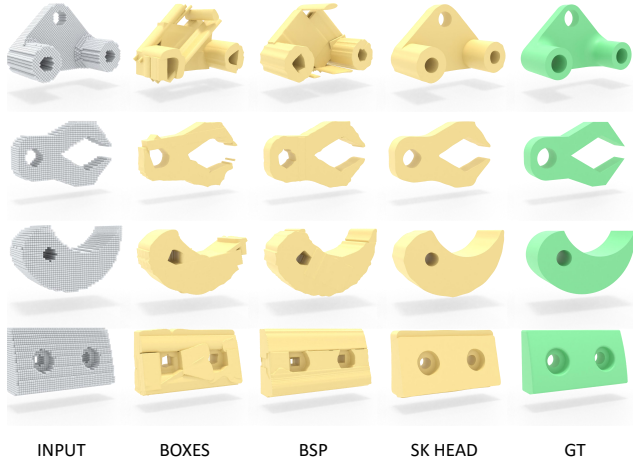


Figure 8. Visual comparison results for ablation study on sketch representation.

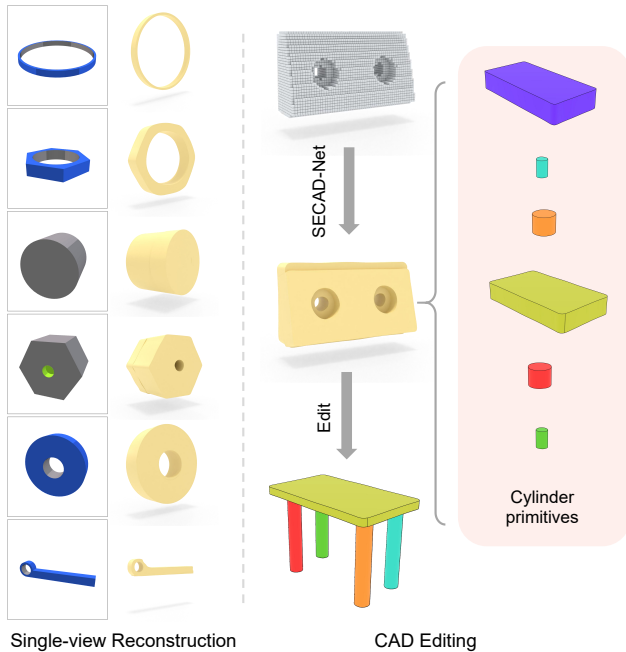


Figure 9. SECAD-Net can aid in more applications. Left: the results of single-view reconstruction. Right: a subsequent CAD editing by changing the predicted cylinder primitives.

represent 2D sketches. For box-like primitives, 24 rectangles are predicted. We divide them into two subsets in half, take the union operation separately, and subtract the other from one of the union results. The numerical and visual comparison results are shown in Table 4 and Fig. 8, respectively. It can be seen that our implicit field can represent the smoothest shape while obtaining the best reconstruction results.

5.5. Other Applications

By replacing the voxel encoder, SECAD-Net is flexible to reconstruct CAD models from other input shape representations, *e.g.*, images and point clouds. Fig. 9 shows the results of SECAD-Net in solving single-view reconstruction (SVR) task. Following the training strategy of previous work [8, 9], we first use voxel data to complete the training of the 3D auto-encoding task, and then train an image encoder with the feature encoding of each shape as the target. The voxels and input images used for the SVR task are obtained directly from Fusion 360. Replacing more input representations, while feasible and meaningful, is not the focus of this paper and we leave it to future research.

Finally, the parameters of both 2D sketches and 3D cylinders are available, thus the CAD results output from SECAD-Net can be directly loaded into existing CAD software for further editing. As shown in the right side of Fig. 9, interpretable CAD variations can be produced via specific editing operations, such as sketch-level curve editing, primitive-level displacement, rotation, scaling, and Boolean operations between primitives.

6. Conclusion and Future Work

We have presented a novel neural network that successively learns shape sketch and extrusion without any expensive annotations of shape segmentation and labels as the supervision. Our approach is able to learn smooth sketches, followed by the differentiable extrusion to reconstruct CAD models that are close to the ground truth. We evaluate SECAD-Net using diverse CAD datasets and demonstrate the advantages of our approach by ablation studies and comparing it to the state-of-the-art methods. We further demonstrate our method’s applicability in single-image CAD reconstruction. Additionally, the CAD shapes generated by our approach can be directly fed into off-the-shelf CAD software for sketch-level or cylinder primitive-level editing.

In future work, we plan to extend our approach to learn more CAD-related operations such as *revolve*, *bevel*, and *sweep*. Besides, we find that current deep learning models perform poorly on datasets with large differences in shape geometry and structure. Therefore, another promising direction is to explore how to improve the generalization of neural networks and enhance the realism of the generated shapes by learning structural and topological information.

Acknowledgments. We thank the anonymous reviewer for their valuable suggestions. This work is partially funded by the National Natural Science Foundation of China (U22B2034, 62172416, U21A20515, 62172415), and the Youth Innovation Promotion Association of the Chinese Academy of Sciences (2022131).

References

- [1] Panos Achlioptas, Olga Diamanti, Ioannis Mitliagkas, and Leonidas Guibas. Learning representations and generative models for 3d point clouds. In *International conference on machine learning*, pages 40–49. PMLR, 2018. 1
- [2] Matan Atzmon and Yaron Lipman. SAL: Sign agnostic learning of shapes from raw data. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 2565–2574, June 2020. 2, 4
- [3] Roseline Bénéière, Gérard Subsol, Gilles Gesquière, François Le Breton, and William Puech. A comprehensive process of reverse engineering from 3d meshes to cad models. *Computer-Aided Design*, 45(11):1382–1393, 2013. 1
- [4] Pál Benkő, Ralph R Martin, and Tamás Várady. Algorithms for reverse engineering boundary representation models. *Computer-Aided Design*, 33(11):839–851, 2001. 2
- [5] Pál Benkő and Tamás Várady. Segmentation methods for smooth point regions of conventional engineering objects. *Computer-Aided Design*, 36(6):511–523, 2004. 2
- [6] Francesco Buonamici, Monica Carfagni, Rocco Furferi, Lapo Governi, Alessandro Lapini, and Yary Volpe. Reverse engineering modeling methods and tools: a survey. *Computer-Aided Design and Applications*, 15(3):443–464, 2018. 1, 2
- [7] Dan Cascaval, Mira Shalah, Phillip Quinn, Rastislav Bodik, Maneesh Agrawala, and Adriana Schulz. Differentiable 3d cad programs for bidirectional editing. In *Computer Graphics Forum*, volume 41, pages 309–323, 2022. 2
- [8] Zhiqin Chen, Andrea Tagliasacchi, and Hao Zhang. BSP-Net: Generating compact meshes via binary space partitioning. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 45–54, 2020. 2, 4, 7, 8
- [9] Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 5939–5948, 2019. 1, 2, 6, 8
- [10] Julian Chibane, Thiemo Alldieck, and Gerard Pons-Moll. Implicit functions in feature space for 3d shape reconstruction and completion. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 6970–6981, 2020. 2
- [11] David Cohen-Steiner, Pierre Alliez, and Mathieu Desbrun. Variational shape approximation. In *ACM SIGGRAPH 2004 Papers*, pages 905–914, 2004. 2
- [12] Frank Dellaert and Lin Yen-Chen. Neural volume rendering: Nerf and beyond. *arXiv preprint arXiv:2101.05204*, 2020. 2
- [13] Boyang Deng, Kyle Genova, Soroosh Yazdani, Sofien Bouaziz, Geoffrey Hinton, and Andrea Tagliasacchi. CvxNet: Learnable convex decomposition. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 31–44, 2020. 4
- [14] Paul Dierckx. Algorithms for smoothing data with periodic and parametric splines. *Computer Graphics and Image Processing*, 20(2):171–184, 1982. 5
- [15] Tao Du, Jeevana Priya Inala, Yewen Pu, Andrew Spielberg, Adriana Schulz, Daniela Rus, Armando Solar-Lezama, and Wojciech Matusik. Inversecsg: Automatic conversion of 3d models to csg trees. *ACM Trans. Graph.*, 37(6):1–16, 2018. 2
- [16] Kyle Genova, Forrester Cole, Daniel Vlasic, Aaron Sarna, William T Freeman, and Thomas Funkhouser. Learning shape templates with structured implicit functions. In *IEEE International Conference on Computer Vision (ICCV)*, pages 7153–7163, 2019. 2
- [17] Haoxiang Guo, Shilin Liu, Hao Pan, Yang Liu, Xin Tong, and Baining Guo. Complexgen: CAD reconstruction by B-rep chain complex generation. *ACM Trans. Graph.*, 41(4):1–18, 2022. 2
- [18] Pradeep Kumar Jayaraman, Aditya Sanghi, Joseph G Lambourne, Karl DD Willis, Thomas Davies, Hooman Shayani, and Nigel Morris. Uv-net: Learning from boundary representations. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 11703–11712, 2021. 2
- [19] Kacper Kania, Maciej Zieba, and Tomasz Kajdanowicz. UCSG-NET-Unsupervised discovering of constructive solid geometry tree. In *Advances in Neural Information Processing Systems*, pages 8776–8786, 2020. 2, 4, 6
- [20] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015. 5
- [21] Sebastian Koch, Albert Matveev, Zhongshi Jiang, Francis Williams, Alexey Artemov, Evgeny Burnaev, Marc Alexa, Denis Zorin, and Daniele Panozzo. Abc: A big cad model dataset for geometric deep learning. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 9601–9611, 2019. 2, 5
- [22] Joseph George Lambourne, Karl Willis, Pradeep Kumar Jayaraman, Longfei Zhang, Aditya Sanghi, and Kamal Rahimi Malekshan. Reconstructing editable prismatic cad from rounded voxel models. In *SIGGRAPH Asia 2022 Conference Papers*, pages 1–9, 2022. 3
- [23] Joseph G Lambourne, Karl DD Willis, Pradeep Kumar Jayaraman, Aditya Sanghi, Peter Meltzer, and Hooman Shayani. BRepNet: A topological message passing system for solid models. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 12773–12782, 2021. 2
- [24] Frank C Langbein, A David Marshall, and Ralph R Martin. Choosing consistent constraints for beautification of reverse engineered geometric models. *Computer-Aided Design*, 36(3):261–278, 2004. 2
- [25] Eric-Tuan Lê, Minhuk Sung, Duygu Ceylan, Radomir Mech, Tamy Boubekeur, and Niloy J Mitra. CPFN: Cascaded primitive fitting networks for high-resolution point clouds. In *IEEE International Conference on Computer Vision (ICCV)*, pages 7457–7466, 2021. 2
- [26] Changjian Li, Hao Pan, Adrien Bousseau, and Niloy J Mitra. Sketch2cad: Sequential cad modeling by sketching in context. *ACM Trans. Graph.*, 39(6):1–14, 2020. 2
- [27] Lingxiao Li, Minhuk Sung, Anastasia Dubrovina, Li Yi, and Leonidas J Guibas. Supervised fitting of geometric primitives to 3d point clouds. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 2652–2660, 2019. 2
- [28] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks:

- Learning 3d reconstruction in function space. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 4460–4470, 2019. 1, 2
- [29] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision (ECCV)*, 2020. 2
- [30] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. DeepSDF: Learning continuous signed distance functions for shape representation. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 165–174, 2019. 1, 2
- [31] Despoina Paschalidou, Ali Osman Ulusoy, and Andreas Geiger. Superquadrics revisited: Learning 3d shape parsing beyond cuboids. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 10344–10353, 2019. 2
- [32] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. PointNet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017. 1
- [33] Daxuan Ren, Jianmin Zheng, Jianfei Cai, Jiatong Li, Haiyong Jiang, Zhongang Cai, Junzhe Zhang, Liang Pan, Mingyuan Zhang, Haiyu Zhao, et al. CSG-stump: A learning friendly CSG-like representation for interpretable shape parsing. In *IEEE International Conference on Computer Vision (ICCV)*, pages 12478–12487, 2021. 2, 3, 4, 6
- [34] Daxuan Ren, Jianmin Zheng, Jianfei Cai, Jiatong Li, and Junzhe Zhang. Extrudenet: Unsupervised inverse sketch-and-extrude for shape parsing. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part II*, pages 482–498, 2022. 3, 6
- [35] Aristides AG Requicha and Jarek R Rossignac. Solid modeling and beyond. *IEEE computer graphics and applications*, 12(5):31–44, 1992. 1
- [36] Ruwen Schnabel, Roland Wahl, and Reinhard Klein. Efficient RANSAC for point-cloud shape detection. In *Comput. Graph. Forum*, volume 26, pages 214–226, 2007. 2
- [37] Ari Seff, Yaniv Ovadia, Wenda Zhou, and Ryan P. Adams. SketchGraphs: A large-scale dataset for modeling relational geometry in computer-aided design. In *ICML 2020 Workshop on Object-Oriented Learning*, 2020. 2
- [38] Jami J Shah. Designing with parametric cad: Classification and comparison of construction techniques. In *International Workshop on Geometric Modelling*, pages 53–68. Springer, 1998. 2
- [39] Gopal Sharma, Rishabh Goyal, Difan Liu, Evangelos Kalogerakis, and Subhransu Maji. CSGNet: Neural shape parser for constructive solid geometry. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 5515–5523, 2018. 2
- [40] Gopal Sharma, Difan Liu, Subhransu Maji, Evangelos Kalogerakis, Siddhartha Chaudhuri, and Radomír Měch. ParSeNet: A parametric surface fitting network for 3d point clouds. In *European Conference on Computer Vision (ECCV)*, pages 261–276, 2020. 2
- [41] Li-Yong Shen, Meng-Xing Wang, Hong-Yu Ma, Yi-Fei Feng, and Chun-Ming Yuan. A framework from point clouds to workpieces. *Visual Computing for Industry, Biomedicine, and Art*, 5(21):1–18, 2022. 2
- [42] Vincent Sitzmann, Michael Zollhöfer, and Gordon Wetzstein. Scene representation networks: Continuous 3d-structure-aware neural scene representations. *Advances in Neural Information Processing Systems*, 2019. 2
- [43] C-H Teh and Roland T. Chin. On the detection of dominant points on digital curves. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(8):859–872, 1989. 5
- [44] Shubham Tulsiani, Hao Su, Leonidas J Guibas, Alexei A Efros, and Jitendra Malik. Learning shape abstractions by assembling volumetric primitives. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 2635–2643, 2017. 2
- [45] Shubham Tulsiani, Tinghui Zhou, Alexei A Efros, and Jitendra Malik. Multi-view supervision for single-view reconstruction via differentiable ray consistency. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 2626–2634, 2017. 2
- [46] Mikaela Angelina Uy, Yen-Yu Chang, Minhyuk Sung, Purvi Goel, Joseph G Lambourne, Tolga Birdal, and Leonidas J Guibas. Point2Cyl: Reverse engineering 3D objects from point clouds to extrusion cylinders. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 11850–11860, 2022. 3, 4, 6
- [47] Tamas Varady, Ralph R Martin, and Jordan Cox. Reverse engineering of geometric models—an introduction. *Computer-aided design*, 29(4):255–268, 1997. 1
- [48] Xiaogang Wang, Yuelang Xu, Kai Xu, Andrea Tagliasacchi, Bin Zhou, Ali Mahdavi-Amiri, and Hao Zhang. PIE-Net: Parametric inference of point cloud edges. *Advances in Neural Information Processing Systems*, pages 20167–20178, 2020. 2
- [49] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph CNN for learning on point clouds. *ACM Trans. Graph.*, 38(5):1–12, 2019. 1
- [50] Karl DD Willis, Yewen Pu, Jieliang Luo, Hang Chu, Tao Du, Joseph G Lambourne, Armando Solar-Lezama, and Wojciech Matusik. Fusion 360 gallery: A dataset and environment for programmatic cad construction from human design sequences. *ACM Trans. Graph.*, 40(4):1–24, 2021. 2, 5
- [51] Rundi Wu, Chang Xiao, and Changxi Zheng. Deepcad: A deep generative network for computer-aided design models. In *IEEE International Conference on Computer Vision (ICCV)*, pages 6772–6782, 2021. 2
- [52] Qiangeng Xu, Weiyue Wang, Duygu Ceylan, Radomir Mech, and Ulrich Neumann. DISN: Deep implicit surface network for high-quality single-view 3d reconstruction. *Advances in Neural Information Processing Systems*, 2019. 2
- [53] Xianghao Xu, Wenzhe Peng, Chin-Yi Cheng, Karl DD Willis, and Daniel Ritchie. Inferring cad modeling sequences using zone graphs. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 6062–6070, 2021. 2
- [54] Xiang Xu, Karl DD Willis, Joseph G Lambourne, Chin-Yi Cheng, Pradeep Kumar Jayaraman, and Yasutaka Furukawa.

- Skexgen: Autoregressive generation of cad construction sequences with disentangled codebooks. In *International Conference on Machine Learning*, 2022. 2
- [55] Dong-Ming Yan, Wenping Wang, Yang Liu, and Zhouwang Yang. Variational mesh segmentation via quadric surface fitting. *Computer-Aided Design*, 44(11):1072–1082, 2012. 2
- [56] Siming Yan, Zhenpei Yang, Chongyang Ma, Haibin Huang, Etienne Vouga, and Qixing Huang. Hpnet: Deep primitive segmentation using hybrid representations. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2753–2762, 2021. 2
- [57] Yuezhi Yang and Hao Pan. Discovering design concepts for CAD sketches. In *Advances in Neural Information Processing Systems*, 2022. 2
- [58] Lior Yariv, Yoni Kasten, Dror Moran, Meirav Galun, Matan Atzmon, Basri Ronen, and Yaron Lipman. Multiview neural surface reconstruction by disentangling geometry and appearance. In *Advances in Neural Information Processing Systems*, 2020. 2
- [59] Fenggen Yu, Zhiqin Chen, Manyi Li, Aditya Sanghi, Hooman Shayani, Ali Mahdavi-Amiri, and Hao Zhang. CAPRI-Net: Learning compact cad shapes with adaptive primitive assembly. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 11768–11778, 2022. 2, 3, 4, 5, 6
- [60] Long Zhang, Jianwei Guo, Jun Xiao, Xiaopeng Zhang, and Dong-Ming Yan. Blending surface segmentation and editing for 3D models. *IEEE Trans. on Vis. and Comput. Graph.*, 28(8):2879–2894, 2022. 2