Deep Sketch-Based 3D Modeling: A Survey

Alberto Tono[1,2,*] Jiajun Wu[1] Gordon Wetzstein[1] Iro Armeni[1] Hariharan Subramonyam[1] James Landay[1] Martin Fischer[1]

[1] Stanford University, USA, atono, jiajunw, gordonwz, iarmeni, harihars, landay, fischer@stanford.edu
[2] Computational Design Institute, USA, alberto.tono@cd.institute
[*] Corresponding author: atono@stanford.edu

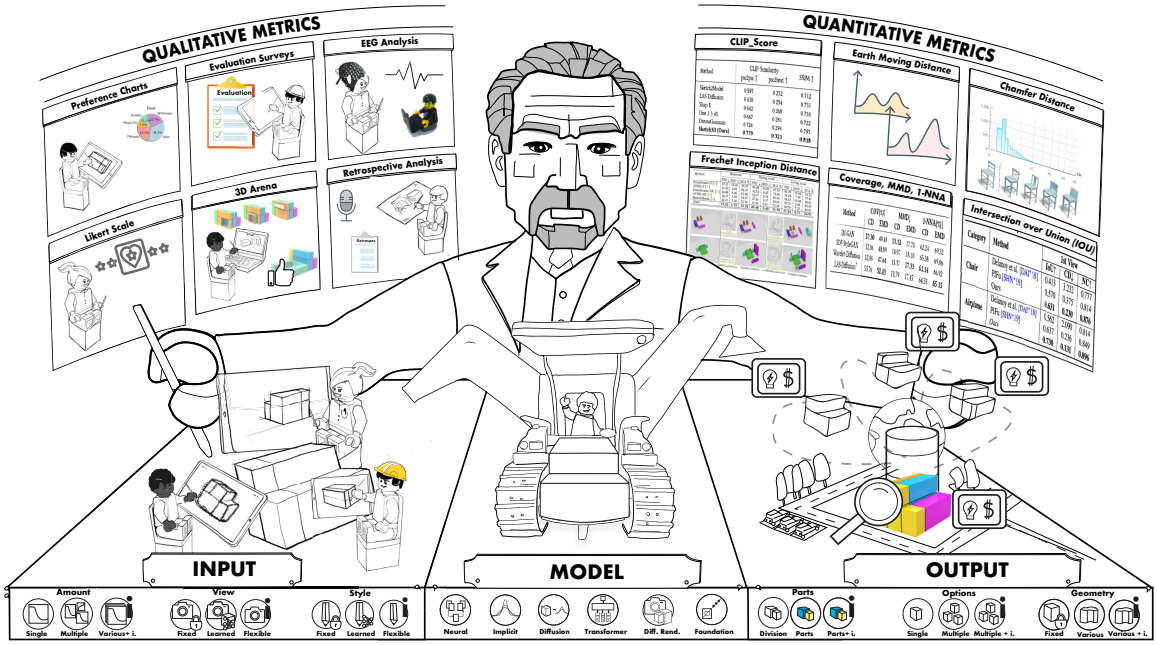


Figure 1: *Illustrates our design space* MORPHEUS *organized around the Input-Model-Output (IMO) framework used to survey deep sketch-based 3D modeling (*DS-3DM*) methods. The Input is categorized into the amount of sketches, their type of viewpoint, and sketch styles with their respective subcategories. The Model is categorized based on the architecture type: neural networks, implicit functions, diffusion models, transformers, differentiable renderers, and foundation models. The Output is categorized by the 3D output compositions, such as part-based semantic, number of options, and geometric topology as well as their corresponding subcategories. The background displays the qualitative and quantitative metrics used to evaluate* DS-3DM*. The figure encapsulates the paper's goal: enabling a user-controlled process driven by human-centric metrics to support informed design.*

Abstract

*In the past decade, advances in artificial intelligence have revolutionized sketch-based 3D modeling, leading to a new paradigm known as Deep Sketch-Based 3D Modeling (*DS-3DM*). DS-3DM offers data-driven methods that address the long-standing challenges of sketch abstraction and ambiguity.* DS-3DM *keeps humans at the center of the creative process by enhancing the flexibility, usability, faithfulness, and adaptability of sketch-based 3D modeling interfaces. This paper contributes a comprehensive survey of the latest* DS-3DM *within a novel design space:* ***MORPHEUS****. Built upon the Input-Model-Output (IMO) framework,* ***MORPHEUS*** *categorizes* ***M****odels outputting* ***O****ptions of 3D* ***R****epresentations and* ***P****arts, derived from* ***H****uman-inputs (varying in quantity and modality), and* ***E****valuated across diverse* ***U****ser-views and* ***S****tyles. Throughout* MORPHEUS *we highlight limitations and identify opportunities for interdisciplinary research in Computer Vision, Computer Graphics, and Human-Computer Interaction, revealing a need for controllability and information-rich outputs. These opportunities align design processes more closely with user' intent, responding to the growing importance of user-centered approaches.*

CCS Concepts

• ***Computing methodologies*** → ***Computer vision representations; Graphics input devices;*** • ***Human-centered computing*** → ***User interface toolkits;***

1. Introduction

In the iconic scene from *The Matrix Reloaded* movie, when Neo (Thomas A. Anderson) meets 'The Architect,' the camera transitions from a panoramic view of the galaxy to a pencil. This cinematic moment encapsulates the essence of sketch modeling: a universal method capable of translating the Architect's intent from the physical world to the digital one. Indeed, 2D sketches are a simple yet effective tool for rapidly communicating complex and abstract concepts [HE17b]. Today, the challenge lies in bridging the gap between the physical and digital medium, especially for 3D modeling, by removing intrinsic sketch ambiguities and ensuring that the user's intent is accurately captured and conveyed to the final 3D digital representation. Addressing this gap will be a significant step toward democratizing the design process, thereby empowering individuals to articulate their concepts through sketches. A critical component of this effort involves evaluating the alignment between the user's intent and the output. This alignment requires the development and iteration of human-centric metrics that assess how well the generated 3D models reflect the user's original vision. Key questions arise: How can we quantify the fidelity of this alignment? What specific metrics should be used to evaluate both the semantic richness and geometric accuracy of the output? How do we ensure that the information embedded in the 3D model, such as annotations, context, or material properties, enhances usability and supports decision-making? Addressing these questions is essential for ensuring an intuitive and effective transition from physical sketches to digital 3D representations while fostering a more inclusive and user-centered approach to 3D modeling. To support this endeavor, we provide insights into these metrics and potential research directions in Section 6.3, and in Table 6.

Sketch-Based Interfaces for Modeling (SBIM) [LM01, LM95a, LM95b, SG00, Sut63, IMT99, ZHH96, NJC*22, OSCSJ09, DL16, BAC*19, FAZ21, LEL*25] predominantly utilize 2D hand-drawn sketches as input, with limited exploration of 3D sketching interfaces [LCX*23, GSL*25, XKK*24]. The input modalities include hand-drawn doodles (abstract drawings made by amateur designers in a few seconds), scaffolding-based sketches [HLW*17], one-, two-, and three-point perspectives [THAF24], axonometric sketches [LPBM22], freehand sketches [WLY*20], scribbles [SJR*23], contour-based sketches [JFD20], and design sketches [ZQG*20]. These modalities leverage the human's innate ability to communicate design ideas [DDS*09] visually, which is deeply rooted in human cognition and culture [OSCSJ09]. However, while sketches are a powerful tool for expression, they capture only partial representations of the users' comprehensive design intent, as illustrated in Figure 2. These inherently ambiguous and incomplete representations impede current SBIM methods from accurately translating users' intent into precise 3D geometric outcomes. As researchers, we might wonder how SBIM can solve this problem, capturing the unsketched knowledge and adequately communicating the user's intent to an information-rich 3D representation [YSR*20]. To address this, modern approaches are increasingly relying on data-driven methods [LOM*18] to infer missing information from datasets of paired shapes and sketches. To ensure that designers' intentions are effectively translated and understood by all, this paper explores how SBIM can capture unsketched knowledge and predict the user's intent.

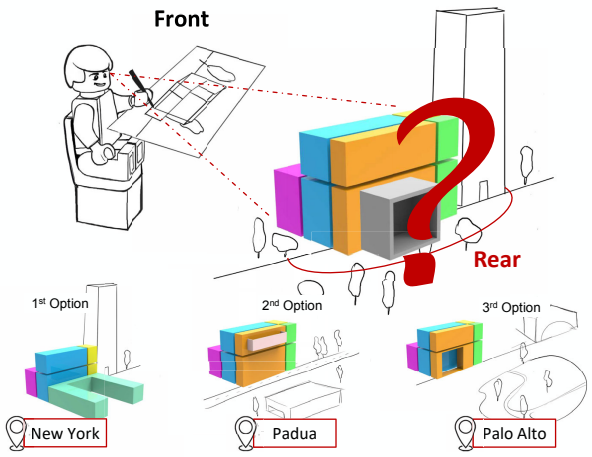


Figure 2: *The initial sketch in the architectural design process contains partial massing, volumetric, and geometric information [THAF24]. This early representation is incomplete because it represents a building only from its front view, conveying only partial information lacking a comprehensive 3D understanding. This single perspective leaves out other buildings' details, particularly the rear [YSR*20]. Moreover, this missing information is influenced by extrinsic and intrinsic factors. Extrinsic factors rely on contextual information such as location, surrounding neighborhood, and local climate. For example, the building will have different shapes if it is designed in New York, Padua, or Palo Alto. Intrinsic factors, such as building typology (office, school, house), design constraints, materials, textures, and appearance reside in the architect's vision of how to translate client requirements into physical spaces.*

Some SBIMs ensure that the designers' intentions are translated in the 3D space by learning the relationship between sketches and 3D shapes, modeled as a conditional distribution between these two modalities in the training dataset. This learned mapping enables the model to infer, predict, and complete the overall form of an object from partial sketch inputs [MON*19, THAF24]. For example, given a frontal sketch of a couch, the model can use prior knowledge to generate the unseen sides of the couch—similar to how humans make accurate and informed predictions about unseen parts [EHA12, HLW*17]. While these methods excel at capturing the overall semantics of a simple black-and-white sketch, they struggle to infer additional details—such as colors, materials, or textures—that are absent from the drawing. Critical to the final design, these visual elements remain beyond the reach of such models when not explicitly provided.

Other SBIMs tackle these representational limitations through sketch-guided text-to-3D generative models [LSC24, LZC*24], enabling users to provide additional details through text or speech, such as colors and materials. However, while these models often produce precise and geometrically accurate representations, they still struggle with language-driven ambiguities and need improvement in effectively conveying intricate design ideas [LFLG24, CYW*24, ZXC*24]. Accurately capturing intent and resolving ambiguities relies on a harmonious interaction between humans and

machines. Even a perfectly calibrated sketch can still harbor multiple ambiguities, making a seamless interaction crucial for bridging the gap between creative intent and precise execution. Since not everyone possesses advanced sketching skills, there is a need for methods that augment the human ability to communicate 3D representations through a level of abstraction [KBS*23] usually captured by simple and quick doodles [BKD*24, HE17b].

The concept of augmenting human capabilities through interfaces predates SketchPad [Sut63] and can be traced back to 1962 with Douglas Engelbart's seminal work, where he articulated the need for a *"clerk"* that enhances human intelligence by supporting informed decision-making and aligning with the user's intent [Eng62a]. In his example, the *"clerk"* needed to support an architect or builder during their design process, hence our analogy and focus throughout the paper on the building environment.

Even if Engelbart perfectly anticipated the advent of the personal computer, he overlooked the potential of novel SBIM for 3D content generation. In fact, the growing demand for 3D user-generated content for gaming and design platforms has underscored the urgent need for advancements in real-time multimodal generation [CLT*23, JMB*22]. To answer this need for novel multimodal generative models, this report explores novel SBIM technologies, providing researchers with future research directions and equipping the industry to commercialize these advancements effectively by aligning design processes more closely with user intent. The information needed for this alignment varies across sectors and design phases—for instance, precise spatial geometry in architecture, dynamic character models in gaming, or ergonomic and functional details in industrial design—underscoring the need for flexible, user-centered approaches tailored to diverse applications. To achieve this flexibility, we examine methods that combine sketch-based and learning-based approaches with a particular focus on sketch-to-3D object generation, called throughout the literature **Deep Sketch-Based 3D Modeling (DS-3DM)** [ZGZS20, ZQG*20]. To facilitate a conversation around this topic, we introduce MORPHEUS, a design space [CMR90] for DS-3DM methods [IIC*13]. Built upon the Input-Model-Output (IMO) framework, **MORPHEUS** categorizes **M**odels outputting **O**ptions of 3D **R**epresentations and **P**arts, derived from **H**uman-inputs (varying in quantity and modality), and **E**valuated across diverse **U**ser-views and **S**tyles. MORPHEUS is a simple and structured design space that identifies the key components of these methods. While previous sketch-based surveys [OSCSJ09, CA09] provided early insights into 3D sketch interfaces, they lacked sections on single-image 3D reconstruction and generation models, since large sketch and 3D datasets were not yet available at that time. This resulted in coverage limited to computationally intensive geometric approaches integrated within traditional WIMP (Window, Icon, Menu, Pointer)-based interfaces. More recent surveys [LB25] have focused primarily on 2D sketch processing, briefly covering mid-air 3D strokes for AR/VR applications. In contrast, we introduce these components and group these methods into categories and related subcategories, as detailed in Section 3. Overall, Section 4 explores the diversity of sketch input modalities, such as doodles, freehand drawings, and scribbles, showing the ongoing efforts within the community to expand the flexibility of the input sketches that can be drawn by different users with different styles and from different viewpoints. Section 5 categorizes the DS-3DM methods following a simple division and analysis based on the type of AI model architecture. Section 6 presents a set of qualitative and quantitative metrics, categorizing outputs based on their adaptability and alignment with user intent. Specifically, these metrics evaluate the method's capability to produce geometrically accurate and topologically appropriate shapes that meet user requirements, as well as its ability to generate multiple output options accompanied by relevant information to facilitate user selection. Finally, Section 8 summarizes unresolved challenges presented throughout the report and highlights future research directions. Our design space and related insights, highlighting the intersection of human-computer interaction (HCI) [LJJ*24] and computer graphics, play a critical role in developing novel DS-3DM to allow users to fully convey their design intents.

2. Scope

Since research on DS-3DM is interdisciplinary, this report serves a diverse audience with expertise in 3D content creation. Sketch-modeling touches many industries such as automotive [GRYF21, YAB*22, NSF*22], architecture [THAF24, TSNA*21, TTZ20, SAMS*21, NKR*22, YJK*23, DLP*23, XKK*24], industrial design [KGC*17, PMKB23, NBS*24, SKR*24, CYH*25, CCL*24], interior design [CDZ*24], fashion [ZZX24, YZM*23, SZZZ23, FRH*21], comics [CBK*24, GYZ23], media and entertainment [XNW*24, SSC*24, WWZ*24], biology [OCM*23], and many others [MPA24, WZW*24, SLX*25]. Specifically, designers use sketches to design 3D representations such as garments [CWC*22, LMG23, GTC*25, GZW*25], digital avatars [HGY17, WCHW24, USB22, LCD*23, WWH*25a], animals [LPL*18, LZZ*21, LCD*23], heads [DHF*22], hair [SZF*21, ZLX*25], smoke [KHW*22], faces [HGY17, LWL*22, LDZ*24, GLC*23, WZY*24], trees [MCEG23], and many other shapes. Although our readers may come from diverse backgrounds, we anticipate that they possess a foundational understanding of deep learning, computer vision, computer graphics, and human-computer interaction. We invite the readers to review previous 3D content generation surveys [LZK*24, LHH*24, XX23, Ebe24], for image [XX23, LZC*24, SPX*23] or text to 3D generation [CRX*19, CG23, CLT*23, HLHF24, LSC24].

To keep this report clear and concise, we have intentionally limited its scope. This report focuses exclusively on DS-3DM for 3D static objects, as presented in Table 2, which shows datasets used in this line of research [CFG*15, DSS*22]. We deliberately exclude animated shapes (animals, humans, faces, smoke, hair, clothing) and 3D scene generation, as these domains introduce additional complexity requiring physics simulations, temporal dynamics, spatial relationships, and specialized frameworks that warrant separate analysis. These 3D static objects are paired with user sketches drawn on a flat surface: paper or tablet. We do not consider non-learning based methods [Pra04, SG00, IMT99, KH06, NISA07, TZF04, AGB04, SWSJ06, KHR02, XCS*14, ML07, LPL*17, SAG*13, HGSB22, YAB*22] such as inflatable techniques [DSC*20] or skeleton-based [MWLZ22, YCYW20] approaches. We also do not consider 3D immersive reality sketch-modeling [Cha24, LCX*23, CLPK24] or purely retrieval-based

methods [QGS*21, BKK*22, ERB*12, BG23, WKL15, CBS*23, WQWF18, XHH*22].

MORPHEUS offers a simple and unified design space [CMR90] to survey DS-3DM methods. We grouped each method into one of the following: reconstruction, generation, and editing [HPG*22], as outlined in the Supplementary material in Table 7; however, discussions and considerations specific to editing are beyond the scope of this report. The overarching goal of DS-3DM is to democratize 3D content creation through controlled and informed design processes. MORPHEUS achieves this by: 1) identifying DS-3DM' limitations and 2) highlighting future research areas. Our proposed design space provides a simple yet unified framework to survey DS-3DM, facilitating both industry and academic use cases. For industry practitioners, this design space can guide the selection of methods most suited to their applications. Meanwhile, academic researchers can use it to identify limitations in existing approaches and explore potential directions for future research.

A central focus of this design space is the role of information in enriching 3D content creation. Information extends beyond visual attributes like color, texture, or geometry to encompass complex processes such as cost estimation, meshing, or aerodynamic performance [Eng62b]. Given that DS-3DM methods are often highly application-specific, our analysis reveals a growing trend toward incorporating contextual factors, external influences that shape a design's functionality and environment. Examples include urban surroundings for buildings, road conditions for vehicles, or placement settings for furniture.

Embedding such insights into DS-3DM ensures the creation of functional and contextually relevant designs. These external considerations are critical for tailoring the final design to its intended use case, carefully considering the surrounding environment [SDM*24, ZCM*24, CZSY25, CSB*22].

3. A Design Space for Deep Sketch-Based 3D Modeling

MORPHEUS assists readers in navigating the rapidly evolving field of DS-3DM. MORPHEUS is designed to serve multiple audiences: researchers seeking to understand current limitations and identify future opportunities, and industry leaders aiming to align their solutions with state-of-the-art methodologies. With MORPHEUS, we present a comprehensive analysis of DS-3DM, structured within the input-model-output framework. MORPHEUS introduces a simple categorization detailed in Sections 4, 5, and 6, and visually summarized in Figures 1 and 3. This categorization emphasizes both human-centric metrics and informed design, providing a structured foundation to guide future developments in sketch-based methods.

- **Input**: The input section (Section 4) emphasizes the diverse modalities and categorizes approaches based on their flexibility to handle various user-provided data. This reflects the research community's growing interest in aligning input processing capabilities with diverse user intents, minimizing constraints on the amount of sketches, their type of viewpoints, and their sketching style.
- **Model**: The model section (Section 5) investigates the trade-offs between different architectures, providing a chronological visualization and a coherent explanation of such evaluation over time. This section maps the transformation from input to output, forming the paper's conceptual core. The architectural design—specifically the encoder, decoder, and loss functions—is critical in translating user intent into precise computational representations, bridging the gap between conceptual design and generated outcomes.
- **Output**: The output section (Section 6) highlights the growing emphasis on diversity, adaptability, and user-centric evaluation of DS-3DM methods. This section explores the methods' ability to provide users with informative 3D part-based representations, with various numbers of options to generate diverse and complex geometries. These outputs are evaluated using both qualitative and quantitative metrics, which are currently evolving to better reflect user needs. A unifying theme is the drive to integrate perceptual and task-based metrics, bridging the gap between subjective user satisfaction and objective performance measurements.

Each section has a table that facilitates a structured discussion and streamlined comparisons. These tables (Tables 1, 3 and 5) highlight trends, gaps, and future opportunities by organizing the design space with detailed input and output subcategories. This structured framework facilitates novel direction identification and encourages methods that better support informed, user-centered design processes.

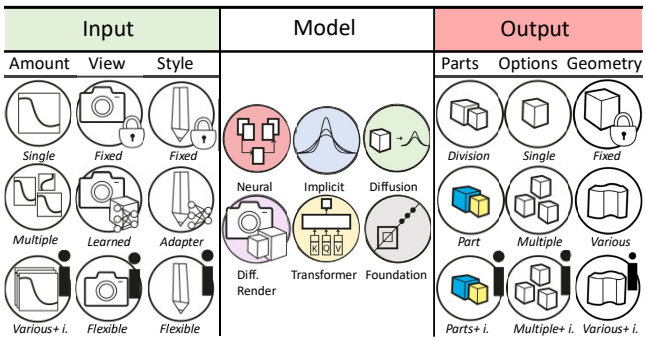


Figure 3: *Illustrates* MORPHEUS *and its overall structure. The input is divided into three main aspects: (1) the quantity of sketches, including multiple sketches, single sketches, and single or multiple sketches with additional information (e.g., text); (2) the viewpoint, specifying whether a fixed viewpoint, learned camera parameters, or a view-independent approach is used; and (3) the style, categorized as fixed sketching styles, style adapters, or flexible styles. The model section highlights six key techniques—neural models, implicit representations, diffusion models, differentiable renderers, transformers, and foundation models—with methods often combining multiple techniques. The output is divided into: (1) parts and semantics, encompassing individual element divisions, part-based segmentation, and parts with related semantics (e.g., material properties); (2) geometric genus, which ranges from fixed representations to flexible genus and enriched geometry; and (3) options, indicating whether the method produces a single output, multiple outputs, or multiple outputs with additional information.*

4. Input: I_{sketch}

In this section, we present the findings from a comprehensive analysis of DS-3DM methods. DS-3DM methods aim to improve the user experience in generating a desired 3D output ($\hat{X}_{\text{shape}}$) by minimizing the need for complex 3D spatial input. Instead, they tend to rely on just a single sketch (I_{sketch}) with additional text (see Section 4.1) and more flexibility to both the sketch's viewpoint (ΘV) and the user's style (ΨS), as described in Figure 4 and respective Sections 4.2 and 4.3. This trend became visible after analyzing Table 1, which organizes these key aspects of input handling, facilitating discussions throughout the literature. Therefore, our discussion is structured to follow the table from left to right, with rows arranged in chronological order.

4.1. Amount

The column "Amount" (I_{sketch}) in Table 1 categorizes the number of input sketches I_{sketch} into three main groups: multiple sketches, single sketches, and single sketches with additional information. The categorization is based on whether the model:

- works with a single sketch or strokes,
- requires multiple sketches, and/or
- works with a single or multiple sketches and information *i*.

For information *i*, DS-3DM is increasingly focusing on methods that utilize a single sketch supplemented by non-geometrical related information, such as appearance [WWF*23] and contextual information [THAF24, CDZ*24], often provided through text description [CYW*24, ZXC*24, LFLG24]. However, there are exceptions, particularly for industrial designers who need multiple drawings [DZX24, LPBM22, LPBM20] to effectively reduce geometric-related ambiguity [FQS*24]. Our analysis primarily examines DS-3DM from a single-user perspective. However, we acknowledge that more complex considerations are necessary for collaborative settings, where multiple stakeholders contribute to the design process. For instance, GroundUp [USGB24] facilitates collaboration between urban planners and architects by incorporating additional top-view sketches provided by planners. This approach helps mitigate sketch ambiguity and enhances the accuracy of city-scale 3D generation, demonstrating the potential benefits of integrating multi-user inputs in DS-3DM systems as shown in Figure 4. Similar considerations are taken when considering the methods' ability to operates with different type of viewpoint (Section 4.2) or style (Section 4.3).

4.2. View

The column "View" (ΘV) in Table 1 examines if the type of viewpoint uses:

- fixed viewpoints,
- learned camera parameters, and/or
- flexible to the viewpoint, by leveraging additional information.

Early DS-3DM addressed the challenge of view ambiguity in

Paper	Amount			View			Style		
Nishida et al. [NGDA*16]									
Delanoy et al. [DBA*17]									
ShapeMVD [LGK*17]									
Contour3D [JFD20]									
DeepSketch [ZQG*20]									
Sketch2CAD [LPBM20]									
SketchDiff [XWJ*20]									
FreeHandRec [WLY*20]									
Sketch2Model [ZGG21]									
Sketch2Mesh [GRYF21]									
Free2CAD [LPBM22]									
SS2Mesh [BBD22]									
GeoCode [PLH*22]									
SketchSampler [GYS*22]									
LAS-Diffusion [ZPW*23]									
Sketch-A-Shape [SJR*23]									
SKED [MPS*23]									
CLIPXPlore [HHL*23]									
D3DSketch+ [CFZ*23]									
Control3D [CPL*23]									
Re3DSketch [CDZ*24]									
Sketch2Point [KWQ23]									
GA-Sketching [ZLY*23]									
S2PointCol [WWF*23]									
Sketch2Vox [Wan24]									
SketchDream [LFLG24]									
SENS [BHSH*24]									
DY3D [BKD*24]									
Vitruvio [THAF24]									
MVControl [LCZL25]									
SHLine [FQS*24]									
M3DSketch [ZHD*24]									
Sketch2NeRF [CYW*24]									
DualShape [DZX24]									
Sketch3D [ZXC*24]									

Table 1: *Categorization of* DS-3DM *methods' Input to ensure its flexibility. As a disclaimer, we do recognize that this categorization could not properly capture the intent of the paper;* : *multiple sketches.* : *single sketches.* : *single or multiple sketch and information,* : *explicitly using camera parameters,* : *differentiable render (learnable camera parameters),* : *view-independent, and* : *the method has specific assumptions about the style (canny, suggestive contours, or others).* : *style specific, it has a specific method that accounts for style.* : *style independent. The "i"s on the icons signify the type of information described in Section 3. Specifically, when referring to information provided in the input sketch, it indicates that the method can accommodate additional prompt-based inputs, such as text or speech modalities, alongside the sketch itself. The red arrows show the trends that we discussed in our paper.*

sketches by relying on strict view assumptions rather than developing methods capable of accommodating flexible viewpoints. These assumptions include fixed viewpoints, such as axonometric views [LPBM22], frontal views [LGK*17], or other predefined camera positions, to simplify the mapping between the 2D sketch and the resulting 3D geometry. For example, Free2CAD [LPBM22] assumed an axonometric perspective, sidestepping the inherent ambiguity of user-provided sketches (see Figure 8).

In more recent work, researchers have shifted toward actively tackling sketch ambiguity by learning view-related parameters and developing view-aware methods [ZGG21]. View-aware approaches [ZGG21, ZLY*23, ZPW*23] incorporate explicit camera parameters or geometric cues to infer 3D shapes more accurately, acquiring knowledge of the sketch's viewpoint. These approaches leverage camera information, such as position, field of view, and perspective, for the 3D generation. The 3D clues derived by specific camera information are added to the sketch during the generation process. For example, depth-guided warping [Feh04, Som20] moves the sketch [LFLG24] to a depth-based space (2.5D) where it can be quickly wrapped to generate additional views and geometric information removing sketch's ambiguity and improving faithfulness. SketchSampler [GWY*24] instead goes directly to 3D; its sketch translator module extracts spatial information and generates a 3D point cloud that conforms to the shape depicted in the sketch. LAS-diffusion [ZPW*23] uses a view-aware local attention mechanism to match path images to 3D volume features as displayed in Figure 7. Furthermore, recently few viewpoint independent methods [CBS*23, BKD*24] emerged. They handle sketches independently of specific viewpoints, using features learned from foundation models and implicit representations to resolve ambiguities. Both approaches remain sensitive to the sketch's position on the canvas and require a proper alignment for optimal translation. For example, Zhong et al. [ZQG*20] addressed this centering issue by employing spatial transformer networks to automatically center and align sketches. The spatial transformer network ensures consistent and accurate model predictions irrespective of the sketch's initial placement. This evolution reflects a transition from rigid assumptions to more flexible and robust techniques for interpreting sketches. In fact, viewpoint independent methods, instead, are methods in which the network does not explicitly consider the camera parameters and positions, presenting more flexible methods [BKD*24] that leverage textual information [HZZ*24].

4.3. Style

The column "Style" (ΨS) in Table 1 focuses on the sketching style, considering the styles used to train the AI models and those that yield the best performance during inference. The subcategories are divided based on when methods:

- use a fixed sketching style during training,
- account for style, with adapters, and/or
- are flexible to the style, by leveraging additional information.

As observed from Table 1, handling sketch style remains a challenging aspect of DS-3DM. Achieving domain-free adaptation to sketch styles, meaning the model must be invariant to distribution shifts caused by variations in sketching styles, as discussed in Section 4 (denoted as ΨS) requires large and diverse datasets of real sketches. However, collecting such datasets is currently prohibitive [CGL*23]. Among various sketching styles, doodles offer an effective medium for conveying abstract semantic information through visual representation [FHWG20, VACOS23, AFCO*25] . Doodles represent an abstract style of sketches, typically created by humans in under 20 seconds (as exemplified in the QuickDraw dataset [HE17b]). To replicate this level of semantic abstraction digitally, CLIPasso [VPB*22] pioneered a method for synthetically generating doodle-like sketches using only a few strokes. While CLIPasso has proven effective for enhancing abstraction modeling in synthetic edgemaps [BKD*24], whether it serves as an adequate proxy for authentic human drawing behavior remains an open research question. While the term "sketch" broadly encompasses both synthetic and natural representations, real sketches are typically hand-drawn on digital or physical surfaces (e.g., paper). In contrast, synthetic sketches are algorithmically generated and can take on various forms, as illustrated in Figure 5. A further distinction is related to the sketch input format to the neural network which influences the encoder design choice. There are mainly two formats: **bitmap** $\boldsymbol{I}_s$ and **vector** inputs $\boldsymbol{I}_{\vec{s}}$. Each category includes both real and synthetic sketch representations. A real sketch refers to one created by a human [ZQG*20], while a synthetic sketch is generated by a machine using techniques designed to mimic a sketch-like form [Can86, VPB*22]. These sketches, paired with 3D shapes, are used to train DS-3DM models (see Dataset Table 2), by learning to associate a user's sketch with specific 3D representations. Since sketches can be drawn by different users and from various perspectives, DS-3DM must be flexible to accommodate this diversity.

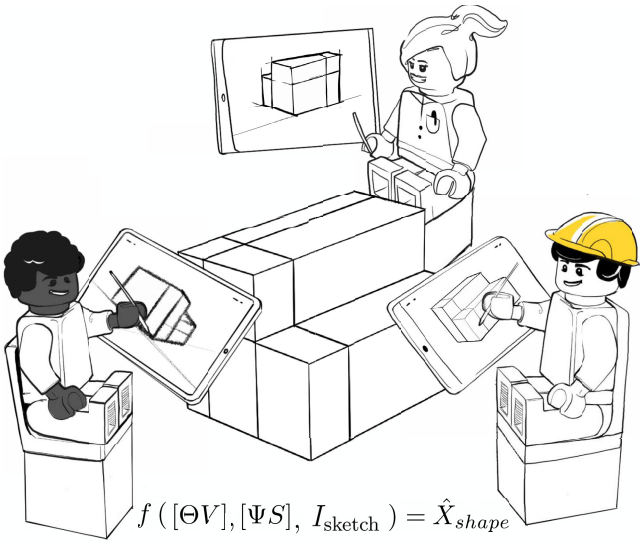


Figure 4: DS-3DM *are designed to accommodate a diverse array of sketching styles [XSL*22] and viewpoints [ZGZS22], including bird's-eye, street-level, front, top, side, axonometric, and perspective views [CGL*23]. These methods are robust to variations in sketch characteristics, whether the lines are wavy or straight, single or multiple, and whether the sketches are shaded or unshaded.*

During *inference*, input sketches can be drawn by both *novice* and *professional* designers [ZQG*20]. Independent of their style, these sketches can be captured in bitmap format if an image is taken from the physical drawing, or in vector format [HGB19] if the sketch is performed directly on a digital interface. However, when differentiating between novice and professional sketches, novice sketches tend to be less structured and more abstract, while expert designers often use *scaffolds* [GSH*19] to guide their drawings, resulting in more precise and detailed designs [HLW*17]. This variation requires DS-3DM to be flexible enough to handle both simple and complex inputs.

During *training*, collecting large datasets of paired 3D shapes X_{3D} and corresponding hand-drawn sketches $I_{sketches}$ is prohibitive. To address this limitation, the research community has developed synthetic proxies, primarily through non-photorealistic rendering (NPR), applicable as a 2D filter or 3D renderer.
Filter-based NPR methods translate 2D images into sketch-like formats, common techniques include Line Rendering [ST90], Apparent Ridges [JDA07], Canny Edge [Can86], Salient Outline [WN19], Hollistically Nested-Edge - HED [XT15], Sobel Edge [SF68], Difference of Gaussians (XDoG) [WOG12], Sketch Simplification via GANs [SSGII16], PaintsUNDO, and others [LPH*24, LHG*23]. *Synthetic Sketch 3D renderer*: these NPR techniques rely on 3D representation to produce more geometrically plausible synthetic sketches, such as Suggestive Contours [DFRS03a, DFRS03b] or Sketch Simplification [HGY17]. For *vector formats*, methods like CLIPasso [VPB*22] represent sketches as sequences of user input strokes [HE17b, LPBM20], which can be grouped into primitives, as demonstrated in Free2CAD [LPBM22] showed in Fig. 11. CLIPasso has been used [BKD*24] to pair these more abstract synthetic sketches and 3D shapes [CBS*23], mitigating the domain shift between synthetic sketches and real sketches.

After analyzing various datasets based on their respective sketch styles (refer to Table 2), we observed that most synthetic datasets predominantly rely on widely-used sketch styles, such as Canny Edges or Suggestive Contours. This raises an important question: "Do these synthetic sketch styles accurately represent the unique styles used in X?" Here, X could refer to a specific design discipline (e.g., architecture, industrial design) or a particular 3D object category (e.g., chairs, buildings, or trees).

While some researchers have attempted to address this question [XSL*22], the growing demand for more refined synthetic representations [KCH*20, LPH*24] underscores the importance of developing domain-specific sketching styles. These styles must better reflect the unique characteristics and requirements of particular fields, emphasizing the need for tailored datasets that capture the nuances of sketch-based design within specific contexts.

Datasets: For a broader survey on sketch datasets, we direct the reader to the work of [XHY*23]. Table 2 summarizes key datasets and their characteristics, focusing on 3D objects used in DS-3DM. Many of these methods require generating custom datasets to address the challenge of domain shift between images and sketches [BKD*24]. This shift occurs because models trained on photographs often perform poorly with sketch inputs without proper fine-tuning; a synthetic-trained mapper does not generalize to human drawings that are vague and ambiguous, lacking geometric and perspective clues [LPH*24, WLY*20]. For instance, Sketch2Model [ZGG21] collected 1,300 sketches of ShapeNet objects from novices and professionals. Vitruvio [THAF24] not only generated 1,000 3D building [CWL*21, YJK*23, SNL*21] shapes as occupancy functions [MON*19] (Section 5.2) but rendered them producing a total of 24,000 synthetic sketches using suggestive contours rendered in Blender. Sketch2CAD [LPBM20] generated 50,000 3D shapes, with different permutations of CAD modeling operations [JMM*24, WPL*21, SLX*25, MNAA25], and corresponding sketches to address the domain shift issue and the lack of a CAD dataset. GeoCode [PLH*22] used Blender Geometry Nodes to create models for chairs, vases, and tables as the dataset, with 59, 39, and 36 human-interpretable parameters as input, respectively, and paired them with NPR style. Sketch3D [ZXC*24, CYW*24, LFLG24, OCK*24] used larger datasets [DSS*22], and re-rendering all these 3D shapes in Suggestive Contours style become computational expensive, therefore other 2D-filter based methods have been used. While previous methods primarily relied on synthetic sketch styles, Deepsketch [ZQG*20] was the first to incorporate 1,500 professional hand-drawn sketches of ShapeNet objects [CFG*15], highlighting the importance of accounting for these nuances in sketch-based modeling [WQF*21]. This shift underscores the need to differentiate between freehand sketches and those generated through synthetic means to better capture the intricacies of human input.

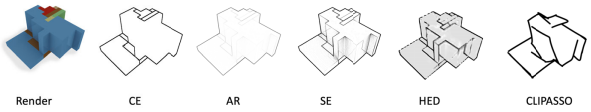


Figure 5: *BuildingGAN [CCL*21] with different sketching styles. CLIPASSO [VPB*22], Apparent Ridges [JDA07] (AE), Canny Edge [Can86] (CE), Hollistically Nested-Edge [XT15] (HED), Sobel Edge [SF68] (SE). These have been generated using OpenCV, and filters have been applied to the initial render.*

4.4. Input Summary

Table 1 reveals that multimodal inputs, particularly combining sketches with text, effectively resolve style and viewpoint ambiguities. Text prompts provide crucial context like *"a chair"* for semantic clarity, *"drawn from a frontal view"* for viewpoint specification, or *"convert this doodle done by a 5-year-old into a chair"* for style information, enhancing neural sketch methods' expressiveness and accuracy. Recent research trends toward advanced techniques including deformation-aware, 3D-aware, and physics-informed losses [UKS*21, ZQG*20], plus equivariant methods from geometric deep learning [STD*21, DLD*21], promising improved robustness for sketch-based 3D modeling.

5. Generative AI Model

In this section, we provide a chronological overview of DS-3DM developments driven by key innovations in neural network architec-

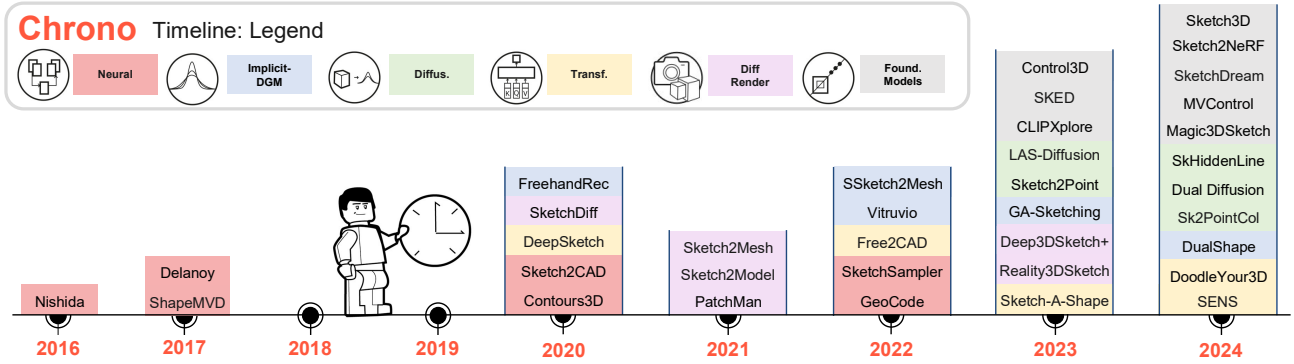


Figure 6: ***Sketch-Modeling Deep Learning Based Methods***. *This timeline illustrates the evolution of* DS-3DM *methods, highlighting key innovations and breakthroughs in the field. Methods are organized chronologically and color-coded by year to facilitate cross-referencing with Table 3, enabling identification of specific research patterns and emerging directions.*

Dataset	Category	Shapes	Sketches	Style	Views
Nishida [NGDA*16]	Param. Prim.	4	47,000	SC	1
ShapeCOSEG [DBA*17]	Chair, Vase	700	5,600	SC	8
OpenSketch [GSH*19]	Product Design	12	400	H	3
ShapeNet	13 Categories	43,783	39,423	Canny/H	48
Manhattan 1k [THAF24]	Buildings	1,000	24,000	SS	24
ShapeNet-S3D [ZXC*24]	ShapeNet+Text	11,000	220,000	Canny	20
Objaverse [DSS*22]	3D Objects	400,000	-	Canny	30
OmObject3D [WZF*23]	20 Categories	-	-	HED	–
Sketch3D Hand [Wan24]	9 Categories	-	12,877	-	-
SpeedTracer [WQF*21]	Various	63+	1,498	H	Various
DifferSketching [XSL*22]	9 Categories	136	3,620	H	2-3

Table 2: *Overview of datasets containing paired sketches and 3D shapes. While ShapeNet [CFG*15] serves as a foundation for many methods, each approach [WLY*20, ZGG21, ZQG*20, ZGZS20, GRYF21, BKD*24] employs a specific subset as detailed in the Supplementary materials. In this table, "H" denotes human-created sketches, "SC" refers to suggestive contours style, "SS" indicates unspecified synthetic sketches, and "HED" represents Holistically Nested-Edge style as shown in Figure 5. "Param. Prim." abbreviates Parametric Primitive. Nishida [NGDA*16] employs procedural techniques to generate multiple variations from four base 3D objects using underlying grammar rules.*

ture (Figure 6). Since these methods often combine multiple innovations simultaneously, such as transformers for structural understanding, diffusion models for geometry generation, and differentiable rendering for 2D-3D alignment, we categorize them based on their primary underlying innovation to simplify our analysis. These fundamental innovations integrate sketch-modeling with these six techniques: neural models (see Section 5.1), deep generative models and novel implicit representation (see Section 5.2), diffusion models (see Section 5.3), differentiable renderers (see Section 5.5), transformers (See Section 5.4), and optimization and pre-trained-based (CLIP) models (see Section 5.6). Since some of the DS-3DM adopt several of these innovations, we provide a comprehensive summary of these methods and overall table (see Table **??**) in the Supplementary 11.1.

5.1. Neural Models:

Neural models in Table 3 establish direct correspondences between 2D image coordinates and 3D representations or via parametric shape modeling or via direct coordinate mapping. In parametric approaches, neural models predict specific parameters that define a 3D shape program (see previous neurosymbolic STAR [RGJ*23]), often relying on representations that consist of procedural operations, like shape grammars or parametric models [NBA18,HKYM17,SBS21]. **Nishida** et al. [NGDA*16] used a cascade of CNNs to classify partial sketches into grammar snippets and estimate their parameters. This approach allows for generating a wide variety of buildings by combining different grammar rules and adjusting building parameters. **Free2CAD** [LPBM22] predicts stroke groupings that correspond to CAD operations, followed by parameter optimization for each group to match the strokes. The parameters are optimized to reproduce the strokes provided as input. Instead, **GeoCode** [PLH*22] predicts parameters for constructive solid geometry (CSG) [SGL*18, YBP*24] operations and extrusion heights, aiming for interpretable shape programs for 3D reconstruction. **Sketch2CAD** [LPBM20], similar to Free2CAD, utilizes regression to predict parameters for CAD operations. It focuses on sequential CAD modeling, parsing user sketches into a series of commands and their associated parameters, limiting their overall shape (see Section 6.3 for more information). However other methods succeed in modeling complex topologies by providing a deformable parametric template composed of Coons NURBS patches to the decoder [SBS21].

In coordinate mapping approaches, models learn direct transformations from image pixel locations to their corresponding 3D spatial positions. **Delanoy et al.** [DBA*17, DCLB19] uses an online method; an updater-CNN iteratively maps the input sketch to a voxel-based representation by utilizing fixed camera viewpoints and perspectives, thus determining which voxels are occu-

Paper	Models
Nishida et al. [NGDA*16]	
Delanoy et al. [DBA*17]	
ShapeMVD [LGK*17]	
Contour3D [JFD20]	
DeepSketch [ZQG*20]	
Sketch2CAD [LPBM20]	
SketchDiff [XWJ*20]	
FreeHandRec [WLY*20]	
Sketch2Model [ZGG21]	
Sketch2Mesh [GRYF21]	
Free2CAD [LPBM22]	
SS2Mesh [BBD22]	
GeoCode [PLH*22]	
SketchSampler [GYS*22]	
LAS-Diffusion [ZPW*23]	
Sketch-A-Shape [SJR*23]	
SKED [MPS*23]	
CLIPXPlore [HHL*23]	
D3DSketch+ [CFZ*23]	
Control3D [CPL*23]	
Re3DSketch [CDZ*24]	
Sketch2Point [KWQ23]	
Sketch2Vox [Wan24]	
GA-Sketching [ZLY*23]	
S2PointCol [WWF*23]	
SketchDream [LFLG24]	
SENS [BHSH*24]	
DY3D [BKD*24]	
Vitruvio [THAF24]	
MVControl [LCZL25]	
SHLine [FQS*24]	
M3DSketch [ZHD*24]	
Sketch2NeRF [CYW*24]	
DualShape [DZX24]	
Sketch3D [ZXC*24]	

Table 3: *Categorization of* DS-3DM*'s methods alongside their corresponding model architectures, emphasizing the primary contributions. The colored cell in each row represents what we identify as the main contribution, following the color scheme in the Figures 3 and 6. The light gray cells indicate the other technical aspects present in each paper.*

pied or empty. Other methods use a two-step approach and multiple sketches: **ShapeMVD** [LGK*17] first predicts depth and normal maps from input sketches. Then, it fuses depth and averages it into a point cloud. Similarly, **DeepSketch** [ZGZS20, ZQG*20] generates normal maps, 2.5D depth maps [WWX*17], and foreground masks from sketches, and then fuses to the 3D space via mapping. While these approaches rely on 2.5D information, they perform poorly with more complex geometries.

5.2. Deep Generative Models and/or Implicit Rep.:

This section focuses on methods based on deep generative models and/or implicit representations presented in Table 3. For a more in-depth understanding of these topics, we invite the reader to review [BTLLW22, DLCS*23]. A 3D implicit representation is defined by a volumetric function $f : \mathbb{R}^3 \rightarrow \mathbb{R}$. For a point cloud, this function takes as input a point in 3D such as $\mathbf{p} = (x, y, z) \in \mathbb{R}^3$ and passes it through a neural network to produce a scalar output in $\mathbb{R}$ written as $\mathbf{p} \mapsto f(\mathbf{p})$, where $f(\mathbf{p})$ encodes information about the 3D shape at point $\mathbf{p}$. In DS-3DM methods, this function can be represented by a Signed Distance Function (SDF) [GRYF21], Occupancy Function [THAF24], or more broadly, Neural Implicit Representation [GRYF21, BBD22, DZX24] composed of a neural network that is typically a multi-layer perceptron (MLP) and used as a continuous function (Deep Implicit Fields). **Sketch2Mesh** [GRYF21] uses [RLR*20] a MeshSDF encoder-decoder architecture to represent and refine a SDF to match the target external contour by using a differentiable renderer (Section 5.5). These methods tend to produce a compact implicit representation that benefits DS-3DM speed [GRYF21, THAF24](see the online column in the Supplementary Table 7). Sketch2Mesh combines view- and contour-aware methods with implicit representations, increasing the flexibility to viewpoint and sketching style. However, this flexibility comes at the cost of sensitivity, thus requiring high precision to accurately capture fine geometric details. Methods like **Sketch2Model** [ZGG21] and **ShapeMVD** [LGK*17] also use implicit representations, but they do not employ deep generative models. Methods like **Vitruvio** [THAF24] and **DeepSketch** [ZQG*20, ZGZS20] leverage deep generative models as Variational AutoEncoder (VAE) [MON*19] and Generative Adversarial Network (GAN) [WWX*17] respectively. If Vitruvio directly maps a single sketch to 3D shapes, DeepSketch (3DSkVP) uses the GAN to translate multiple sketches to a 2.5D domain with normal depth and mask maps fused to generate the 3D mesh via Iterative Closest Point (ICP) and Poisson surface reconstruction [KBH06]. **FreeHandRec** [WLY*20] uses a pix2pix-based [IZZE17] sketch standardization module to reduce the style variability. Unlike Sketch2Mesh, it does not use differentiable rendering techniques to map sketches to 3D space; rather, it uses a view-aware 3D reconstruction network. FreeHandRec encodes the 3D shape with global latent space, which fails to capture fine-grained details of local parts. To overcome this limitation **LAS-Diffusion** [ZPW*23, LJJ*24] controls these local parts with local patch features provided by a view-aware locally attentional SDF (see Figure 7 and Section 5.3).

LAS-Diffusion does not excel in hand-drawn human sketches that have highly distorted lines, cluttered linework, or inconsistent perspectives. To mitigate this issue, **Doodle Your 3D** [BKD*24] leverages a latent diffusion model with a part-disentangled decoder (Figure 9). It establishes correspondence among CLIPasso [VPB*22] (see Section 4) semantic edge maps, and projected 3D part regions. This part-aware Neural Implicit Shape modeling is initiated by SPAGHETTI [HPG*22]. SPAGHETTI's decoder reconstructs a part-disentangled latent space via inversion. The alignment between the input sketch and output parts is trained with a diffusion model, enabling semantic edits that are crucial to capturing fine-grained level details.

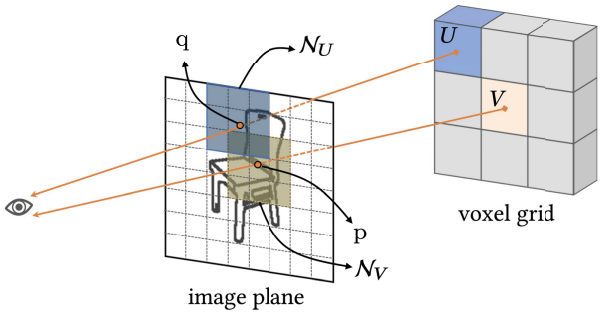


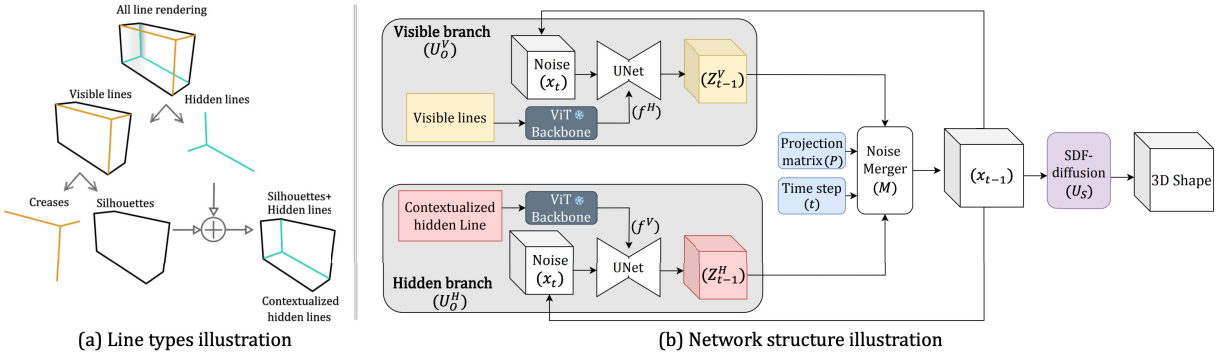
Figure 7: *Demonstration of LAS-Diffusion [ZPW*23] approach for local view-based attention. The center of voxel V is mapped to the image plane at point p using a pre-determined perspective projection. The image patch features in the vicinity of point p (highlighted in yellow) are used to interact with the voxel features of V within the U-Net architecture through cross-attention. The same process is applied to other voxels, such as U. LAS-Diffusion uses 2D image patch features to guide 3D voxel feature learning methods. A voxel grid is projected to the viewer's image plane. Once this relationship is established, patch image features f_N (ViT) interact with the voxel features f_V via a U-Net and one-layer multi-head cross-attention [VSP*17].*

5.3. Diffusion Models:

In this section, we analyze papers leveraging diffusion models to implement their DS-3DM methods, as displayed in Table 3. For a more in-depth survey about diffusion models, refer to [BWpC*24, PYG*24]. Diffusion Models (DMs) are a class of deep generative models (see Section 5.2) that learn to generate 3D shapes $\hat{x}_{shape}$ (for brevity $\hat{x}_{3D}$) from a data distribution $p(x)$ by first learning a forward diffusion process that gradually adds Gaussian noise ϵ to training data x_{shape} (for brevity x_{3D}) over discrete time steps $t = 1, 2, ..., T$ until it becomes pure noise. The model then learns to reverse this noising process, starting from random noise and iteratively denoising to generate new 3D shapes $\hat{x}_{3D}$ [HJA20]. Training typically uses an MSE loss between the actual Gaussian noise and the predicted noise $\hat{\epsilon}$, formulated as $\mathcal{L}_{MSE} = |\hat{\epsilon}_t - \epsilon_t|_2^2$. This gradual corruption of the input can be realized in the data domain, or in the latent space z to reduce the dimensionality of the problem. Throughout the literature, different approaches have been adopted to consider the sketch I_{sketch} (for brevity I_s) in the corruption process of the clean input x_{3DS} as a form of conditioning [HS22] (refer to [CZSY25, ZCM*24]). This learned noise can be represented as a function of the time t, the conditioning sketch I_s and the corrupted shape x_{3DS} at the time t as $\epsilon_\phi(x_{3DS}; t, I_s)$.

Control3D [CPL*23] uses a 2D latent diffusion model [ZRA23] optimized on Neural Radiance Field (NeRF) after converting the input sketch I_s to a colored image I_c. Control3D used view-dependent prompting; they added camera specifications to the prompt y transforming the noise to $\epsilon_\phi(x_{3DS}; t, y, I_c)$. **Doodle Your 3D** [BKD*24] operates directly on the 3D latent representation z [HJA20]. It uses a 3D part-latent diffusion model, where the implicit 3D shape representation $\hat{x}_{3DS_{implicit}}$ is divided in m parts ($Z \in \mathbb{R}^{m \times d}$) aligned with the sketch-parts $E(I_s) \in \mathbb{R}^{m \times d_s}$ [VPB*22, CDI22] via multi-head attention blocks as cross attention modules [VSP*17, ZPW*23] (see Section 5.4 and 5.2); this allows to capture part-specific 3D details that in Control3D have been overlooked due to the global latent space of the NeRF. [BKD*24] establish this part-alignment for only a single category (category-specific model like [ZGG21]) because each part carries specific semantic meaning, and the latent variable Z is aligned across all shapes within the chair category. Other DS-3DM methods capture finer details while maintaining the ability to generalize across multiple categories. They operate directly on the global latent space ($Z \in \mathbb{R}^d$) of the 3D representation x_{3DS}. **LAS-Diffusion** [ZPW*23] tackles the quality gap capturing fine-grained details by employing a two-stage 3D diffusion process that leverages discrete signed distance function (SDF) representation. The first stage, called occupancy diffusion, generates a coarse discrete occupancy function [THAF24, MON*19] to approximate the shape's shell using a 3D U-Net [DBA*17, DCLB19]. The forward process [KSPH21] is represented as: $\mathbf{x_{3DS_t}} = \sqrt{\alpha_t}\mathbf{x_{3DS_0}} + \sqrt{1-\alpha_t}\boldsymbol{\epsilon}$, where $\mathbf{x_{3DS}}$ is the initial 3D shape used for self-conditioning [CZH23]. The second stage, SDF-diffusion, refines the occupied voxels from the first stage to produce a high-resolution SDF. For controllability, LAS-Diffusion uses 2D image patch features to guide 3D voxel feature learning. A voxel grid is projected to the image plane of the viewer; once this relationship is established, patch image features f_N (ViT) interact with the voxel ones f_V via a U-Net and one-layer multi-head cross-attention in the decoder (see usual formulation $f_V^{\text{new}} = \text{MH-Attention}(Q, K, V, \mathcal{M})$ with respectively $Q = f_V W^Q$, $K = f_N W^K$, $V = f_N W^V$). LAS-Diffusion achieves a higher level of control over fine-grained details of the local geometry but lacks details about the color, material, or textures.

To capture color-related details **Sketch2PointCol** [WWF*23], the encoder for the diffusion process uses a capsule attention mechanism to encode the sparsity of the pixels in the sketch, and then a multimodal text-sketch joint embedding to reduce the text-based ambiguities. The geometry diffusion stage employs a U-Net conditioned on sketch and text features embedding C_g to generate the point cloud geometry; in this case, the noise is produced by the 3D point cloud $\left\|\epsilon - \epsilon_\theta\left(\mathbf{x_{3DS_t}}, C_g, t\right)\right\|^2$ The method also uses convolutional layers, capsule networks, and attention mechanisms. The capsule network identifies and focuses on the informative pixels within the sketch, ignoring the background. The text feature fusion uses BERT [DCLT19] and multi-head attention fusion to produce the joint embedding that guides the diffusion process. However, the users with these methods lack control over the entire shape. The users draw from a frontal view and do not provide information about the rear, as shown in Figure 2. Suppose these approaches leverage data to complete the partial information the sketch provides. In that case, SHLine [FQS*24] aligns with the users' needs, especially in automotive and industrial design industries, where 3D scaffolding is provided, and precision is paramount. **SHLine** [FQS*24] extends LAS-Diffusion to leverage hidden lines commonly found in technical drawings. SHLine distinguishes creases and silhouette lines from the hidden ones. The model has two parallel occupancy-based diffusion networks that converge to a CNN-based noise merger before entering the SDF-diffusion network. One processes visible lines, and the other processes hidden lines



(a) Line types illustration (b) Network structure illustration

Figure 8: *Fukushima et al., [FQS*24] captures also hidden lines. It captures the back information (see Figure 2) leveraging hidden lines common in several practices [YSR*20].*

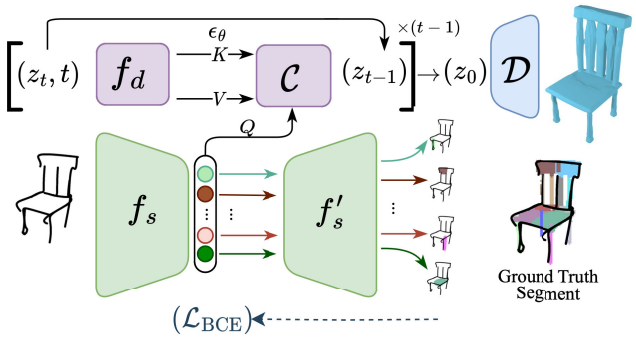
and silhouette lines for context. It tackles geometry and appearance separately. Both branches employ occupancy-diffusion networks, denoted as I_{s_v} for visible lines and I_{s_h} for hidden lines. They predict denoised occupancy grids in a unified noise volume. This noise merger network, M, combines the denoised outputs from both branches, as displayed in Figure 8.

5.4. Transformer-Based Models:

In this section, we analyze papers that introduce transformers and attentions-based methods to DS-3DM methods, as shown in Table 3. Transformers [VSP*17] excel at processing sequential data, making them well-suited for handling the ordered nature of sketch strokes [HE17b]. They capture long-range dependencies within the input sequence, which is crucial for understanding the relationships between strokes in a complex sketch and for accurately grouping them into meaningful units [LPBM20]. For a more detailed survey on transformers, see [LWLQ22]. Transformers exploit self-attention mechanisms. **DeepSketch** [ZQG*20] adopted and re-purposed this self-attention mechanism to ensure that an automatically generated attention map aligns with that of the ground-truth 3D shape. To solve previous problems, such as the one introduced at the end of Section 5.5, of shape misalignment across multiple view optimization, DeepSketch learns a global non-linear geometric transformation between an input sketch and its 3D shape counterpart via a spatial transform network (STN). Through their attention mechanism, transformers can selectively focus on relevant parts of the sketch, mitigating the impact of ambiguities. For instance, when encountering a sketch with unclear viewpoint information, a transformer-based model can learn to prioritize strokes that provide stronger cues about the intended 3D shape, leading to more accurate reconstructions. If they improved the performance of previous methods, they did not fully solve the ambiguity problem related to the style and viewpoint. To reduce this ambiguity, using a fixed viewpoint and a predefined sketching style proved to help.

For example, **Free2CAD** [LPBM22] uses an axonometric fixed viewpoint to better leverage the transformer's ability to handle sequential data, introducing a novel stroke grouping task to learn CAD operations. Free2CAD encodes the strategic knowledge of CAD modeling by using the pen strokes as complete input drawings, grouping them [YZF*21], and fitting geometric parameters. It has been trained on synthetically-generated data, but the model adapts to novice users. It takes as input the complete drawing. It mitigates the exposure bias [MM19,HCX*21] provided by the synthetic sequence of operation with specific transformer-based geometric fitting. To avoid strong assumptions on a fixed axonometric viewpoint, **Sketch-a-Shape** [SJR*23] conditions a 3D Discrete Auto-Encoder (Vector Quantized Variational Auto-encoder SkexGen VQ-VAE [XWL*22]) on the CLIP features of Non Photorealist Render (NPR) from Canny edge [Can86]. Showing the robustness of CLIP (obtained from a frozen large pre-trained vision model) [VPB*22] also in 3D. The model outputs multiple shapes per sketch query thanks to the Masked Transformer, positional encoding, and cross-attention mechanism. The 3D shapes are encoded into a sequence of discrete indices Z, pointing to a shape dictionary, whose distributions are then modeled later with the bi-directional masked transformer. [CZJ*22]. However, Sketch-a-Shape does not leverage parts awareness of the different shapes.

To solve this problem, **SENS** [BHSH*24] uses transformers to process visual embeddings of sketch patches and a set of part queries to predict latent part vectors, which are subsequently used to reconstruct the 3D shape. This transformer-based approach facilitates a part-aware generation process, enabling the model to handle complex shapes and avoid mere retrieval from a database of existing models. SENS generates neural implicit shapes via hand-drawing sketches. It is based on SPAGHETTI [HPG*22] that maps input parts to the latent space. These approaches necessitate a large dataset of sketches and pair 3D data. Another SPAGHETTI-based approach has introduced a proposed solution: **Doodle Your 3D** [BKD*24]. Doodle Your 3D leverages multi-head attention with a latent diffusion model for better part disentanglement in the decoder. This approach establishes a correspondence between semantic edge maps generated via CLIPasso [VPB*22] and projected 3D part regions, reducing dependence on human sketch-3D shape paired datasets [CBS*23] and approximating more natural human sketching patterns. The part-level modeling is initiated from SPAGHETTI [HPG*22]; the decoder inversion reconstructs a part-disentangled latent space. The alignment between the input sketch and output parts is trained with a diffusion model, enabling semantic edits.



(b) Sketch Conditioned Diffusion

Figure 9: *Doodle Your 3D [BKD*24] processes sketches by encoding them into part-disentangled representations using an encoder f_s, where the sketches are segmented into maps representing individual parts. A shared decoder f'_s is applied to these segmented maps. The resulting sketch representations are then passed into the attention module $\mathcal{C}$ as the Query, while the intermediate diffusion outputs from f_d serve as the Key-Value pairs.*

5.5. Differentiable Rendering Models:

Differentiable rendering (DR) seeks to bridge the gap between 2D images and 3D scenes by incorporating the rendering process into neural network training pipelines for end-to-end learning, as grouped in Table 3. This enables a bidirectional interaction between 2D and 3D representations, facilitated by the power of gradient-based optimization techniques, which are also applied later in this report (see Section 5.6). By making the rendering process differentiable, it establishes a direct connection between the pixels in the input image and the 3D parameters that define the scene [RRN*20, JSL*19, OVK*19, JSR*22]. For an exhaustive survey about differentiable rendering, please see [KBM*20]. In short, DR parameters are material ϕ_m, light ϕ_l, geometry ϕ_s, and camera ϕ_c. These parameters are used as inputs to output a colored image. However, this information is not present in a single black-and-white rapid scribble. So one may wonder, how can DR be integrated and support DS-3DM methods? DR needs an overall function, summarized as $\phi = \phi_s, \phi_m, \phi_c, \phi_l$, which generates an image defined as I with color and depth information. DS-3DM methods on the contrary start from the sketch I_{sketch} to generate a 3D shape Φ_s. If this DR works with colored 2.5D (with depth information) images, it fails with sketches due to the lack of color and depth information. To mitigate this limitation, **Sketch2Mesh** [GRYF21] employs a hybrid approach, combining a traditional encoder/decoder architecture for initial 3D mesh generation with a differentiable rendering-based refinement step. This process helps align the projected mesh with the input sketch accurately. Sketch2Mesh uses two differentiable rendering variants: *Render* and *Chamfer*. *Render* uses a differentiable rasterizer (SoftRas) [RRN*20] to generate a corresponding binary mask to back-propagate the information to the 3D representation. *Chamfer* directly minimize the difference between sketch and the projection of the 3D mesh's external contours, moving the optimization in the sketch contour space similar concept applied to Control3D [CPL*23], to improve the sketch-fidelity via its sketch-consistency loss $\mathcal{L}_{\text{sketch}}$ (see Section 5.6). *Chamfer* approach outperforms *Render* for its simplicity and robustness to style changes. This robustness to the style holds only for professional sketches [ZQG*20].

To accommodate inputs from nonprofessional users, Smirnov et al. [SBS21] (referred to as PatchMan in the table) use sketch augmentation strategies in the dataset preparation. The authors used occluding contours and sharp edges (using the Arnold Toon Shader in Autodesk Maya). They augmented these drawings via special techniques (vectorization and augmentation by stochastic split and curve truncation). Even with this augmentation, their initial sketch still requires a good initialization from a correct perspective, which is uncommon in doodles [BKD*24]. To address the aforementioned style limitations and make the method compatible with novice hand drawings or doodles, **Sketch2Model** [ZGG21] disentangle the latent codes for shape and viewpoint. This disentanglement allows the model to be view-aware and explicitly leverage the viewpoint information during generation. Here, the SoftRas DR generates the silhouette from a specific viewpoint. Then a silhouette loss L_s compares the generated silhouette S_1 to the ground truth S_2, summarizing $L_s = \mathcal{L}_{\text{iou}}(S_1, S_2)$. A problem arises when the method tries to leverage silhouettes from multiple views. A single silhouette is insufficient to capture all the information of a 3D object. **Deep3DSketch+** [CFZ*23] addresses this by sampling multiple camera viewpoints and generating silhouettes from each. It introduces a structural-aware adversarial training strategy. This strategy includes a Stroke Enhancement Module (SEM) to capture the structural information and facilitate learning realistic and detailed shape structures thanks to a Shape Discriminator (SD). SD allows more consistency in the 3D representation to ensure it is used as a reference in each optimization step. Similar approaches are used in Section 5.6 with Score Distillation Losses in 3D.

5.6. Pre-Trained Optimization-Based Models (Found.):

Developments of diffusion models [DN21, PYG*24] brought to popularity text-based 3D content generation [PJBM23, SWY*24, SZS*24], for more see text-to-3D surveys [LSC24, LHH*24, SWG24, XX23]. In this section, we survey optimization-based generation methods that use pre-trained large language and vision models [OCK*24] as displayed in the last column of Table 3, represented as Foundation model [BHA*21]. These methods leverage techniques from differentiable render (see Section 5.5) to optimize both geometry and texture via specific loss function, differently from the initial image-conditioned ones based on NeRF optimizations [PJBM23, WLW*23] we focused on sketch-conditioned methods. For example Mikaeili et al. propose **SKED** [MPS*23], introducing the ability to edit the NeRF representation via multiple 3D consistent sketch-based contour masks. SKED for editing uses (Instant-NGP) and $\mathcal{L}_{SDS}$ for the usual SDS 3D consistency loss (as highlighted in Table 4). $\mathcal{L}_{pres}$ for the preservation loss, this loss uses a distance-based weighting mechanism to preserve the base 3D object's original content selectively. This approach ensures that the edits are localized to the regions specified by the user's sketches while maintaining the fidelity of the original object elsewhere. $\mathcal{L}_{sil}$ ensures that the density added during the editing pro-

cess occupies the regions specified by the user's sketches. Instead of directly comparing the sketch with the edited 3D object, which is difficult due to their different representations, SKED cleverly utilizes object mask renderings. This loss penalizes the model if the rendered object mask does not closely match the provided sketch masks. Minimizing this loss encourages the generated 3D edits to fill the areas outlined in the sketches accurately. $\mathcal{L}_{sp}$ for the sparsity to minimize the entropy with the sketch masks for each view. This loss is focused on preserving the shape of the added content to ensure doesn't alter the overall initial shape. If SKED focuses on partial edits later models, such as **Control3D** [CPL*23], generate the entire 3D objects from initial sketches. Chen et al. with Control3D introduce these sketch-conditioned capabilities with the Score Jacobian Chaining (SJC) [WDL*23] and enforce the sketch-fidelity via a novel sketch-consistency loss $\mathcal{L}_{\text{sketch}}$. This loss optimize the correspondences between the original sketch and the photo-to-sketch model G [LLM*19] via a dot product of these two images embedding produced by the CLIP-encoder [RKH*21] E: $\mathcal{L}_{\text{sketch}} = -E(G(x))^T E(I_s)$. Here the goal is to output photorealistic representation and the loss is specifically designed to achieve high-fidelity outputs. However, a notable limitation is the lack of explicit mechanisms to capture and preserve the user's original creative intent. While the technical focus on visual fidelity drives impressive results, the optimization process doesn't necessarily include constraints or guidance to ensure the generated content aligns with the conceptual goals or stylistic preferences the user had in mind when creating the initial sketch.

These limitations persist in subsequent approaches, including Li et al. with **MVControl** [LCZL25] and related works [CYW*24, WYZ*24]. Notably, MVControl diverges from using hand-drawn sketches, instead relying on NPR edge maps derived through Canny edge filters applied to images. MVControl employs dual optimization objectives: the standard Score Distillation Sampling loss ($\mathcal{L}_{\text{SDS}}$) and a hybrid variant ($\mathcal{L}_{SDS}^{\text{hybrid}}$) that integrates both 2D ($\mathcal{L}_{SDS}^{2D}$) and 3D ($\mathcal{L}_{SDS}^{3D}$). Similarly, **Sketch2NeRF** [CYW*24] uses synthetic sketches from OmniObject3D-Sketch dataset [WZF*23] and implements a camera-aware reconstruction loss. This composite loss function incorporates perceptual loss ($\mathcal{L}_{\text{LPIPS}}$) [ZIE*18] to capture high-level structural and stylistic image features, alongside L1 loss ($\mathcal{L}_{\text{L1}}$) to preserve pixel-level sketch details. Beyond these image-space objectives, Sketch2NeRF adopts geometric regularization techniques from DreamFusion [PJBM23] that constrain the underlying 3D structure, while also implementing random viewpoint regularization to ensure consistent 3D representation. This comprehensive approach prevents overfitting to input viewpoints and produces coherent geometry from novel angles, effectively addressing common issues such as view-dependent ambiguities, floating artifacts, and near-plane geometry distortions typical of NeRF-based approaches. More advances techniques evolved with the progress made in single image-based 3D generation [ZYG*24] via Guassian Splatting [KKLD23]. In fact, **Sketch3D** outputs Gaussian Splatting (survey [WYZ*24]) and SDS optimization strategies. The sketch loss $\mathcal{L}_{\text{sketch}}$ is a similarity loss ($\mathcal{L}_2$ loss) between the two CLIP-based (ResNet101) encoders with 4 layers. It uses a weighted color loss ($\mathcal{L}_{\text{Col}}$) to ensure consistent quality across viewpoints, and a structural loss $\nabla_\theta \mathcal{L}_{\text{S-SDS}}$) to align 3D structure with the sketch. These losses still focus on achieving photorealist results and do not include user's intent such as fabricability of the object or its costs.

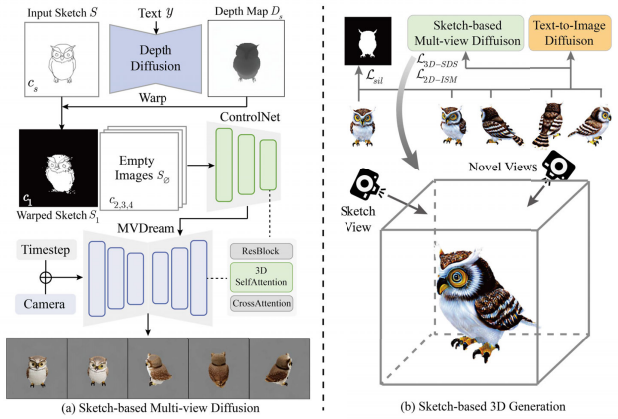


Figure 10: *SketchDream [LFLG24] shows a text-based generation augmented by the sketch. The Score Distillation Sampling is facilitated by the guidance provided by the sketch.*

Paper	Speed (min)	GPU	FT Dataset	Input	Losses
[MPS*23]	40*	3090	-	SketchMask	$\mathcal{L}_{SDS} + \mathcal{L}_{pres} + \mathcal{L}_{sil} + \mathcal{L}_{sp}$
[CPL*23]	60	V100	-	AS**	$\mathcal{L}_{sketch}$
[OCK*24]	-	-	-	Canny	$\mathcal{L}_{SDS}$
[LCZL25]	-	A100	Obja/LAION	Canny	$\mathcal{L}_{SDS}^{2D} + \mathcal{L}_{SDS}^{3D}$
[CYW*24]	120	3090	Omni+THum-S	-	$\mathcal{L}_{sketch} + \mathcal{L}_{reg} + \mathcal{L}_a$
[LFLG24]	75	A100	Objaverse	-	$\mathcal{L}_{SDS} + \mathcal{L}_{sil} + \mathcal{L}_{ISM}$
[ZXC*24]	3	4090	SS3D	Canny	$\mathcal{L}_{\text{sketch}} + \mathcal{L}_{\text{Col}} + \mathcal{L}_{\text{SDS}}$

Table 4: *Presents the evaluation performance of different methods. AS: Abstract Sketch. "*": fixed views. **: The specific abstract sketch style is not specified. FT Datasets: dataset used to fine-tune the model. SS3D: ShapeNet-Sketch3D dataset [ZXC*24]. The SketchDream loss 2D Interval Score Matching (ISM) is from LucidDreamer [LYL*23]. Obja: Objaverse. Omni+THum-S: OmniObject3D-S+THuman-S. "-": not specified.*

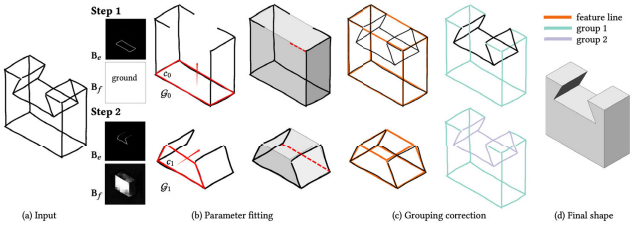


Figure 11: *Free2CAD [LPBM22] used transformers to group sketch input to specific CAD sequences. The geometry is divided in different, but they do not carry any semantic meaning, they are related to the CAD sequence of commands* (TB).

5.7. Models Summary

Examining Table 3 and Table 4, we notice that while newer methods deliver superior results, they make significant trade-offs in on-device real-time interaction capabilities. As illustrated in Figure 6, the evolution of DS-3DM methods has closely followed advancements in single-image to 3D generation techniques. This progression reveals distinct developmental phases, each with its own strengths and limitations: Neural network models initially offered precise parameter control but imposed rigid constraints on program-based output shapes, limiting creative flexibility. Deep generative models subsequently expanded output diversity but struggled to accurately capture design intent and generate high-quality 3D shapes. The introduction of diffusion models marked significant improvement in output quality, though these approaches often prioritized visual fidelity over practical design considerations. Transformer architectures brought important advances by effectively capturing relationships between sketch elements, while differentiable rendering techniques ensured view consistency across perspectives. Most recently, pre-trained optimization methods have achieved remarkable photorealistic results, albeit without adequately accounting for functional intent. This developmental trajectory reveals a consistent limitation across all approaches: optimization objectives overwhelmingly prioritize visual appearance over user intent factors such as functionality, fabricability, and other practical considerations. The predominant loss functions ($\mathcal{L}_{\text{SDS}}$, $\mathcal{L}_{\text{sketch}}$, $\mathcal{L}_{\text{Col}}$) reflect this emphasis on visual accuracy rather than addressing practical design considerations. For example, when generating a couch from a sketch, current systems optimize primarily for visual appeal rather than the designer's intent for factors like ease of assembly, cost-effective fabrication, or ergonomic functionality. Some emergent research has begun addressing these limitations, with novel approaches becoming more body-aware for wearable objects [GTC*24] and incorporating fabrication-aware constraints [MVDD24]. The high cost of training 3D foundation models has catalyzed a shift in DS-3DM research towards leveraging multimodal inputs. By combining sketches, images, and text, a new generation of models can generate editable 3D shapes. These approaches are broadly categorized by their reliance on strong 2D priors [LFLG24, ZHD*24], 3D priors [SKR*24], or program synthesis via Large Language 3D Modelers (LL3M) [LCD*25]. This paradigm facilitates the use of diverse generative architectures, such as autoregressive, flow, and diffusion models, which are designed to exploit both the sequential nature of text and sketches [HE17b] and the global properties of images. A particularly promising direction, especially for discrete representations like Computer-Aided Design (CAD), is Discrete Diffusion [SHW*25, SAG*24]. These mask-based models are adept at infilling tasks without a fixed generation order, making them ideal for part-based 3D representations. They are also optimized for limited-data regimes and can operate directly on discrete data, reducing the risk of overfitting common in autoregressive models due to their Evidence Lower Bound (ELBO) type of loss. To summarize the research community continues to prioritize two critical areas: developing information-rich part-aware representations and enabling 3D foundation models to ingest multiple modalities to generate several options that accurately reflect the user's intent as we will see in the next Section.

6. Outputs

In this section, we introduce Table 5 and structure the discussion around the outputs of DS-3DM. More importantly, we focus on how these outputs have been evaluated, including the metrics used to assess their quality and their alignment with user intent. Table 5 is organized chronologically along the vertical axis, while the horizontal axis examines three key aspects: part-based semantic information (see Section 6.1), the quantity of generated shapes (see Section 6.2), and the geometric characteristics of the generated shapes (see Section 6.3).

All these considerations are directly connected to how the output is evaluated via qualitative and quantitative metrics. Furthermore, Section 6.1, 6.2, and 6.3 present the relation between the user's intent and **metrics** used to evaluate the 3D output generated by DS-3DM methods. For a more in-depth report on different 3D representations, we invite the reader to review [XX23]. The 3D output representation is critical for the design of the DS-3DM model's architecture (see Supplementary 11 and 11.1). Therefore, to adopt end-to-end DS-3DM methods, the 3D representation must be differentiable (see Supplementary for an overview of the different 3D representations). Another critical aspect concerns the information embedded in the final 3D outputs.

6.1. Part-based semantics

This section categorizes DS-3DM methods based on how they structure output components, as shown in Table 5. This subcategorization indicates the:

- geometry division,
- part-based representation, or
- part-based representation with semantic information.

The first approach focuses on decomposing output into fundamental geometric entities. As recent works demonstrate, these decompositions manifest through various geometric forms, including topological features such as holes, and constructive geometrical operations such as extrusions. [FQS*24, LPBM22]. Other methods employ geometric primitives, such as Gaussians [XABP24, ZXC*24], cuboids [LDT*22, TSG*17], superquadrics [PUG19a, PUG19b] or mathematical surfaces like Coons patches [SBS21]. Additional geometric decomposition strategies are embraced in neural fields (NeRF) [CYW*24,CPL*23] and point cloud representations [WWF*23]. Notably, even when NeRF-based approaches produce visually distinct parts through color variation, these distinctions typically do not carry structural information. This consideration is expanded in Section 8, when we discuss editing capabilities for these parts.

The second approach incorporates higher-level structural meaning into component representations [MZC*19, KHA*22, LPG24]. These methods segment shapes into functionally meaningful parts [PLH*22, BHSH*24], such as the back, the seat, and the legs of a chair. This structural decomposition enhances understanding of component relationships, facilitates more intuitive editing and manipulation, improving the alignment with design intent.

The third approach extends part-based representations by incorporating additional semantic attributes. These attributes include

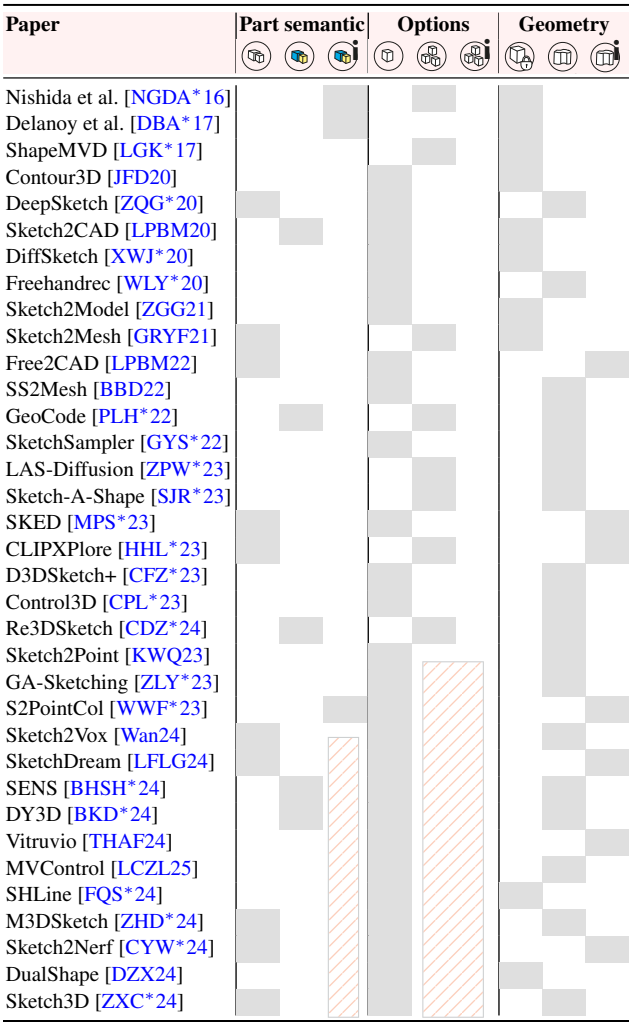


Table 5: DS-3DM *methods' Output.* : *part division but without annotations,* : *with annotations.* : *with annotation and info.* : *single output.* : *multiple options.* : *multiple options with info.* : *limited topology.* : *various topology.* : *Various topology with info. Red hatches show research gaps.*

appearance details [WWF*23], material properties, cost estimate, data, color information, and domain-specific metadata. This semantic enrichment enhances the usability and contextual relevance of generated models, supporting more informed design decisions.

Additionally, empty rows in Table 5 indicate when the shape is represented by a single unified entity [THAF24, SJR*23] as a overall mesh, constructive solid geometry, or another non-decomposable structure.

Metrics: These approaches are evaluated with metrics that capture both geometric accuracy and semantic meaningfulness. For a meaningful part-level segmentation, ad-hoc metrics like mean Intersection over Union (mIoU) are used. In Sketch2PointColor [WWF*23], mIoU quantify how point cloud part segmentation captures color features. Here, distinct colors represent different parts, two for table and three for an airplane, car, and chair. Another critical evaluation dimension is the structural stability assessment. GeoCode [PLH*22] implements a dual-criteria stability metric comprising structural integrity verification and physics-based simulation for practical stability assessment. This approach acknowledges that geometrically accurate reconstructions might not translate to physically stable objects, for example generated vessels with bases that are too narrow. Similar to Sketch2PointColor [WWF*23], NeRF-based approaches [LFLG24, CYW*24] often employ a two-stage methodology to enable editability. This two-stage approach is evaluated with perceptual metrics. Additional specialized metrics exist for component-specific evaluation, such as the retrieval-based metrics in DualShape [DZX24]. To ensure a valuable output, the 3D representation and its constituent parts must align with user intent and enable a controlled and informed design process [WS19]. User intent can vary widely; for instance, a user generating a "chair" may have specific goals for its parts. They might desire a more functional "armrest", a more cost-effective "seat" made from a different material, a more comfortable "backrest", or more structurally sound "legs". By defining these goals as the initial user intent, a part-based 3D model can provide performance feedback for each component. For example, it could calculate the maximum weight the "legs" can sustain, predict the comfort level of the "seat" with specific metrics, or estimate the aesthetic appeal for a target demographic. This performance-informed feedback empowers the user to control the design and drive further iterations effectively. See Table 6 for more evaluation metrics that are often used to evaluate the output as a whole.

6.2. Amount of Output Options

This section examines the number of output options generated by DS-3DM, as shown in Table 5. A fundamental challenge in DS-3DM arises from the inherent ambiguity of sketch inputs, as a single sketch captures only partial information about the intended 3D representation see Figure 2. Methods have addressed this challenge through techniques that enable sampling and generating multiple 3D variations. However, the ability to produce multiple high-quality options for users to evaluate, compare, and select from remains an ongoing challenge. Significant bottlenecks persist, particularly in model architecture and computational efficiency, limiting the feasibility of real-time, diverse, and high-quality output generation. The approaches to output generation can be categorized into three distinct categories:

- one 3D shape,
- multiple 3D shapes, or
- multiple 3D shapes with information.

Metrics: Recent work has begun to address the evaluation of multiple shape outputs and their diversity. For instance, Sketch-A-Shape [SJR*23] implements a comprehensive evaluation framework that measures accuracy across various input modalities. Their evaluation encompasses multiple datasets, including ImageNet-Sketch (IS-Acc), TU-Berlin Sketch (TU-Acc), ShapeNet-Sketch (SS-Acc), and QuickDraw (QD-Acc). Notably, Sketch-A-Shape

demonstrates the capability to generate multiple shapes per sketch query, with reported metrics based on the mean performance across five sampled shapes for each sketch input.

6.3. Geometry

This section categorizes output shapes based on their geometric complexity and topological properties, specifically considering the genus of the geometry (the number of "holes" or handles), as detailed in Table 5. We classify models according to their capability to generate:

- generates 3D shapes with *limited topologies*,
- generates 3D shapes with various topologies, or
- generates 3D shapes with various topologies and information.

This categorization represents a progression from low-entropy to increasingly complex representations. Early approaches focused on simple models with fixed parametric ranges, requiring regression of basic geometric properties such as height, width, and depth values [LPBM22, NGDA*16]. More sophisticated methods have evolved to generate complex parametric representations or shape programs [PLH*22, SBS21]. Other methods do not depend on initial templates or programs; instead, they rely on alternative output representations such as point clouds [WWF*23], unsigned distance functions (uSDF) [THAF24], signed distance functions (SDF) [ZGZS20], and other geometric representations [GRYF21, ZGG21]. The output of these models is typically a single complex 3D shape; however, recent work has achieved part-based representations with complex part geometry in dense objects directly from sketches, as demonstrated by SENS [BHSH*24] and DoodleYour3D [BKD*24]. However, critical information regarding part materials, costs, weights, and other practical attributes remains largely unexplored in current sketch-to-3D approaches. Indeed, in Vitruvio [THAF24], the authors envision future models capable of generating Building Information Models (BIMs) in Universal Scene Description (USD) formats, where detailed information about each building component is preserved. This includes maintenance specifications for windows, fire safety properties, thermal transfer coefficients for walls in energy analysis, and other critical parameters necessary for shape optimization and building performance evaluation.

Metrics: Some DS-3DM metrics presented here answers some of the initial questions presented in the introduction 1. Others [ZGZS20, MPS*23, LFLG24] evaluate how effectively the methods capture the geometric details of the sketch (see Table 6). How accurately must the sketch convey geometric information for the model to interpret it correctly? Finally, does the model account for the context of the sketch, including user preferences for color, texture, materials, and even broader design considerations? Does it understand the user's intent beyond the overall 3D shape? Does it integrate more nuanced elements like physics [RCDB23] for more functional design [GTC*24]? All these questions need to be translated in both qualitative and quantitative metrics capable to answer them. Therefore, to provide a clearer structure, this section focuses primarily on quantitative and qualitative metrics, as they are more directly related to the overall 3D geometry and connected with the user's intent. Based on insights from previous surveys, we further categorize the **quantitative** metrics discussion into reconstruction and generation-based metrics, reflecting the distinct evaluation approaches used in existing methods [YYW25]. After that we introduce the qualitative metrics.

Reconstruction-based metrics evaluate the final model output in direct relation to the dataset used to train DS-3DM models; here, we report a list of metrics with their strengths and weaknesses in correlating the user's intent with the output. For a more exhaustive and in-depth list, we recommend the reader to review [FAZ21]. *Normal consistency (NC)* measures the accuracy and completeness of the shape normals [ZLY*23, DZX24]. While it performs well for smooth surfaces, *NC* is sensitive to noise and insufficient for capturing global shape fidelity and topology. To address this limitation, *Earth Mover's Distance (EMD)* and *Chamfer Distance (CD)* have been adopted, as they effectively capture global features and reduce noise sensitivity. They measure the distance between points in two point clouds, and use these measurements to compare the reconstructed point cloud to a ground truth or the cost of moving point density from one point cloud to another [DZX24, ZGZS20, ZQG*20, BKD*24, WLY*20, ZPW*23, KWQ23]. These metrics capture the geometric output qualities for reconstruction tasks but are sensitive to outliers. There are metrics less prone to this sensitivity [ZQG*20] such as F-Score, Intersection over Union (IoU), Accuracy, and others. Specifically, F-Score evaluates the overlap between predicted and ground-truth point clouds based on precision and recall. *IoU* is used in retrieval tasks and provides an even more global evaluation. IoU measures how well the defined volumes overlap but overlooks fine details. IoU shows the model's accuracy in retrieving shapes from the dataset. While these metrics emphasize the accurate geometric reconstruction of shapes, they often fail to capture the subtleties of how humans perceive visual quality. Furthermore, these metrics primarily evaluate the generated output in isolation, without establishing a direct connection to the input sketch, assessing output quality by comparing it to similar shapes within the dataset, rather than considering its alignment with the user's original intent. As a result, they are more suited for retrieval-based tasks rather than capturing the nuances of sketch-to-3D generation, where user intent plays a more central role. To address disconnection with user's intent, novel metrics evaluate the similarity also in the image space: perceptual metrics shift the focus from strict geometric accuracy to more subjective qualities, such as realism and visual coherence. These perceptual metrics evaluate the model's output regarding shape fidelity and how well the generated images align with human perception. For example, *Structural Similarity (SSIM)* [MPS*23, ZXC*24] measures the similarity between two images, considering luminance, contrast, and structure. To add additional information and remove further ambiguity sketches have been paired with textual descriptions, and with the advent of CLIP [RKH*21, PJBM23, JMB*22], novel metrics now also focus on the alignment of generated images with text inputs [WWY25], enhancing the evaluation of models in tasks that involve both visual and textual data [HHL*23, MPS*23, LFLG24, CYW*24, XABP24]. *CLIP-score* measures fidelity to the user's input by combining CLIP-T, CLIP-I, and CLIP-S, which assess the similarity between generated shapes and a text

Paper	Qualitative Metrics					Quantitative Metrics		
	People	Age	Test	Topics	Scale	Reconstruction-Based	Generation-Based	Modeling-Based
Delanoy et al. [DBA*17]	6	-	2s	-	Likert	Speed, IoU	-	-
Sketch2CAD [LPBM20]	6	-	3s	-	Likert	-	-	Operator Accuracy
GeoCode [PLH*22]	12	-	Proc	-	Likert	CD	-	Robustness
GA-Sketching [ZLY*23]	8	18-28	-	U/A	SUS, NASA-TLX	CD, IoU, NC	-	CD-viewpoint
Control3D [CPL*23]	6	-	v	U	Pref. Score	-	-	-
CLIPXplore [HHL*23]	10	-	50v	Q/S	Likert	FID, CD, CLIP-Score	-	-
LAS-Diffusion [ZPW*23]	-	-	-	-	-	FID, CD, IoU, CLIP-Score	COV, MMD, 1-NNA	-
Sketch3D [ZXC*24]	9	-	50v	S/A	Likert	CD-sp2p, CD-sp2t, SSIM		-
DreamSketch [LFLG24]	41	18-40	25	Q/S/A	Likert	CD-mean, CD-std	-	-
Sketch2Vox [Wan24]	18	21-26	36s	Q/S	Likert	mIoU	-	-

Table 6: *Shows the qualitative and quantitative evaluation metrics used in different* DS-3DM *methods. In the 'Test' column, subjects evaluated the 3D shape, primarily represented as a 360° video, with the number of videos per shape indicated by the letter (v), or shape (s) generated with the related interface. In other studies, a series of images was used for evaluation, or 'Proc' for procedural parameters. The qualitative metrics include: Q for Quality, S for Similarity, A for Alignment, and U for Usability. 'SUS' for System Usability Scale questionnaire and NASA-TLX for the NASA Task Load Index. The Likert scale used for evaluation ranges from 1 to 5, except for Sketch2Vox and Delanoy et al. that adopted a 1-7 scale. Quantitative metrics reference the division based on the Table 7, regarding reconstruction, generation, and modeling focus of these methods. Reconstruction metrics include FID (Fréchet Inception Distance), CD (Chamfer Distance), CD-sp2p (Chamfer score between 2 points), CD-sp2t (Chamfer score between shape and target), SSIM (Structural Similarity Index), NC (Normal Consistency), CD-mean (mean Chamfer Distance), CD-std (standard deviation of Chamfer Distance), CD-viewpoint (specific for different viewpoints, with different elevation and azimuth values. Generation metrics incluse COV (Coverage measured with CD and EMD that stands for Earth Moving Distance), MMD (Minimum Matching Distance), 1-NNA (1-Nearest Neighbor Accuracy). For more details, see the Supplementary material.*

prompt in the latent space of the CLIP embedding. CLIP-I measures the similarity between generated shapes and a reference image. The CLIP-Similarity (CLIP-S) measures the consistency of generated shapes with given conditions in the CLIP space. The text-guided mode assesses the similarity between a rendered image of the shape and the target text. The sketch-guided mode compares the rendered sketch of the generated shape with the input sketch [MPS*23, ZXC*24, HHL*23, ZPW*23, CYW*24]. *Fréchet Inception Distance (FID)* [LCZL25, HHL*23, ZXC*24] evaluates the distance between the distributions of generated shapes and real shapes in a feature space learned by a pre-trained Inception network.

As established in the literature, the reconstruction-based metrics primarily quantify text alignment with user intent, aiming to recreate shapes that users have previously encountered. However, these metrics do not directly assess the ability to generate entirely novel and original shapes that fall outside the distribution of the training dataset. To address this, **generation-based metrics** have been introduced, specifically designed to evaluate the diversity of newly generated shapes [WWF*23, BKD*24]. These metrics measure how effectively deep generative models sample from learned distributions, capturing variations beyond those present in the original dataset. Hence, it is difficult to compare these distributions, the original dataset distribution, and the generated sample distributions. These distributions should be similar, representing that the generated 3D shapes preserve similar characteristics, but the individual shapes should be slightly different in their details. This requirement is also a limitation of current models, which still do not generalize for out-of-distribution tasks and for this specific evaluation need access to the original training dataset. These DS-3DM face challenges in generating entirely novel 3D shapes, which is closely tied to the inherent challenge of quantifying such novelty. To properly measure the similarity between these two distributions, the DS-3DM should sample an amount of 3D shapes comparable with the number of 3D shapes present in the test set. Coverage (COV) measures the percentage of test samples covered by generated samples. A test sample is covered if it is the nearest neighbor of a generated sample [HPG*22, ZPW*23]. Minimum Matching Distance (MMD) measures the distance between test samples and generated samples [NKR*22]. 1-Nearest Neighbor Accuracy (1-NNA) measures how well test samples and generated samples are mixed by penalizing samples from the test set and generated set that have their nearest neighbor in the same set. COV, MMD, 1-NNA cannot be used without a proper test set, as in the case of pre-trained foundation models.

The main limitation of quantitative metrics is that they are disconnected with the user's intent, and often they do not consider the human interaction with the sketch-based interface, lacking a deep connection between input sketch and output shape. To address this limitation surveys and user studies are used to evaluate DS-3DM methods **qualitatively** (see Table 6). They often rely on common perception studies from the Human-Computer Interaction (HCI) community [OK14]. For example, DreamSketch [LFLG24] uses the one-way ANOVA test and box plots and paired T-test to evaluate its model's performance. Participants used a 1–5 Likert scale to evaluate text faithfulness, sketch faithfulness, geometry quality, texture quality, and overall quality. The participants reviewed 20 cases; each was created by pairing a 3D shape with the sketch and text inputs used to generate it, a similar pairing is performed in Sketch2Vox [Wan24]. To evaluate the 3D shapes, some DS-3DM

methods allow a free rotation of the 3D digital object, while other methods present videos [CPL*23] (the video is captured with a full rotation in azimuth and a fixed elevation of 15 degrees). Evaluating the final 3D shape based on someone else's sketch and text fails to capture the original user's intent and the abstract vision they had in mind while creating the object. This disconnect can obscure how accurately the generated 3D model reflects the initial creative process. Additionally, most of these methods conducted qualitative evaluations on shapes with a similar genus, which may not fully showcase the model's awareness capabilities for topology diversification. DeepSketch [ZGZS20] aimed to evaluate the shape diversity that could be outputted from their model. They selected questions and evaluation metrics guided by these principles: easy to sketch, generality, view differentiability, shape genius higher than 1, and sizable inter-category variance. They generated three shape categories from ShapeNet [CFG*15] with distinctive styles (naive, stylized, and style-unified). They considered professional drawings. However not all DS-3DM users are professional designers, therefore, how can DS-3DM be adapted to novice designers. For example, one method that considers amateur or novice sketching style is Doodle your 3D [BKD*24], the authors asked 30 users to draw ten sketches each on the demo canvas (Gradio) and rate the generated shapes based on how well they matched their expectations. This approach addresses a key limitation in the evaluation method of DreamSketch [LFLG24] showed in Fig. 10, where the participants in the evaluation of the 3D outputs were different from the users providing the initial sketch, making it impossible to track intent accurately. To improve future studies, we recommend that researchers have participants use their novel DS-3DM methods for generating and evaluating these shapes, ensuring a more direct alignment between input and output. This intent-aligned evaluation captures if the user's concepts and ideas are faithfully translated into 3D shapes. However, the final 3D representation should also be assessed in terms of its editing capabilities, which we do not cover in this report. For instance, Sketch2CAD [LPBM20] enrolled six novices in their user studies. These participants were asked to evaluate whether the DS-3DM method accurately translated their intended CAD operations and parameters properly, given the iterative nature of the Sketch2CAD method (see the Supplementary materials 11.2 for more). An intent-aligned evaluation example is provided by DualShape [DZX24]. DualShape's authors asked, "Did the generated car shell models meet your expectations?" While a simple Yes or No answer may not fully evaluate the method, it is a step toward better understanding and addressing user needs.

6.4. Output Summary

This section presents three key insights derived from analyzing the output representations in Table 5 and evaluation metrics in Table 6 across DS-3DM methods. Most approaches generate holistic 3D shapes without explicit part decomposition, limiting their utility for downstream editing tasks. Part-aware representations require models to distinguish individual components through one of several strategies: (1) part-level latent codes [KYNS23], where each part occupies a distinct region in latent space [BHSH*24, HPG*22, LJK*25]; (2) neurosymbolic decomposition [PLH*22, RGJ*23], which represents shapes as executable programs [MB21]; or (3) hierarchical attention mechanisms that parse part boundaries. However, these part-aware architectures introduce substantial computational overhead, explaining their limited adoption despite clear advantages for interactive modeling workflows.

Second, current neural sketch methods rarely support multi-candidate generation, hindering iterative design workflows. Only two methods [SJR*23, GRYF21] generate multiple output variations from a single sketch, and none provide auxiliary information such as material cost, and engineering performance. This contrasts with procedural generative design pipelines [MWH*23], which routinely produce several design alternatives in parallel for comparative evaluation. Procedural approaches achieve this scalability by leveraging computationally efficient parametric models that can generate variants through parameter sampling rather than full neural inference.

Finally, methods are shifting toward texture-enriched representations, but evaluation frameworks lag behind. Recent approaches generate geometry alongside surface appearance [LFLG24], yet existing metrics only assess geometric accuracy. This creates a gap in evaluating whether generated textures align with user intent. Texture synthesis fundamentally requires optimization-based architectures with differentiable rendering pipelines that iteratively refine appearance through gradient descent, enabling explicit control over material parameters. As discussed in Section 8, scenarios like game development, and architecture need perceptual evaluations measured via user preference studies, to properly validate the texture and material information outputted.

7. Ethical Considerations

Deep sketch-based 3D modeling necessitates a human-centric paradigm: creation begins with a user's sketch and additional information, making full automation neither feasible nor desirable. Key ethical challenges lie in defining the scope of human control: What constitutes a "sufficiently detailed" sketch? What additional information should be provided? And for which audiences and application domains? Historically, the relationship between sketch and 3D shapes has followed a specific direction: observing three-dimensional reality and translating it into two-dimensional sketches [Cav05] (Table 7, where initial approaches focused on reconstruction). Producing an effective sketch demands mastery of proportion, perspective, and observational acuity, skills codified by figures such as Brunelleschi, Leon Battista Alberti, and Leonardo da Vinci [Cav05, Fan15, LHHW25]. Sketching, like writing, is not merely a communicative act but also a cognitive one [FBCW23]; it shapes perception and thought [KHY*25, AS01, FHWG20]. Now, with these novel deep sketch-based 3D modeling approaches, sketches become the basis for generating 3D forms. This inversion raises ethical questions: Should these tools simply output 3D shapes, or should they also foster user learning, helping users refine sketching skills to better communicate ideas and participate in collaborative creation? Should we study and analyze the consequences of these novel tools and how they affect our creative processes, how they augment us? Implementing a human-centered approach while keeping these questions and considerations in mind could support the development of novel metrics more aligned with keeping humans in control and defining the level of information required as input and the specific information

to provide as output. As these tools become widespread, should DS-3DM's methods augment humans, reinforcing their sense of community, feeling of belonging, and creative empowerment? If so, how [WWH*25b, SSG*23, WHE*24, WKYS23, U.S24]?

8. Discussion and Future Directions

This section demonstrates how the MORPHEUS design space can guide research in DS-3DM by presenting concrete application scenarios. Each case study illustrates an integrated workflow from input to output, with entry points at different stages of the IMO-based design space.

First, consider the challenge of generating production-ready 3D assets from concept sketches in game development scenarios. Researchers can begin by examining the **Model** dimension, noting the historical shift [OSCSJ09, DL16, BAC*19] to data-driven approaches enabled by breakthroughs in computer vision and graphics (Figure 6). Using MORPHEUS to constrain the design space (fixing Input to "multiple sketches and annotations" typical of professional storyboards and Output to "high-fidelity 3D assets with various topologies and material information") reveals a clear architectural trend. Our comprehensive comparison table (Supplementary Material) shows that recent methods predominantly employ pre-trained foundation models with optimization-based generation [ZXC*24, CYW*24, ZLZ*25], producing Neural Radiance Fields or Gaussian Splatting representations. This landscape analysis directly informs evaluation strategy. For game development, the critical trade-off lies between visual quality and runtime performance: lower polygon meshes enable fluid gameplay but may sacrifice fidelity. Proper assessment therefore requires both quantitative geometry metrics and user-centered perceptual studies that measure whether quality degradation impacts the player experience.

Second, in preliminary design phases, generating multiple plausible options is essential for productive client-designer dialogue. Analyzing the **Output** categories in Table 5 through MORPHEUS reveals a critical gap: while some methods generate multiple geometric variations [SJR*23] and others provide part-level semantics [WWF*23], no existing method efficiently produces multiple options with associated metadata (e.g., cost estimates, performance metrics) in real-time (as discussed in Section 6.4). This gap defines a clear research opportunity. Consider an architect sketching an office building layout. An ideal DS-3DM system would generate multiple design alternatives, each annotated with energy performance, construction cost, material specifications, and buildability constraints. To address this requirement, researchers can use MORPHEUS to specify the target output: *multiple semantic options* represented as lightweight graph-based room layouts with cost metadata. Consulting the framework's model taxonomy identifies discrete diffusion models as strong candidates, given their success in graph-based architectural generation [CCL*21, SNL*21] and discrete representation spaces [IKSS*23, SHW*25]. With this output configuration fixed, researchers can identify appropriate baselines sharing similar input modalities and establish evaluation metrics. These metrics must assess both diversity across generated options and fidelity to the original design intent [VSZ*25], alongside domain-specific measures like spatial efficiency and regulatory compliance.

Finally, researchers can also begin from the **Input** dimension in Table 1 by targeting a specific audience [ZLD16, DPS15, PKM*11]. For example, let's consider democratizing industrial design for novice users, such as users designing their custom footwear. The research problem begins with accommodating *flexible sketch styles* and *multiple views* to resolve geometric ambiguity. Consulting Table 1 reveals that methods supporting both "Multiple Sketches" and "Flexible Style" are sparse or absent, immediately isolating a high-value research direction. Even without direct competitors, researchers can identify partial baselines sharing one of these traits for comparison. This input constraint propagates downstream: outputs must be fabrication-ready CAD formats [LPBM22, PLH*22] or Signed Distance Functions better suited for Finite Element Analysis [LEL*25, LPBM20, GRYF21, DZX24, BHSH*24], excluding implicit representations like NeRF that resist physical simulation. The model architecture should therefore employ differentiable rendering with optimization losses tuned for manufacturing constraints such as material cost and structural durability.

The diverse requirements across these scenarios, ranging from environmental performance and fabrication constraints to player experience [DZX24, XNW*24, USGB24], underscore the necessity for comprehensive **human-centered metrics**. The evaluation of DS-3DM presents unique challenges: while existing approaches prioritize reconstruction fidelity, they often fail to capture the critical human factors that determine interface success, such as interaction fluidity and cognitive load. As input technologies evolve from standard tablets to novel brain-computer interfaces—for instance, integrating EEG data [BWpC*24] in evaluation protocols could be considered. Future research should aim to standardize these user experience metrics, establishing benchmarks that balance technical geometric precision with user satisfaction and intent alignment across varying levels of expertise.

9. Conclusion

We survey previous work in **Deep Sketch-Based 3D Modeling** (DS-3DM) and propose MORPHEUS, a design space designed around the input-model-output framework. Through MORPHEUS we highlight limitations and find venues for future research. DS-3DM techniques combine sketch modeling with deep learning to generate 3D representations for design applications. We emphasize the need for a controlled and informed design where the user's intent can be better-captured thanks to novel evaluation metrics. Furthermore, MORPHEUS facilitates the evaluation and categorization of DS-3DM methods, highlighting their trends: higher level of 3D control and information-richer outputs. Finally, MORPHEUS helps researchers to identify limitations of previous methods, thus offering future research directions, and supports industry-focused readers in selecting the most suitable method and metrics for their use cases.

10. Acknowledgment

We acknowledge Karen Liu, Yael Vinker, Judith Fan, Alexandra Bonnicci, Dima Smirnov, Elena Colombini and Paul Guerrero for their feedback. Furthermore, the authors thank Hannah Luxenberg-Tono, Alberto Tauiti, Simge Girgin, Eleni Alexandraki,

Luc Houriez, Allie Cemalovic, Bochen Zhang, Alissa Cooperman, Andrej Krevl, Simi Aluko, Robyn Brinks Lockwood, Lisa Modifica, Veronica Augustina Bot, Samantha Bennett, Tara Srirangarajan, Collin Anthony Chen, and Yulia Gryaditskaya for their inspiration, suggestions, reviews, and support throughout the publication. The work is supported by CIFE Seed Grants, the Wu Tsai Neurosciences Institute and the Koret Human Neurosciences Community Lab (HNCL), Amazon (AWS), NVIDIA, Adobe, Google, the McCoy Family Center for Ethics in Society, and Stanford HAI.

References

[AFCO*25] ARAR E., FRENKEL Y., COHEN-OR D., SHAMIR A., VINKER Y.: SwiftSketch: A Diffusion Model for Image-to-Vector Sketch Generation. In Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers (New York, NY, USA, 2025), SIGGRAPH Conference Papers '25, Association for Computing Machinery. URL: https://doi.org/10.1145/3721238.3730612, doi:10.1145/3721238.3730612. 6

[AGB04] ALEXE A., GAILDRAT V., BARTHE L.: Interactive modelling from sketches using spherical implicit functions. In Proceedings of the 3rd International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa (New York, NY, USA, 2004), AFRIGRAPH '04, Association for Computing Machinery, p. 25–34. URL: https://doi.org/10.1145/1029949.1029953, doi:10.1145/1029949.1029953. 3

[AS01] AGRAWALA M., STOLTE C.: Rendering effective route maps: improving usability through generalization. In Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (New York, NY, USA, 2001), SIGGRAPH '01, Association for Computing Machinery, p. 241–249. URL: https://doi.org/10.1145/383259.383286, doi:10.1145/383259.383286. 18

[BAC*19] BONNICI A., AKMAN A., CALLEJA G., CAMILLERI K. P., FEHLING P., FERREIRA A., HERMUTH F., ISRAEL J. H., LANDWEHR T., LIU J., ET AL.: Sketch-based interaction and modeling: where do we stand? Artificial Intelligence for Engineering Design, Analysis and Manufacturing 33, 4 (2019), pp.370–388. doi:10.1017/S0890060419000349. 2, 19

[BBD22] BHARDWAJ N., BHARADWAJ D., DUBEY A.: SingleSketch2Mesh : Generating 3D Mesh model from Sketch, 2022. URL: https://arxiv.org/abs/2203.03157, arXiv:2203.03157. 5, 9, 15, 31, 32

[BG23] BERARDI G., GRYADITSKAYA Y.: Fine-Tuned but Zero-Shot 3D Shape Sketch View Similarity and Retrieval. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops (October 2023), pp. 1775–1785. 4

[BHA*21] BOMMASANI R., HUDSON D. A., ADELI E., ALTMAN R., ARORA S., VON ARX S., BERNSTEIN M. S., BOHG J., BOSSELUT A., BRUNSKILL E., BRYNJOLFSSON E., BUCH S., CARD D., CASTELLON R., CHATTERJI N., CHEN A. S., CREEL K., DAVIS J. Q., DEMSZKY D., DONAHUE C., DOUMBOUYA M., DURMUS E., ERMON S., ETCHEMENDY J., ETHAYARAJH K., FEI-FEI L., FINN C., GALE T., GILLESPIE L., GOEL K., GOODMAN N. D., GROSSMAN S., GUHA N., HASHIMOTO T., HENDERSON P., HEWITT J., HO D. E., HONG J., HSU K., HUANG J., ICARD T., JAIN S., JURAFSKY D., KALLURI P., KARAMCHETI S., KEELING G., KHANI F., KHATTAB O., KOH P. W., KRASS M. S., KRISHNA R., KUDITIPUDI R., ET AL.: On the Opportunities and Risks of Foundation Models. arXiv pre-print (2021). URL: https://arxiv.org/abs/2108.07258, arXiv:2108.07258, doi:https://doi.org/10.48550/arXiv.2108.07258. 12

[BHSH*24] BINNINGER A., HERTZ A., SORKINE-HORNUNG O., COHEN-OR D., GIRYES R.: SENS: Part-Aware Sketch-based Implicit Neural Shape Modeling. Computer Graphics Forum 43, 2 (2024), e15015. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.15015, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.15015, doi:https://doi.org/10.1111/cgf.15015. 5, 9, 11, 14, 15, 16, 18, 19, 31, 32, 33

[BKD*24] BANDYOPADHYAY H., KOLEY S., DAS A., BHUNIA A. K., SAIN A., CHOWDHURY P. N., XIANG T., SONG Y.-Z.: Doodle Your 3D: From Abstract Freehand Sketches to Precise 3D Shapes. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2024), 9795 – 9805. doi:10.1109/CVPR52733.2024.00935. 3, 5, 6, 7, 8, 9, 10, 11, 12, 15, 16, 17, 18, 32, 33

[BKK*22] BHUNIA A. K., KOLEY S., KHILJI A. F. U. R., SAIN A., CHOWDHURY P. N., XIANG T., SONG Y.-Z.: Sketching Without Worrying: Noise-Tolerant Sketch-Based Image Retrieval. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022), pp. pp.999–1008. doi:10.1109/CVPR52688.2022.00107. 4

[BTLLW22] BOND-TAYLOR S., LEACH A., LONG Y., WILLCOCKS C. G.: Deep Generative Modelling: A Comparative Review of VAEs, GANs, Normalizing Flows, Energy-Based and Autoregressive Models. IEEE Transactions on Pattern Analysis and Machine Intelligence 44, 11 (Nov 2022), 7327–7347. doi:10.1109/TPAMI.2021.3116668. 9

[BWpC*24] BAI Y., WANG X., PEI CAO Y., GE Y., YUAN C., SHAN Y.: DreamDiffusion: High-Quality EEG-to-Image Generation with Temporal Masked Signal Modeling and CLIP Alignment. Computer Vision – ECCV 2024: 18th European Conference, Milan, Italy, September 29–October 4, 2024, Proceedings, Part XXXI (2024), 472–488. doi:10.1007/978-3-031-72751-1_27. 10, 19

[CA09] COOK M. T., AGAH A.: A survey of sketch-based 3-D modeling techniques. Interacting with Computers 21, 3 (05 2009), 201–211. URL: https://doi.org/10.1016/j.intcom.2009.05.004, arXiv:https://academic.oup.com/iwc/article-pdf/21/3/201/2353735/iwc21-0201.pdf, doi:10.1016/j.intcom.2009.05.004. 3

[Can86] CANNY J.: A Computational Approach to Edge Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8, 6 (Nov 1986), 679–698. doi:10.1109/TPAMI.1986.4767851. 6, 7, 11

[Cav05] CAVANAGH P.: The artist as neuroscientist. Nature 434, 7031 (March 2005), 301–307. URL: https://doi.org/10.1038/434301a, doi:10.1038/434301a. 18

[CBK*24] CHEN D., BHUNIA A., KOLEY S., SAIN A., CHOWDHURY P., SONG Y.: DemoCaricature: Democratising Caricature Generation with a Rough Sketch. In 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (Los Alamitos, CA, USA, jun 2024), IEEE Computer Society, pp. 8629–8639. URL: https://doi.ieeecomputersociety.org/10.1109/CVPR52733.2024.00824, doi:10.1109/CVPR52733.2024.00824. 3

[CBS*23] CHOWDHURY P. N., BHUNIA A. K., SAIN A., KOLEY S., XIANG T., SONG Y.-Z.: Democratising 2D Sketch to 3D Shape Retrieval Through Pivoting. In 2023 IEEE/CVF International Conference on Computer Vision (ICCV) (Oct 2023), pp. 23218–23229. doi:10.1109/ICCV51070.2023.02127. 4, 6, 7, 11

[CCL*21] CHANG K., CHENG C., LUO J., MURATA S., NOURBAKHSH M., TSUJI Y.: Building-GAN: Graph-Conditioned Architectural Volumetric Design Generation. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2021), pp.11956–11965. URL: https://arxiv.org/abs/2104.13316, arXiv:2104.13316, doi:https://doi.org/10.48550/arXiv.2104.13316. 7, 19

[CCL*24] CHEN T., CAO R., LI Z., ZANG Y., SUN L.: Deep3DSketch-im: rapid high-fidelity AI 3D model generation by single freehand sketches. Frontiers of Information Technology and Electronic

Engineering 25, 1 (2024), 149 – 159. doi:10.1631/FITEE.2300314. 3

[CDI22] CHAN C., DURAND F., ISOLA P.: Learning to generate line drawings that convey geometry and semantics. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022). arXiv:2203.12691, doi:10.1109/CVPR52688.2022.00776. 10, 33

[CDZ*24] CHEN T., DING C., ZHU L., ZANG Y., LIAO Y., LI Z., SUN L.: Reality3DSketch: Rapid 3D Modeling of Objects from Single Free-hand Sketches. IEEE Transactions on Multimedia 26 (2024), 4859 – 4870. doi:10.1109/TMM.2023.3327533. 3, 5, 9, 15, 32

[CFG*15] CHANG A. X., FUNKHOUSER T., GUIBAS L., HANRAHAN P., HUANG Q., LI Z., SAVARESE S., SAVVA M., SONG S., SU H., XIAO J., YI L., YU F.: ShapeNet: An Information-Rich 3D Model Repository. arXiv pre-print (2015). doi:https://doi.org/10.48550/arXiv.1512.03012. 3, 7, 8, 18, 33

[CFZ*23] CHEN T., FU C., ZANG Y., ZHU L., ZHANG J., MAO P., SUN L.: Deep3DSketch+: Rapid 3D Modeling from Single Free-hand Sketches. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 13834 LNCS (2023), 16 – 28. doi:10.1007/978-3-031-27818-1_2. 5, 9, 12, 15, 32

[CG23] CHAO C.-K. T., GINGOLD Y.: Text-guided Image-and-Shape Editing and Generation: A Short Survey, 2023. URL: https://arxiv.org/abs/2304.09244, arXiv:2304.09244. 3

[CGL*23] COLE F., GOLOVINSKIY A., LIMPAECHER A., STODDART BARROS H., FINKELSTEIN A., FUNKHOUSER T., RUSINKIEWICZ S.: Where Do People Draw Lines?, 1 ed. Association for Computing Machinery, New York, NY, USA, 2023. URL: https://doi.org/10.1145/3596711.3596756. 6

[Cha24] CHATTERJEE S.: Free-form Shape Modeling in XR: A Systematic Review, 2024. URL: https://arxiv.org/abs/2401.00924, arXiv:2401.00924. 3

[CLPK24] CHOI C., LEE J., PARK J., KIM Y. M.: 3Doodle: Compact Abstraction of Objects with 3D Strokes. ACM Trans. Graph. 43, 4 (jul 2024). 3

[CLT*23] CHENG Y.-C., LEE H.-Y., TULYAKOV S., SCHWING A., GUI L.: SDFusion: Multimodal 3D Shape Completion, Reconstruction, and Generation, 2023. arXiv:2212.04493. 3

[CMR90] CARD S. K., MACKINLAY J. D., ROBERTSON G. G.: The design space of input devices. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (New York, NY, USA, 1990), CHI '90, Association for Computing Machinery, p. 117–124. URL: https://doi.org/10.1145/97243.97263, doi:10.1145/97243.97263. 3, 4

[CPL*23] CHEN Y., PAN Y., LI Y., YAO T., MEI T.: Control3D: Towards Controllable Text-to-3D Generation. In Proceedings of the 31st ACM International Conference on Multimedia (New York, NY, USA, 2023), MM '23, Association for Computing Machinery, p. 1148–1156. 5, 9, 10, 12, 13, 14, 15, 17, 18, 31, 32

[CRX*19] CHAUDHURI S., RITCHIE D., XU K., JIAJUN W., ZHANG H. R.: Learning Generative Models of 3D Structures. In Eurographics 2019 - Tutorials (2019), Jakob W., Puppo E., (Eds.), The Eurographics Association. doi:10.2312/egt.20191038. 3

[CSB*22] CHOWDHURY P. N., SAIN A., BHUNIA A. K., XIANG T., GRYADITSKAYA Y., SONG Y.-Z.: FS-COCO: Towards Understanding of Freehand Sketches of Common Objects in Context. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 13668 LNCS (2022), 253 – 270. doi:10.1007/978-3-031-20074-8_15. 4

[CWC*22] CHOWDHURY P., WANG T., CEYLAN D., SONG Y., GRYADITSKAYA Y.: Garment Ideation: Iterative View-Aware Sketch-Based Garment Modeling. In 2022 International Conference on 3D Vision (3DV) (Los Alamitos, CA, USA, sep 2022), IEEE Computer Society, pp. 22–31. URL: https://doi.ieeecomputersociety.org/10.1109/3DV57658.2022.00015, doi:10.1109/3DV57658.2022.00015. 3

[CWL*21] CONGCONG W., WENYU H., LAZARUS C., YAN LIANG T., CHAN S. L., HANG Z., CHEN F.: RealCity3D: A Large-scale Georeferenced 3D Shape Dataset of Real-world Cities. arXiv pre-print (2021). 7

[CYH*25] CHEN T., YU C., HU Y., LI J., XU T., CAO R., ZHU L., ZANG Y., ZHANG Y., LI Z., SUN L.: Img2CAD: Conditioned 3-D CAD Model Generation From Single Image With Structured Visual Geometry. IEEE Transactions on Industrial Informatics (2025). doi:10.1109/TII.2025.3584476. 3

[CYW*24] CHEN M., YUAN W., WANG Y., SHENG Z., HE Y., DONG Z., BO L., GUO Y.: Sketch2NeRF: Multi-view Sketch-guided Text-to-3D Generation, 2024. URL: https://arxiv.org/abs/2401.14257, arXiv:2401.14257. 2, 5, 7, 9, 13, 14, 15, 16, 17, 19, 32, 33

[CZH23] CHEN T., ZHANG R., HINTON G.: Analog bits: Generating discrete data using diffusion models with self-conditioning. In The Eleventh International Conference on Learning Representations (2023). URL: https://openreview.net/forum?id=3itjR9QxFw. 10

[CZJ*22] CHANG H., ZHANG H., JIANG L., LIU C., FREEMAN W. T.: MaskGIT: Masked Generative Image Transformer, June 2022. 11

[CZSY25] CAO P., ZHOU F., SONG Q., YANG L.: Controllable Generation with Text-to-Image Diffusion Models: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence (2025), 1–20. doi:10.1109/TPAMI.2025.3646548. 4, 10

[DBA*17] DELANOY J., BOUSSEAU A., AUBRY M., ISOLA P., EFROS A. A.: What You Sketch Is What You Get: 3D Sketching using Multi-View Deep Volumetric Prediction. Proceedings of the Conference on Human Factors in Computing Systems Extended Abstracts (CHI EA) (2017). URL: https://doi-org.stanford.idm.oclc.org/10.1145/3290607.3312847, doi:10.1145/3290607.3312847. 5, 8, 9, 10, 15, 17, 32, 33

[DCLB19] DELANOY J., COEURJOLLY D., LACHAUD J.-O., BOUSSEAU A.: Combining voxel and normal predictions for multi-view 3D sketching. Computers & Graphics 82 (2019), 65–72. URL: https://www.sciencedirect.com/science/article/pii/S0097849319300858, doi:https://doi.org/10.1016/j.cag.2019.05.024. 8, 10

[DCLT19] DEVLIN J., CHANG M.-W., LEE K., TOUTANOVA K.: BERT: Pre-training of deep bidirectional transformers for language understanding. NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference 1 (2019), 4171 – 4186. 10

[DDS*09] DENG J., DONG W., SOCHER R., LI L.-J., LI K., FEI-FEI L.: Imagenet: A large-scale hierarchical image database. Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2009), pp.248–255. doi:10.1109/CVPR.2009.5206848. 2

[DFRS03a] DECARLO D., FINKELSTEIN A., RUSINKIEWICZ S., SANTELLA A.: Suggestive contours for conveying shape. ACM Transactions on Graphics (TOG) 22, 3 (2003), 848–855. 7

[DFRS03b] DECARLO D., FINKELSTEIN A., RUSINKIEWICZ S., SANTELLA A.: Suggestive Contours for Conveying Shape. ACM Transactions on Graphics (SIGGRAPH) 22, 3 (2003), pp.848–855. URL: https://doi-org.stanford.idm.oclc.org/10.1145/882262.882354, doi:10.1145/882262.882354. 7

[DHF*22] DU D., HAN X., FU H., WU F., YU Y., CUI S., LIU L.: SAniHead: Sketching Animal-Like 3D Character Heads Using a View-Surface Collaborative Mesh Generative Network. IEEE Transactions on Visualization and Computer Graphics 28, 6 (June 2022), 2415–2429. doi:10.1109/TVCG.2020.3030330. 3

[DL16] DING C., LIU L.: A survey of sketch based modeling systems. Frontiers of Computer Science 10, 6 (2016), 985–999. As 3D technology, including computer graphics, virtual reality, and 3D printing, has been rapidly developed in the past years, 3D models are gaining an increasingly huge demand. URL: https://doi.org/10.1007/s11704-016-5422-9, doi:10.1007/s11704-016-5422-9. 2, 19

[DLCS*23] DE LUIGI L., CARDACE A., SPEZIALETTI R., RAMIREZ P. Z., SALTI S., DI STEFANO L.: DEEP LEARNING ON IMPLICIT NEURAL REPRESENTATIONS OF SHAPES. In 11th International Conference on Learning Representations, ICLR 2023 (2023). Cited by: 13. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85162174705&partnerID=40&md5=343d4bd2729e5650b93ff1fb7233edd1. 9

[DLD*21] DENG C., LITANY O., DUAN Y., POULENARD A., TAGLIASACCHI A., GUIBAS L. J.: Vector Neurons: A General Framework for SO(3)-Equivariant Networks. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2021), pp.12180–12189. URL: https://doi.org/10.48550/arXiv.2104.12229, doi:2104.12229. 7

[DLP*23] DENG Z., LIU Y., PAN H., JABI W., ZHANG J., DENG B.: Sketch2PQ: Freeform Planar Quadrilateral Mesh Design via a Single Sketch. IEEE Transactions on Visualization and Computer Graphics 29, 9 (2023), 3826 – 3839. doi:10.1109/TVCG.2022.3170853. 3

[DN21] DHARIWAL P., NICHOL A.: Diffusion Models Beat GANs on Image Synthesis. NIPS '21, Curran Associates Inc. 12

[DPS15] DE PAOLI C., SINGH K.: SecondSkin: sketch-based construction of layered 3D models. ACM Trans. Graph. 34, 4 (July 2015). URL: https://doi.org/10.1145/2766948, doi:10.1145/2766948. 19

[DSC*20] DVOROŽŇÁK M., SÝKORA D., CURTIS C., CURLESS B., SORKINE-HORNUNG O., SALESIN D.: Monster Mash: A Single-View Approach to Casual 3D Modeling and Animation. ACM Transactions on Graphics 39, 6 (2020). 3

[DSS*22] DEITKE M., SCHWENK D., SALVADOR J., WEIHS L., MICHEL O., VANDERBILT E., SCHMIDT L., EHSANI K., KEMBHAVI A., FARHADI A.: Objaverse: A Universe of Annotated 3D Objects, 2022. URL: https://arxiv.org/abs/2212.08051, arXiv:2212.08051. 3, 7, 8

[DZX24] DU X., ZHANG T., XIE H.: DualShape: Sketch-Based 3D Shape Design With Part Generation and Retrieval. IEEE Access 12 (2024), 18888–18900. doi:10.1109/ACCESS.2024.3361659. 5, 9, 15, 16, 18, 19, 31, 32

[Ebe24] EBERT D.: 3D Arena, 2024. URL: https://huggingface.co/spaces/dylanebert/3d-arena. 3

[EHA12] EITZ M., HAYS J., ALEXA M.: How do humans sketch objects? ACM Trans. Graph. 31, 4 (July 2012). URL: https://doi.org/10.1145/2185520.2185540, doi:10.1145/2185520.2185540. 2, 31

[Eng62a] ENGELBART D.: Augmenting Human Intellect: A Conceptual Framework. In Summary Report AFOSR-3223 (1962). Published by Stanford Research Institute, Menlo Park, CA. URL: https://dougengelbart.org/content/view/138. 3

[Eng62b] ENGELBART D. C.: Augmenting Human Intellect: A Conceptual Framework. Air Force Office of Scientific Research, AFOSR-3233, www.bootstrap.org/augdocs/friedewald030402/augmentinghumanintellect/ahi62index.html, 1962. 4

[ERB*12] EITZ M., RICHTER R., BOUBEKEUR T., HILDEBRAND K., ALEXA M.: Sketch-based shape retrieval. ACM Trans. Graph. 31, 4 (July 2012). URL: https://doi.org/10.1145/2185520.2185527, doi:10.1145/2185520.2185527. 4

[Fan15] FAN J. E.: Drawing to learn: How producing graphical representations enhances scientific thinking. Translational Issues in Psychological Science 1, 2 (2015), 170–181. URL: https://doi.org/10.1037/tps0000037, doi:10.1037/tps0000037. 18

[FAZ21] FAHIM G., AMIN K., ZARIF S.: Single-View 3D reconstruction: A Survey of deep learning methods. Computers and Graphics 94 (2021), 164–190. doi:https://doi.org/10.1016/j.cag.2020.12.004. 2, 16

[FBCW23] FAN J. E., BAINBRIDGE W. A., CHAMBERLAIN R., WAMMES J. D.: Drawing as a versatile cognitive tool. Nature Reviews Psychology 2, 9 (September 2023), 556–568. URL: https://doi.org/10.1038/s44159-023-00212-w, doi:10.1038/s44159-023-00212-w. 18

[Feh04] FEHN C.: Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV. In Stereoscopic Displays and Virtual Reality Systems XI (2004), Bolas M. T., Woods A. J., Merritt J. O., Benton S. A., (Eds.), vol. 5291, International Society for Optics and Photonics, SPIE, pp. 93 – 104. URL: https://doi.org/10.1117/12.524762, doi:10.1117/12.524762. 6

[FHWG20] FAN J. E., HAWKINS R. D., WU M., GOODMAN N. D.: Pragmatic Inference and Visual Abstraction Enable Contextual Flexibility During Visual Communication. Computational Brain & Behavior 3, 1 (March 2020), 86–101. URL: https://doi.org/10.1007/s42113-019-00058-7, doi:10.1007/s42113-019-00058-7. 6, 18

[FQS*24] FUKUSHIMA Y., QI A., SHEN I.-C., GRYADITSKAYA Y., IGARASHI T.: 3D Reconstruction from Sketch with Hidden Lines by Two-Branch Diffusion Model. In Eurographics 2024 - Short Papers (2024), Hu R., Charalambous P., (Eds.), The Eurographics Association. doi:10.2312/egs.20241032. 5, 9, 10, 11, 14, 15, 32

[FRH*21] FONDEVILLA A., ROHMER D., HAHMANN S., BOUSSEAU A., CANI M.-P.: Fashion Transfer: Dressing 3D Characters from Stylized Fashion Sketches. Computer Graphics Forum 40, 6 (2021), 466–483. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.14390, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.14390, doi:https://doi.org/10.1111/cgf.14390. 3

[GLC*23] GAO L., LIU F.-L., CHEN S.-Y., JIANG K., LI C.-P., LAI Y.-K., FU H.: SketchFaceNeRF: Sketch-based Facial Generation and Editing in Neural Radiance Fields. ACM Trans. Graph. 42, 4 (jul 2023). URL: https://doi.org/10.1145/3592100, doi:10.1145/3592100. 3

[GRYF21] GUILLARD B., REMELLI E., YVERNAY P., FUA P.: Sketch2Mesh: Reconstructing and Editing 3D Shapes From Sketches. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2021), pp.13023–13032. doi:10.1109/ICCV48922.2021.01278. 3, 5, 8, 9, 12, 15, 16, 18, 19, 32, 33

[GSH*19] GRYADITSKAYA Y., SYPESTEYN M., HOFTIJZER J. W., PONT S., DURAND F., BOUSSEAU A.: OpenSketch: A Richly-Annotated Dataset of Product Design Sketches. ACM Transactions on Graphics (SIGGRAPH Asia) 38 (2019). doi:10.1145/3355089.3356533. 7, 8, 33

[GSL*25] GU S., SONG H., LIU B., YU Q., ZHANG S., JIANG H., HUANG J., TIAN F.: VRsketch2Gaussian: 3D VR Sketch Guided 3D Object Generation with Gaussian Splatting, 2025. 2

[GTC*24] GUO M., TANG M., CHA H., ZHANG R., LIU C. K., WU J.: ShapeCraft: Body-Aware and Semantics-Aware 3D Object Design, 2024. arXiv:2412.03889. 14, 16

[GTC*25] GUO M., TANG M., CHA H., ZHANG R., LIU C. K., WU J.: CRAFT: Designing Creative and Functional 3D Objects. 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) (2025), 7215–7224. doi:10.1109/WACV61041.2025.00701. 3

[GWY*24] GAO C., WANG X., YU Q., SHENG L., ZHANG J., HAN X., SONG Y.-Z., XU D.: 3D Reconstruction from a Single Sketch via View-dependent Depth Sampling. IEEE Transactions on Pattern Analysis and Machine Intelligence (2024), 1–16. doi:10.1109/TPAMI.2024.3424404. 6

[GYS*22] GAO C., YU Q., SHENG L., SONG Y.-Z., XU D.: SketchSampler: Sketch-Based 3D Reconstruction via View-Dependent Depth Sampling. In Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part I (Berlin, Heidelberg, 2022), Springer-Verlag, p. 464–479. URL: https://doi.org/10.1007/978-3-031-19769-7_27, doi:10.1007/978-3-031-19769-7_27. 5, 9, 15, 32

[GYZ23] GAO R., YUAN W., ZHU J.-Y.: Controllable Visual-Tactile Synthesis. Proceedings of the IEEE International Conference on Computer Vision (2023), 7017 – 7029. doi:10.1109/ICCV51070.2023.00648. 3

[GZW*25] GUO J., ZHANG J., WU F., LU H., WANG Q., YANG W., LIM E. G., LU D.: HiGarment: Cross-modal Harmony Based Diffusion Model for Flat Sketch to Realistic Garment Image, 2025. 3

[HCX*21] HE K., CHEN X., XIE S., LI Y., DOLLÁR P., GIRSHICK R.: Masked Autoencoders Are Scalable Vision Learners, 2021. URL: https://arxiv.org/abs/2111.06377, arXiv:2111.06377. 11

[HE17a] HA D., ECK D.: A Neural Representation of Sketch Drawings. CoRR abs/1704.03477 (2017). URL: http://arxiv.org/abs/1704.03477, arXiv:1704.03477. 31

[HE17b] HA D., ECK D.: A Neural Representation of Sketch Drawings . International Conference on Learning Representations (ICLR) (2017). URL: https://openreview.net/forum?id=Hy6GHpkCW, doi:1704.03477. 2, 3, 6, 7, 11, 14

[HGB19] HÄHNLEIN F., GRYADITSKAYA Y., BOUSSEAU A.: Bitmap or Vector? A study on sketch representations for deep stroke segmentation. In Journées Francaises d'Informatique Graphique et de Réalité virtuelle (Marseille, France, November 2019). URL: https://inria.hal.science/hal-02922043. 7

[HGSB22] HÄHNLEIN F., GRYADITSKAYA Y., SHEFFER A., BOUSSEAU A.: Symmetry-driven 3D Reconstruction from Concept Sketches. In ACM SIGGRAPH 2022 Conference Proceedings (New York, NY, USA, 2022), SIGGRAPH '22, Association for Computing Machinery. URL: https://doi.org/10.1145/3528233.3530723, doi:10.1145/3528233.3530723. 3

[HGY17] HAN X., GAO C., YU Y.: DeepSketch2Face: a deep learning based sketching system for 3D face and caricature modeling. ACM Transactions on Graphics 36, 4 (July 2017), 1–12. URL: http://dx.doi.org/10.1145/3072959.3073629, doi:10.1145/3072959.3073629. 3, 7

[HHL*23] HU J., HUI K.-H., LIU Z., ZHANG H., FU C.-W.: CLIPXPlore: Coupled CLIP and Shape Spaces for 3D Shape Exploration. In SIGGRAPH Asia 2023 Conference Papers (New York, NY, USA, 2023), SA '23, Association for Computing Machinery. URL: https://doi.org/10.1145/3610548.3618144, doi:10.1145/3610548.3618144. 5, 9, 15, 16, 17, 32

[HJA20] HO J., JAIN A., ABBEEL P.: Denoising diffusion probabilistic models. In Proceedings of the 34th International Conference on Neural Information Processing Systems (Red Hook, NY, USA, 2020), NIPS '20, Curran Associates Inc. 10

[HKYM17] HUANG H., KALOGERAKIS E., YUMER E., MECH R.: Shape Synthesis from Sketches via Procedural Models and Convolutional Networks. IEEE Transactions on Visualization and Computer Graphics 23, 8 (August 2017), 2003–2013. URL: https://doi.org/10.1109/TVCG.2016.2597830, doi:10.1109/TVCG.2016.2597830. 8

[HLHF24] HUI K.-H., LI R., HU J., FU C.-W.: Neural Wavelet-domain Diffusion for 3D Shape Generation. ACM Trans. Graph. 43, 2 (Jan. 2024). URL: https://doi.org/10.1145/3635304, doi:10.1145/3635304. 3

[HLW*17] HENNESSEY J. W., LIU H., WINNEMÖLLER H., DONTCHEVA M., MITRA N. J.: How2Sketch: generating easy-to-follow tutorials for sketching 3D objects. In Proceedings of the 21st ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (New York, NY, USA, 2017), I3D '17, Association for Computing Machinery. URL: https://doi.org/10.1145/3023368.3023371, doi:10.1145/3023368.3023371. 2, 7

[HPG*22] HERTZ A., PEREL O., GIRYES R., SORKINE-HORNUNG O., COHEN-OR D.: SPAGHETTI: Editing Implicit Shapes Through Part Aware Generation. ACM Trans. Graph. 41, 4 (July 2022). URL: https://doi.org/10.1145/3528223.3530084, doi:10.1145/3528223.3530084. 4, 9, 11, 17, 18

[HS22] HO J., SALIMANS T.: Classifier-free diffusion guidance. arXiv preprint arXiv:2207.12598 (2022). 10

[HZZ*24] HUANG T., ZENG Y., ZHANG Z., XU W., XU H., XU S., LAU R. W. H., ZUO W.: DreamControl: Control-Based Text-to-3D Generation with 3D Self-Prior . In 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (Los Alamitos, CA, USA, June 2024), IEEE Computer Society, pp. 5364–5373. URL: https://doi.ieeecomputersociety.org/10.1109/CVPR52733.2024.00513, doi:10.1109/CVPR52733.2024.00513. 6

[IIC*13] ISENBERG T., ISENBERG P., CHEN J., SEDLMAIR M., MÖLLER T.: A Systematic Review on the Practice of Evaluating Visualization. IEEE Transactions on Visualization and Computer Graphics 19, 12 (Dec 2013), 2818–2827. doi:10.1109/TVCG.2013.126. 3

[IKSS*23] INOUE N., KIKUCHI K., SIMO-SERRA E., OTANI M., YAMAGUCHI K.: LayoutDM: Discrete Diffusion Model for Controllable Layout Generation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2023), pp. 10167–10176. 19

[IMT99] IGARASHI T., MATSUOKA S., TANAKA H.: Teddy: A Sketching Interface for 3D Freeform Design. Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques (1999), pp.409–416. doi:10.1145/1281500.1281532. 2, 3

[IZZE17] ISOLA P., ZHU J., ZHOU T., EFROS A. A.: Image-to-Image Translation with Conditional Adversarial Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017), 1125–1134. doi:10.1109/CVPR.2017.632. 9

[JDA07] JUDD T., DURAND F., ADELSON E.: Apparent ridges for line drawing. ACM Trans. Graph. 26, 3 (jul 2007), 19–es. URL: https://doi.org/10.1145/1276377.1276401, doi:10.1145/1276377.1276401. 7

[JFD20] JIN A., FU Q., DENG Z.: Contour-based 3D Modeling through Joint Embedding of Shapes and Contours. In Symposium on Interactive 3D Graphics and Games (New York, NY, USA, 2020), I3D '20, Association for Computing Machinery. URL: https://doi.org/10.1145/3384382.3384518, doi:10.1145/3384382.3384518. 2, 5, 9, 15, 32

[JMB*22] JAIN A., MILDENHALL B., BARRON J. T., ABBEEL P., POOLE B.: Zero-Shot Text-Guided Object Generation With Dream Fields. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022), pp. pp.867–876. doi:10.1109/CVPR52688.2022.00094. 3, 16

[JMM*24] JIGNASU A., MARSHALL K. O., MISHRA A. K., RILLO L. N., GANAPATHYSUBRAMANIAN B., BALU A., HEGDE C., KRISHNAMURTHY A.: Slice-100K: A Multimodal Dataset for Extrusion-based 3D Printing. Advances in Neural Information Processing Systems 37 (2024). 7

[JSL*19] JATAVALLABHULA K. M., SMITH E., LAFLECHE J.-F., TSANG C. F., ROZANTSEV A., CHEN W., XIANG T., LEBARedian R., FIDLER S.: Kaolin: A PyTorch Library for Accelerating 3D Deep Learning Research. 12

[JSR*22] JAKOB W., SPEIERER S., ROUSSEL N., NIMIER-DAVID M., VICINI D., ZELTNER T., NICOLET B., CRESPO M., LEROY V., ZHANG Z.: Mitsuba 3 renderer, 2022. https://mitsuba-renderer.org. 12

[KBH06] KAZHDAN M., BOLITHO M., HOPPE H.: Poisson Surface Reconstruction. In Proceedings of the Fourth Eurographics Symposium on

Geometry Processing (Goslar, DEU, 2006), SGP '06, Eurographics Association, p. pp.61–70. 9

[KBM*20] KATO H., BEKER D., MORARIU M., ANDO T., MATSUOKA T., KEHL W., GAIDON A.: Differentiable Rendering: A Survey, 2020. URL: https://arxiv.org/abs/2006.12057, arXiv:2006.12057. 12

[KBS*23] KOLEY S., BHUNIA A. K., SAIN A., CHOWDHURY P. N., XIANG T., SONG Y.-Z.: Picture that sketch: Photorealistic image generation from abstract sketches. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2023), 6850–6861. 3

[KCH*20] KIM S., CHI H.-G., HU X., HUANG Q., RAMANI K.: A Large-scale Annotated Mechanical Components Benchmark for Classification and Retrieval Tasks with Deep Neural Networks. Proceedings of the European Conference on Computer Vision (ECCV) (2020). doi:10.1007/978-3-030-58523-5. 7

[KGC*17] KAZI R. H., GROSSMAN T., CHEONG H., HASHEMI A., FITZMAURICE G.: DreamSketch: Early Stage 3D Design Explorations with Sketching and Generative Design. In Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (New York, NY, USA, 2017), UIST '17, Association for Computing Machinery, p. 401–414. URL: https://doi.org/10.1145/3126594.3126662, doi:10.1145/3126594.3126662. 3

[KH06] KARPENKO O. A., HUGHES J. F.: SmoothSketch: 3D free-form shapes from complex sketches. ACM Trans. Graph. 25, 3 (jul 2006), 589–598. URL: https://doi.org/10.1145/1141911.1141928, doi:10.1145/1141911.1141928. 3

[KHA*22] KOO J., HUANG I., ACHLIOPTAS P., GUIBAS L. J., SUNG M.: PartGlot: Learning Shape Part Segmentation From Language Reference Games. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022), pp.16505–16514. doi:10.1109/CVPR52688.2022.01601. 14

[KHR02] KARPENKO O., HUGHES J. F., RASKAR R.: Free-form sketching with variational implicit surfaces. Computer Graphics Forum 21, 3 (2002), 585–594. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-8659.t01-1-00709, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-8659.t01-1-00709, doi:https://doi.org/10.1111/1467-8659.t01-1-00709. 3

[KHW*22] KIM B., HUANG X., WUELFROTH L., TANG J., CORDONNIER G., GROSS M., SOLENTHALER B.: Deep Reconstruction of 3D Smoke Densities from Artist Sketches. Computer Graphics Forum 41, 2 (2022), 97–110. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.14461, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.14461, doi:https://doi.org/10.1111/cgf.14461. 3

[KHY*25] KOSMYNA N., HAUPTMANN E., YUAN Y. T., SITU J., LIAO X.-H., BERESNITZKY A. V., BRAUNSTEIN I., MAES P.: Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task, 2025. URL: https://arxiv.org/abs/2506.08872, arXiv:2506.08872. 18

[KKLD23] KERBL B., KOPANAS G., LEIMKUEHLER T., DRETTAKIS G.: 3D Gaussian Splatting for Real-Time Radiance Field Rendering. ACM Transactions on Graphics 42, 4 (2023). doi:10.1145/3592433. 13

[KSPH21] KINGMA D. P., SALIMANS T., POOLE B., HO J.: Variational Diffusion Models. Advances in Neural Information Processing Systems (NeurIPS) 34 (2021), 21696–21707. 10

[KWQ23] KONG D., WANG Q., QI Y.: A Diffusion-ReFinement Model for Sketch-to-Point Modeling. In Computer Vision – ACCV 2022: 16th Asian Conference on Computer Vision, Macao, China, December 4–8, 2022, Proceedings, Part VII (Berlin, Heidelberg, 2023), Springer-Verlag, p. 54–70. URL: https://doi.org/10.1007/978-3-031-26293-7_4, doi:10.1007/978-3-031-26293-7_4. 5, 9, 15, 16, 32

[KYNS23] KOO J., YOO S., NGUYEN M. H., SUNG M.: SALAD: Part-Level Latent Diffusion for 3D Shape Generation and Manipulation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (October 2023), pp. 14441–14451. 18

[LB25] LIU C., BESSMELTSEV M.: State-of-the-art Report in Sketch Processing. Computer Graphics Forum 44, 2 (2025), e70079. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.70079, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.70079, doi:https://doi.org/10.1111/cgf.70079. 3

[LCD*23] LUO Z., CAI S., DONG J., MING R., QIU L., ZHAN X., HAN X.: RaBit: Parametric Modeling of 3D Biped Cartoon Characters with a Topological-consistent Dataset, 2023. URL: https://arxiv.org/abs/2303.12564, arXiv:2303.12564. 3

[LCD*25] LU S., CHEN G., DINH N. A., LANG I., HOLTZMAN A., HANOCKA R.: LL3M: Large Language 3D Modelers, 2025. URL: https://arxiv.org/abs/2508.08228, arXiv:2508.08228. 14

[LCX*23] LUO L., CHOWDHURY P. N., XIANG T., SONG Y.-Z., GRYADITSKAYA Y.: 3D VR Sketch Guided 3D Shape Prototyping and Exploration. In Proceedings of the IEEE/CVF International Conference on Computer Vision (2023), pp. 9267–9276. 2, 3

[LCZL25] LI Z., CHEN Y., ZHAO L., LIU P.: Controllable Text-to-3D Generation via Surface-Aligned Gaussian Splatting. Proceedings - 2025 International Conference on 3D Vision, 3DV 2025 (2025), 1113 – 1123. doi:10.1109/3DV66043.2025.00107. 5, 9, 13, 15, 17, 32

[LDT*22] LI X., DING H., TONG Z., WU Y., CHEE Y. M.: Primitive3D: 3D Object Dataset Synthesis From Randomly Assembled Primitives. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022), pp. pp.15947–15957. doi:10.1109/CVPR52688.2022.01548. 14

[LDZ*24] LUO Z., DU D., ZHU H., YU Y., FU H., HAN X.: SketchMetaFace: A Learning-based Sketching Interface for High-fidelity 3D Character Face Modeling. IEEE Transactions on Visualization and Computer Graphics 30, 8 (2024), 5260 – 5275. doi:10.1109/TVCG.2023.3291703. 3

[LEL*25] LI H., ERKOC Z., LI L., SIRIGATTI D., ROZOV V., DAI A., NIESSNER M.: MeshPad: Interactive Sketch-Conditioned Artist-designed Mesh Generation and Editing. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2025). 2, 19

[LFLG24] LIU F.-L., FU H., LAI Y.-K., GAO L.: SketchDream: Sketch-based Text-to-3D Generation and Editing. ACM Transactions on Graphics 43, 4 (2024). URL: https://arxiv.org/abs/2405.06461, doi:10.1145/3658120. 2, 5, 6, 7, 9, 13, 14, 15, 16, 17, 18, 31, 32

[LGK*17] LUN Z., GADELHA M., KALOGERAKIS E., MAJI S., WANG R.: 3D Shape Reconstruction from Sketches via Multi-view Convolutional Networks. CoRR abs/1707.06375 (2017). doi:10.1109/3DV.2017.00018. 5, 6, 9, 15, 32

[LHG*23] LEE H., HWANG I., GO H., CHOI W.-S., KIM K., ZHANG B.-T.: Learning Geometry-aware Representations by Sketching. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2023), pp. 23315–23326. 7

[LHH*24] LIU J., HUANG X., HUANG T., CHEN L., HOU Y., TANG S., LIU Z., OUYANG W., ZUO W., JIANG J., LIU X.: A Comprehensive Survey on 3D Content Generation, 2024. URL: https://arxiv.org/abs/2402.01166, arXiv:2402.01166. 3, 12

[LHHW25] LIU F., HUANG L., HUANG Z., WANG Z.: Learning to Draw Is Learning to See: Analyzing Eye Tracking Patterns for Assisted Observational Drawing. In Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers (New York, NY, USA, 2025), SIGGRAPH Conference Papers '25, Association for Computing Machinery. URL: https://doi.org/10.1145/3721238.3730734, doi:10.1145/3721238.3730734. 18

[LJJ*24] LEE S. W., JO T. H., JIN S., CHOI J., YUN K., BROMBERG S., BAN S., HYUN K. H.: The Impact of Sketch-guided vs. Prompt-guided 3D Generative AIs on the Design Exploration Process. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (New York, NY, USA, 2024), CHI '24, Association for Computing Machinery. URL: https://doi.org/10.1145/3613904.3642218, doi:10.1145/3613904.3642218. 3, 9

[LJK*25] LEE S., JUNG H., KOH B., HUANG Q., YOON S., KIM S.: PASTA: Part-Aware Sketch-to-3D Shape Generation with Text-Aligned Prior. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (October 2025), 18585–18595. URL: https://arxiv.org/abs/2503.12834. 18

[LLM*19] LI M., LIN Z., MĚCH R., YUMER E., RAMANAN D.: Photosketching: Inferring contour drawings from images. Proceedings - 2019 IEEE Winter Conference on Applications of Computer Vision, WACV 2019 (2019), 1403 – 1412. doi:10.1109/WACV.2019.00154. 13

[LM95a] LANDAY J. A., MYERS B. A.: Interactive sketching for the early stages of user interface design. In Proceedings of the CHI '95 Conference on Human Factors in Computing Systems (Denver, CO, 1995), ACM Press/Addison-Wesley, pp. 43–50. doi:10.1145/223904.223910. 2

[LM95b] LANDAY J. A., MYERS B. A.: Just draw it! Programming by sketching storyboards. Tech. Rep. CMU-CS-95-199, Carnegie Mellon University, School of Computer Science, November 1995. 2

[LM01] LANDAY J., MYERS B.: Sketching interfaces: toward more human interface design. Computer 34, 3 (2001), 56–64. doi:10.1109/2.910894. 2

[LMG23] LIANG X., MO H., GAO C.: Controllable Garment Image Synthesis Integrated with Frequency Domain Features. Computer Graphics Forum 42, 7 (2023), e14938. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.14938, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.14938, doi:https://doi.org/10.1111/cgf.14938. 3

[LOM*18] LESCOAT T., OVSJANIKOV M., MEMARI P., THIERY J.-M., BOUBEKEUR T.: A Survey on Data-driven Dictionary-based Methods for 3D Modeling. Computer Graphics Forum 37, 2 (2018), 577–601. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13384, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.13384, doi:https://doi.org/10.1111/cgf.13384. 2

[LPBM20] LI C., PAN H., BOUSSEAU A., MITRA N. J.: Sketch2CAD: Sequential CAD Modeling by Sketching in Context. ACM Transactions on Graphics (SIGGRAPH Asia) 39, 6 (2020), pp.1–14. doi:https://doi.org/10.1145/3414685.3417807. 5, 7, 8, 9, 11, 15, 17, 18, 19, 32

[LPBM22] LI C., PAN H., BOUSSEAU A., MITRA N. J.: Free2CAD: Parsing Freehand Drawings into CAD Commands. ACM Transactions on Graphics (SIGGRAPH) 41, 4 (2022), pp.1–16. doi:https://doi.org/10.1145/3528223.3530133. 2, 5, 6, 7, 8, 9, 11, 13, 14, 15, 16, 19, 32

[LPG24] LI S., PASCHALIDOU D., GUIBAS L.: PASTA: Controllable Part-Aware Shape Generation with Autoregressive Transformers. 14

[LPH*24] LIAO Z., PIAO F., HUANG D., LI X., MA Y., FENG P., FANG H., ZENG L.: Freehand Sketch Generation from Mechanical Components. In ACM Multimedia 2024 (2024). URL: https://openreview.net/forum?id=bDjEC9P0sl. 7

[LPL*17] LI C., PAN H., LIU Y., TONG X., SHEFFER A., WANG W.: BendSketch: modeling freeform surfaces through 2D sketching. ACM Trans. Graph. 36, 4 (jul 2017). URL: https://doi.org/10.1145/3072959.3073632, doi:10.1145/3072959.3073632. 3

[LPL*18] LI C., PAN H., LIU Y., TONG X., SHEFFER A., WANG W.: Robust Flow-Guided Neural Prediction for Sketch-Based Freeform Surface Modeling. ACM Transactions on Graphics (SIGGRAPH 37, 6 (2018). doi:10.1145/3272127.3275051. 3

[LSC24] LEE H.-H., SAVVA M., CHANG A. X.: Text-to-3D Shape Generation. Computer Graphics Forum 43, 2 (2024), e15061. doi:10.1111/cgf.15061. 2, 3, 12

[LWL*22] LING J., WANG Z., LU M., WANG Q., QIAN C., XU F.: Structure-aware Editable Morphable Model for 3D Facial Detail Animation and Manipulation. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 13663 LNCS (2022), 249 – 267. URL: https://arxiv.org/abs/2207.09019, doi:10.1007/978-3-031-20062-5_15. 3

[LWLQ22] LIN T., WANG Y., LIU X., QIU X.: A survey of transformers. AI Open 3 (2022), 111–132. URL: https://www.sciencedirect.com/science/article/pii/S2666651022000146, doi:https://doi.org/10.1016/j.aiopen.2022.10.001. 11

[LYL*23] LIANG Y., YANG X., LIN J., LI H., XU X., CHEN Y.: LucidDreamer: Towards High-Fidelity Text-to-3D Generation via Interval Score Matching. IEEE Transactions on Visualization and Computer Graphics 31, 12 (2023), 10640 – 10651. doi:10.1109/TVCG.2025.3611489. 13

[LZC*24] LI C., ZHANG C., CHO J., WAGHWASE A., LEE L.-H., RAMEAU F., YANG Y., BAE S.-H., HONG C. S.: Generative AI meets 3D: A Survey on Text-to-3D in AIGC Era, 2024. URL: https://arxiv.org/abs/2305.06131, arXiv:2305.06131. 2, 3

[LZK*24] LI X., ZHANG Q., KANG D., CHENG W., GAO Y., ZHANG J., LIANG Z., LIAO J., CAO Y.-P., SHAN Y.: Advances in 3D Generation: A Survey, 2024. URL: https://arxiv.org/abs/2401.17807, arXiv:2401.17807. 3

[LZZ*21] LUO Z., ZHOU J., ZHU H., DU D., HAN X., FU H.: SimpModeling: Sketching Implicit Field to Guide Mesh Modeling for 3D Animalmorphic Head Design. In The 34th Annual ACM Symposium on User Interface Software and Technology (New York, NY, USA, 2021), UIST '21, Association for Computing Machinery, p. 854–863. URL: https://doi.org/10.1145/3472749.3474791, doi:10.1145/3472749.3474791. 3

[MB21] MICHEL E., BOUBEKEUR T.: DAG amendment for inverse control of parametric shapes. ACM Trans. Graph. 40, 4 (July 2021). URL: https://doi.org/10.1145/3450626.3459823, doi:10.1145/3450626.3459823. 18

[MCEG23] MANFREDI G., CAPECE N., ERRA U., GRUOSSO M.: TreeSketchNet: From Sketch to 3D Tree Parameters Generation. ACM Transactions on Intelligent Systems and Technology 14, 3 (2023). doi:10.1145/3579831. 3

[ML07] MASRY M., LIPSON H.: A sketch-based interface for iterative design and analysis of 3D objects. In ACM SIGGRAPH 2007 Courses (New York, NY, USA, 2007), SIGGRAPH '07, Association for Computing Machinery, p. 31–es. URL: https://doi.org/10.1145/1281500.1281542, doi:10.1145/1281500.1281542. 3

[MM19] MIHAYLOVA T., MARTINS A. F. T.: Scheduled Sampling for Transformers. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL) (2019), 351–356. doi:10.18653/v1/P19-2049. 11

[MNAA25] MAN B., NEHME G., ALAM M. F., AHMED F.: VideoCAD: A Dataset and Model for Learning Long-Horizon 3D CAD UI Interactions from Video. IEEE Transactions on Instrumentation and Measurement 72 (2025). URL: https://arxiv.org/abs/2505.24838, doi:10.1109/TIM.2022.3229704. 7

[MON*19] MESCHEDER L., OECHSLE M., NIEMEYER M., NOWOZIN S., GEIGER A.: Occupancy Networks: Learning 3D Reconstruction in Function Space. Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019). doi:10.1109/CVPR.2019.00459. 2, 7, 9, 10

[MPA24] MUKTI M., PAWLING R., ANDREWS D.: Computer aided sketching in the early-stage design of complex vessels. Ocean Engineering 305 (2024), 117407. URL: https://www.sciencedirect.com/science/article/pii/S0029801824007443, doi:https://doi.org/10.1016/j.oceaneng.2024.117407. 3

[MPS*23] MIKAEILI A., PEREL O., SAFAEE M., COHEN-OR D., MAHDAVI-AMIRI A.: SKED: Sketch-guided Text-based 3D Editing. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (October 2023), pp. 14607–14619. 5, 9, 12, 13, 15, 16, 17, 32

[MVDD24] MITROPOULOU I., VAXMAN A., DIAMANTI O., DILLENBURGER B.: Fabrication-aware strip-decomposable quadrilateral meshes. Comput. Aided Des. 168, C (March 2024). URL: https://doi.org/10.1016/j.cad.2023.103666, doi:10.1016/j.cad.2023.103666. 14

[MWH*23] MÜLLER P., WONKA P., HAEGLER S., ULMER A., VAN GOOL L.: Procedural Modeling of Buildings, 1 ed. Association for Computing Machinery, New York, NY, USA, 2023. URL: https://doi.org/10.1145/3596711.3596738. 18

[MWLZ22] MA J., WANG J., LI J., ZHANG D.: Real-time Skeletonization for Sketch-based Modeling. Computers and Graphics 102 (2022), 56–66. doi:https://doi.org/10.1016/j.cag.2021.11.005. 3

[MZC*19] MO K., ZHU S., CHANG A. X., YI L., TRIPATHI S., GUIBAS L. J., SU H.: PartNet: A Large-Scale Benchmark for Fine-Grained and Hierarchical Part-Level 3D Object Understanding. Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019). doi:10.1109/CVPR.2019.00100. 14

[NBA18] NISHIDA G., BOUSSEAU A., ALIAGA D.: Procedural Modeling of a Building from a Single Image. Computer Graphics Forum 37 (2018), pp.415–429. doi:10.1111/cgf.13372. 8

[NBS*24] NGHIEM C. M.-G., BOUSSEAU A., SYPESTEYN M., HOFTIJZER J., TSANDILAS T.: Sketch Presentation for Product Design. In Proceedings of the 35th International Francophone Conference on Human-Computer Interaction (New York, NY, USA, 2024), IHM '24, Association for Computing Machinery. URL: https://doi.org/10.1145/3650104.3652907, doi:10.1145/3650104.3652907. 3

[NGDA*16] NISHIDA G., GARCIA-DORADO I., ALIAGA D. G., BENES B., BOUSSEAU A.: Interactive sketching of urban procedural models. ACM Transactions on Graphics 35, 4 (2016). doi:10.1145/2897824.2925951. 5, 8, 9, 15, 16, 31, 32, 33

[NISA07] NEALEN A., IGARASHI T., SORKINE O., ALEXA M.: FiberMesh: designing freeform surfaces with 3D curves. ACM Trans. Graph. 26, 3 (July 2007), 41–es. URL: https://doi.org/10.1145/1276377.1276429, doi:10.1145/1276377.1276429. 3

[NJC*22] NAYA F., JORGE J. A., CONESA J., CONTERA M., GOMIS J. M.: Direct Modelíng: from Sketches to 3D Models. In SIACG2002 - 1st Ibero-American Symposium in Computer Graphics (2022), Pueyo X., dos Santos M. P., Velho L., (Eds.), The Eurographics Association. doi:10.2312/pt.20021409. 2

[NKR*22] NAM G., KHLIFI M., RODRIGUEZ A., TONO A., ZHOU L., GUERRERO P.: 3D-LDM: Neural Implicit 3D Shape Generation with Latent Diffusion Models. arXiv pre-print (2022). arXiv:2212.00842. 3, 17

[NSF*22] NOZAWA N., SHUM H. P. H., FENG Q., HO E. S. L., MORISHIMA S.: 3D car shape reconstruction from a contour sketch using GAN and lazy learning. Vis. Comput. 38, 4 (April 2022), 1317–1330. URL: https://doi.org/10.1007/s00371-020-02024-y, doi:10.1007/s00371-020-02024-y. 3

[OCK*24] OH Y., CHOI J., KIM Y., PARK M., SHIN C., YOON S.: ControlDreamer: Blending Geometry and Style in Text-to-3D. 35th British Machine Vision Conference 2024, BMVC 2024, Glasgow, UK, November 25–28, 2024 (2024). 7, 12, 13

[OCM*23] OLIVIER P., CHABRIER R., MEMARI P., COLL J.-L., CANI M.-P.: Bio-Sketch: A new medium for interactive storytelling illustrated by the phenomenon of infection. In VCBM 2023 - 13th Eurographics Workshop on Visual Computing for Biology and Medicine (Norrköping, Sweden, September 2023), The Eurographics Association, p. 11. URL: https://hal.science/hal-04216965. 3

[OK14] OLSON J. S., KELLOGG W. A.: Ways of Knowing in HCI. Springer Publishing Company, Incorporated, 2014. 17

[OSCSJ09] OLSEN L., SAMAVATI F., COSTA SOUSA M., JORGE J.: Sketch-based modeling: A survey. Computers and Graphics 33 (2009), pp.85–103. 2, 3, 19

[OVK*19] OZTIRELI C., VALENTIN J., KESKIN C., PIDLYPENSKYI P., MAKADIA A., SUD A., BOUAZIZ S.: TensorFlow Graphics: Computer Graphics Meets Deep Learning, 2019. 12

[PJBM23] POOLE B., JAIN A., BARRON J. T., MILDENHALL B.: Dreamfusion: Text-to-3d using 2d diffusion. The Eleventh International Conference on Learning Representations (2023). URL: https://openreview.net/forum?id=FjNys5c7VyY. 12, 13, 16

[PKM*11] PACZKOWSKI P., KIM M. H., MORVAN Y., DORSEY J., RUSHMEIER H., O'SULLIVAN C.: Insitu: sketching architectural designs in context. ACM Trans. Graph. 30, 6 (December 2011), 1–10. URL: https://doi.org/10.1145/2070781.2024216, doi:10.1145/2070781.2024216. 19

[PLH*22] PEARL O., LANG I., HU Y., YEH R. A., HANOCKA R.: GeoCode: Interpretable Shape Programs. Computer Graphics Forum 44, 1 (2022). doi:10.1111/cgf.15276. 5, 7, 8, 9, 14, 15, 16, 17, 18, 19, 32

[PMKB23] PUHACHOV I., MARTENS C., KRY P. G., BESSMELTSEV M.: Reconstruction of Machine-Made Shapes from Bitmap Sketches. ACM Trans. Graph. 42, 6 (dec 2023). URL: https://doi.org/10.1145/3618361, doi:10.1145/3618361. 3

[Pra04] PRANOVICH S.: Structural sketcher:a tool for supporting architects in early design. Automation in Construction - AUTOM CONSTR (01 2004). 3

[PUG19a] PASCHALIDOU D., ULUSOY A. O., GEIGER A.: Superquadrics Revisited: Learning 3D Shape Parsing beyond Cuboids. Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019). 14

[PUG19b] PASCHALIDOU D., ULUSOY A. O., GEIGER A.: Superquadrics Revisited: Learning 3D Shape Parsing beyond Cuboids. In Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (June 2019). 14

[PYG*24] PO R., YIFAN W., GOLYANIK V., ABERMAN K., BARRON J. T., BERMANO A., CHAN E., DEKEL T., HOLYNSKI A., KANAZAWA A., LIU C., LIU L., MILDENHALL B., NIESSNER M., OMMER B., THEOBALT C., WONKA P., WETZSTEIN G.: State of the Art on Diffusion Models for Visual Computing. Computer Graphics Forum 43, 2 (2024), e15063. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.15063, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.15063, doi:https://doi.org/10.1111/cgf.15063. 10, 12

[QGS*21] QI A., GRYADITSKAYA Y., SONG J., YANG Y., QI Y., HOSPEDALES T. M., XIANG T., SONG Y.-Z.: Toward Fine-Grained Sketch-Based 3D Shape Retrieval. Trans. Img. Proc. 30 (January 2021), 8595–8606. URL: https://doi.org/10.1109/TIP.2021.3118975, doi:10.1109/TIP.2021.3118975. 4

[RCDB23] ROSSET N., CORDONNIER G., DUVIGNEAU R., BOUSSEAU A.: Interactive design of 2D car profiles with aerodynamic feedback. Computer Graphics Forum 42, 2 (2023), 427–437. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.14772, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.14772, doi:https://doi.org/10.1111/cgf.14772. 16

[RGJ*23] RITCHIE D., GUERRERO P., JONES R. K., MITRA N. J.,

SCHULZ A., WILLIS K. D. D., WU J.: Neurosymbolic Models for Computer Graphics, 2023. URL: https://arxiv.org/abs/2304.10320, arXiv:2304.10320. 8, 18

[RKH*21] RADFORD A., KIM J. W., HALLACY C., RAMESH A., GOH G., AGARWAL S., SASTRY G., ASKELL A., MISHKIN P., CLARK J., KRUEGER G., SUTSKEVER I.: Learning Transferable Visual Models From Natural Language Supervision. Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event 139 (2021), 8748–8763. 13, 16

[RLR*20] REMELLI E., LUKOIANOV A., RICHTER S. R., GUILLARD B., BAGAUTDINOV T., BAQUÉ P., FUA P.: Meshsdf: Differentiable iso-surface extraction. In Advances in Neural Information Processing Systems (NeurIPS) (2020), vol. 33, pp. 22468–22478. URL: https://proceedings.neurips.cc/paper/2020/file/xxxxxxx.pdf. 9

[RRN*20] RAVI N., REIZENSTEIN J., NOVOTNY D., GORDON T., LO W.-Y., JOHNSON J., GKIOXARI G.: Accelerating 3D Deep Learning with PyTorch3D, 2020. URL: https://arxiv.org/abs/2007.08501, arXiv:2007.08501. 12

[SAG*13] SHTOF A., AGATHOS A., GINGOLD Y., SHAMIR A., COHEN-OR D.: Geosemantic Snapping for Sketch-Based Modeling. Computer Graphics Forum 32, 2pt2 (2013), 245–253. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.12044, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.12044, doi:https://doi.org/10.1111/cgf.12044. 3

[SAG*24] SAHOO S. S., ARRIOLA M., GOKASLAN A., MARROQUIN E. M., RUSH A. M., SCHIFF Y., CHIU J. T., KULESHOV V.: Simple and Effective Masked Diffusion Language Models. In The Thirty-eighth Annual Conference on Neural Information Processing Systems (2024). URL: https://openreview.net/forum?id=L4uaAR4ArM. 14

[SAMS*21] STANISLAVA F., ALBERTO T., MEHER SHASHWAT N., JIAYAO Z., AMIRHOSSEIN A., CECILIA MARIA B., DOMINIK L. M.: Synthetic 3D Data Generation Pipeline for Geometric Deep Learning in Architecture. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLIII-B2-2021 (2021), pp.337–344. URL: https://arxiv.org/abs/2104.12564, arXiv:2104.12564, doi:10.5194/isprs-archives-XLIII-B2-2021-337-2021. 3

[SBS21] SMIRNOV D., BESSMELTSEV M., SOLOMON J.: Learning Manifold Patch-Based Representations of Man-Made Shapes. International Conference on Learning Representations (ICLR) (2021). 8, 12, 14, 16

[SDM*24] SHUAI X., DING H., MA X., TU R., JIANG Y.-G., TAO D.: A Survey of Multimodal-Guided Image Editing with Text-to-Image Diffusion Models, 2024. URL: https://arxiv.org/abs/2406.14555, arXiv:2406.14555. 4

[SF68] SOBEL I., FELDMAN G.: A 3x3 Isotropic Gradient Operator for Image Processing. In Presented at the Stanford Artificial Intelligence Project (SAIL) (1968). Technical Report. URL: https://www.researchgate.net/publication/239398818_A_3x3_Isotropic_Gradient_Operator_for_Image_Processing. 7

[SG00] SCHWEIKARDT E., GROSS M. D.: Digital clay: deriving digital models from freehand sketches. Automation in Construction 9, 1 (2000), 107–115. URL: https://www.sciencedirect.com/science/article/pii/S0926580599000527, doi:https://doi.org/10.1016/S0926-5805(99)00052-7. 2, 3

[SGL*18] SHARMA G., GOYAL R., LIU D., KALOGERAKIS E., MAJI S.: Csgnet: Neural shape parser for constructive solid geometry. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2018). URL: https://openaccess.thecvf.com/content_cvpr_2018/papers/Sharma_CSGNet_Neural_Shape_CVPR_2018_paper.pdf, doi:10.1109/CVPR.2018.00578. 8

[SHW*25] SHI J., HAN K., WANG Z., DOUCET A., TITSIAS M. K.: Simplified and generalized masked diffusion for discrete data. In Proceedings of the 38th International Conference on Neural Information Processing Systems (Red Hook, NY, USA, 2025), NIPS '24, Curran Associates Inc. 14, 19

[SJR*23] SANGHI A., JAYARAMAN P. K., RAMPINI A., LAMBOURNE J., SHAYANI H., ATHERTON E., TAGHANAKI S. A.: Sketch-A-Shape: Zero-Shot Sketch-to-3D Shape Generation, 2023. URL: https://arxiv.org/abs/2307.03869, arXiv:2307.03869. 2, 5, 9, 11, 15, 18, 19, 32

[SKR*24] SANGHI A., KHANI A., REDDY P., RAMPINI A., CHEUNG D., MALEKSHAN K. R., MADAN K., SHAYANI H.: Wavelet Latent Diffusion (Wala): Billion-Parameter 3D Generative Model with Compact Wavelet Encodings, 2024. URL: https://arxiv.org/abs/2411.08017, arXiv:2411.08017. 3, 14

[SLX*25] SUN Y., LI J., XU Z., ZHANG J., LIU X., ZHANG D., LU G.: Sketch2Seq: Reconstruct CAD models from Feature-based Sketch Segmentation. IEEE Transactions on Visualization and Computer Graphics (2025), 1–14. doi:10.1109/TVCG.2025.3566544. 3, 7

[SNL*21] SELVARAJU P., NABAIL M., LOIZOU M., MASLIOUKOVA M., AVERKIOU M., ANDREOU A., CHAUDHURI S., KALOGERAKIS E.: BuildingNet: Learning to Label 3D Buildings. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2021). doi:10.1109/ICCV48922.2021.01023. 7, 19

[Som20] SOMRAJ N.: Pose-Warping for View Synthesis / DIBR, 2020. URL: https://github.com/NagabhushanSN95/Pose-Warping. 6

[SPX*23] SHI Z., PENG S., XU Y., GEIGER A., LIAO Y., SHEN Y.: Deep Generative Models on 3D Representations: A Survey, 2023. URL: https://arxiv.org/abs/2210.15663, arXiv:2210.15663. 3

[SSC*24] SHEN Y., SHEN Y., CHENG J., JIANG C., FAN M., WANG Z.: Neural Canvas: Supporting Scenic Design Prototyping by Integrating 3D Sketching and Generative AI. Conference on Human Factors in Computing Systems - Proceedings (2024). doi:10.1145/3613904.3642096. 3

[SSG*23] SOMEPALLI G., SINGLA V., GOLDBLUM M., GEIPING J., GOLDSTEIN T.: Diffusion Art or Digital Forgery? Investigating Data Replication in Diffusion Models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2023), pp. 6048–6058. 19

[SSGII16] SIMO-SERRA E., GIUNCHI M., IIZUKA S., ISHIKAWA H.: Learning to simplify: Fully convolutional networks for rough sketch cleanup. ACM Transactions on Graphics (TOG) 35, 4 (2016), 1–11. 7

[ST90] SAITO T., TAKAHASHI T.: Comprehensible rendering of 3-D shapes. In Proceedings of the 17th Annual Conference on Computer Graphics and Interactive Techniques (New York, NY, USA, 1990), SIGGRAPH '90, Association for Computing Machinery, p. 197–206. URL: https://doi.org/10.1145/97879.97901, doi:10.1145/97879.97901. 7

[STD*21] SUN W., TAGLIASACCHI A., DENG B., SABOUR S., YAZDANI S., HINTON G. E., YI K. M.: Canonical Capsules: Self-Supervised Capsules in Canonical Pose. In Advances in Neural Information Processing Systems (NeurIPS 2021) (2021). URL: https://proceedings.neurips.cc/paper/2021/hash/d1ee59e20ad01cedc15f5118a7626099-Abstract.html. 7

[Sut63] SUTHERLAND I. E.: Sketchpad: a man-machine graphical communication system. In Proceedings of the May 21-23, 1963, Spring Joint Computer Conference (New York, NY, USA, 1963), AFIPS '63 (Spring), Association for Computing Machinery, p. 329–346. URL: https://doi.org/10.1145/1461551.1461591, doi:10.1145/1461551.1461591. 2, 3

[SWG24] SUN J.-M., WU T., GAO L.: Recent advances in implicit representation-based 3D shape generation. Visual Intelligence 2 (03 2024). doi:10.1007/s44267-024-00042-1. 12

[SWSJ06] SCHMIDT R., WYVILL B., SOUSA M. C., JORGE J. A.: ShapeShop: sketch-based solid modeling with BlobTrees. In ACM SIGGRAPH 2006 Courses (New York, NY, USA, 2006), SIGGRAPH '06, Association for Computing Machinery, p. 14–es. URL: https://doi.org/10.1145/1185657.1185775, doi:10.1145/1185657.1185775. 3

[SWY*24] SHI Y., WANG P., YE J., LONG M., LI K., YANG X.: MVDream: Multi-view Diffusion for 3D Generation, 2024. URL: https://arxiv.org/abs/2308.16512, arXiv:2308.16512. 12

[SZF*21] SHEN Y., ZHANG C., FU H., ZHOU K., ZHENG Y.: DeepSketchHair: Deep Sketch-Based 3D Hair Modeling. IEEE Transactions on Visualization and Computer Graphics 27, 7 (July 2021), 3250–3263. URL: http://dx.doi.org/10.1109/TVCG.2020.2968433, doi:10.1109/tvcg.2020.2968433. 3

[SZS*24] SUN J., ZHANG B., SHAO R., WANG L., LIU W., XIE Z., LIU Y.: DreamCraft3D: Hierarchical 3D Generation with Bootstrapped Diffusion Prior. In Proceedings of the International Conference on Learning Representations (ICLR) 2024 (2024). URL: https://arxiv.org/abs/2310.16818. 12

[SZZZ23] SHI J., ZHANG H., ZHOU D., ZHANG Z.: Toward Intelligent Interactive Design: A Generation Framework Based on Cross-domain Fashion Elements. In Proceedings of the 31st ACM International Conference on Multimedia (New York, NY, USA, 2023), MM '23, Association for Computing Machinery, p. 7152–7163. URL: https://doi.org/10.1145/3581783.3612376, doi:10.1145/3581783.3612376. 3

[THAF24] TONO A., HUANG H., AGRAWAL A., FISCHER M.: Vitruvio: Conditional variational autoencoder to generate building meshes via single perspective sketches. Automation in Construction 166 (2024), 105498. URL: https://www.sciencedirect.com/science/article/pii/S0926580524002346, doi:https://doi.org/10.1016/j.autcon.2024.105498. 2, 3, 5, 7, 8, 9, 10, 15, 16, 32, 33

[TSG*17] TULSIANI S., SU H., GUIBAS L. J., EFROS A. A., MALIK J.: Learning shape abstractions by assembling volumetric primitives. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (July 2017). 14

[TSNA*21] TONO A., SHASHWAT NIGAM M., AHMADNIA A., FEDOROVA S., BOLOGNESI C.: Limitations and Review of Geometric Deep Learning Algorithms for Monocular 3D Reconstruction in Architecture. Augmented reality and Artificial intelligence: Cultural Heritage and Innovative Design (2021). doi:10.3280/oa-686.68. 3

[TTZ20] TONO A., TONO H., ZANI A.: Encoded Memory: Artificial Intelligence and Deep Learning in Architecture. Impact of Industry 4.0 on Architecture and Cultural Heritage (2020). doi:doi.org/10.3280/oa-686.68. 3

[TZF04] TAI C.-L., ZHANG H., FONG J. C.-K.: Prototype Modeling from Sketched Silhouettes based on Convolution Surfaces. Computer Graphics Forum 23, 1 (2004), 71–83. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8659.2004.00006.x, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8659.2004.00006.x, doi:https://doi.org/10.1111/j.1467-8659.2004.00006.x. 3

[UKS*21] UY M. A., KIM V. G., SUNG M., AIGERMAN N., CHAUDHURI S., GUIBAS L. J.: Joint learning of 3d shape retrieval and deformation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021), pp. 11708–11717. URL: https://openaccess.thecvf.com/content/CVPR2021/papers/Uy_Joint_Learning_of_3D_Shape_Retrieval_and_Deformation_CVPR_2021_paper.pdf, doi:10.1109/CVPR46437.2021.01154. 7

[U.S24] U.S. COPYRIGHT OFFICE: Copyright and Artificial Intelligence, 2024. 19

[USB22] UNLU G., SAYED M., BROSTOW G.: Interactive Sketching of Mannequin Poses. In 2022 International Conference on 3D Vision (3DV) (2022), pp. 700–710. doi:10.1109/3DV57658.2022.00080. 3

[USGB24] UNLU G. E., SAYED M., GRYADITSKAYA Y., BROSTOW G.: GroundUp: Rapid Sketch-Based 3D City Massing. In Proceedings of the European Conference on Computer Vision (ECCV) (July 2024). 5, 19

[VACOS23] VINKER Y., ALALUF Y., COHEN-OR D., SHAMIR A.: CLIPascene: Scene Sketching with Different Types and Levels of Abstraction . In 2023 IEEE/CVF International Conference on Computer Vision (ICCV) (Los Alamitos, CA, USA, Oct. 2023), IEEE Computer Society, pp. 4123–4133. URL: https://doi.ieeecomputersociety.org/10.1109/ICCV51070.2023.00383, doi:10.1109/ICCV51070.2023.00383. 6

[VPB*22] VINKER Y., PAJOUHESHGAR E., BO J. Y., BACHMANN R. C., BERMANO A. H., COHEN-OR D., ZAMIR A., SHAMIR A.: CLIPasso: Semantically-Aware Object Sketching. ACM Transactions on Graphics (SIGGRAPH) 41, 4 (2022). URL: https://doi-org.stanford.idm.oclc.org/10.1145/3528223.3530068, doi:10.1145/3528223.3530068. 6, 7, 9, 10, 11, 33

[VSP*17] VASWANI A., SHAZEER N., PARMAR N., USZKOREIT J., JONES L., GOMEZ A. N., KAISER L., POLOSUKHIN I.: Attention is all you need. In Advances in Neural Information Processing Systems (2017), vol. 2017-December, p. 5999 – 6009. Cited by: 66591. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85043317328&partnerID=40&md5=3e5a5c2b862c8979ffea845bb707b3c3. 10, 11

[VSZ*25] VINKER Y., SHAHAM T. R., ZHENG K., ZHAO A., E FAN J., TORRALBA A.: Sketchagent: Language-driven sequential sketch generation. In Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR) (June 2025), pp. 23355–23368. 19

[Wan24] WANG F.: Sketch2Vox: Learning 3D Reconstruction from a Single Monocular Sketch Image. In Proceedings of the European Conference on Computer Vision (ECCV) (2024), Springer. 5, 8, 9, 15, 17, 32

[WCHW24] WANG R., CAO Y., HAN K., WONG K.-Y. K.: A Survey on 3D Human Avatar Modeling – From Reconstruction to Generation, 2024. URL: https://arxiv.org/abs/2406.04253, arXiv:2406.04253. 3

[WDL*23] WANG H., DU X., LI J., YEH R. A., SHAKHNAROVICH G.: Score Jacobian Chaining: Lifting Pretrained 2D Diffusion Models for 3D Generation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2023), pp. 12619–12629. 13

[WGXL19] WANG H., GE S., XING E. P., LIPTON Z. C.: Learning Robust Global Representations by Penalizing Local Predictive Power, 2019. URL: https://arxiv.org/abs/1905.13549, arXiv:1905.13549. 31

[WHE*24] WANG S.-Y., HERTZMANN A., EFROS A. A., ZHU J.-Y., ZHANG R.: Data Attribution for Text-to-Image Models by Unlearning Synthesized Images. In NeurIPS (2024). 19

[WKL15] WANG F., KANG L., LI Y.: Sketch-based 3D Shape Retrieval using Convolutional Neural Networks. CoRR abs/1504.03504 (2015). doi:10.1109/CVPR.2015.7298797. 4

[WKYS23] WU Y., KOUTA M., YUN SUEN P.: OwnDiffusion: A Design Pipeline Using Design Generative AI to preserve Sense Of Ownership. In SIGGRAPH Asia 2023 Posters (New York, NY, USA, 2023), SA '23, Association for Computing Machinery. URL: https://doi.org/10.1145/3610542.3626142, doi:10.1145/3610542.3626142. 19

[WLW*23] WANG Z., LU C., WANG Y., BAO F., LI C., SU H., ZHU J.: ProlificDreamer: High-Fidelity and Diverse Text-to-3D Generation with Variational Score Distillation. In Thirty-seventh Conference on Neural Information Processing Systems (2023). URL: https://openreview.net/forum?id=ppJuFSOAnM. 12

[WLY*20] WANG J., LIN J., YU Q., LIU R., CHEN Y., YU S. X.:

3D Shape Reconstruction from Free-Hand Sketches. arXiv pre-print (2020). URL: https://arxiv.org/abs/2006.09694, doi:2006.09694. 2, 5, 7, 8, 9, 15, 16, 32, 33

[WN19] WU C.-Y., NEUMANN U.: Salient Building Outline Enhancement and Extraction Using Iterative L0 Smoothing and Line Enhancing. In 2019 IEEE International Conference on Image Processing (ICIP) (2019), pp. 944–948. doi:10.1109/ICIP.2019.8803054. 7

[WOG12] WINNEMÖLLER H., OLSEN J., GOOCH B.: XDoG: Advanced image stylization with extended difference-of-Gaussians. ACM Transactions on Graphics (TOG) 31, 6 (2012), 1–10. 7

[WPL*21] WILLIS K. D. D., PU Y., LUO J., CHU H., DU T., LAMBOURNE J. G., SOLAR-LEZAMA A., MATUSIK W.: Fusion 360 Gallery: A Dataset and Environment for Programmatic CAD Construction from Human Design Sequences. The International Conference on Learning Representations (ICLR) 40, 4 (2021). doi:10.1145/3450626.3459818. 7

[WQF*21] WANG Z., QIU S., FENG N., RUSHMEIER H., MCMILLAN L., DORSEY J.: Tracing Versus Freehand for Evaluating Computer-Generated Drawings. ACM Transactions on Graphics (SIGGRAPH) 40, 4 (August 2021). URL: https://doi.org/10.1145/3450626.3459819, doi:10.1145/3450626.3459819. 7, 8

[WQWF18] WANG L., QIAN C., WANG J., FANG Y.: Unsupervised Learning of 3D Model Reconstruction from Hand-Drawn Sketches. In Proceedings of the 26th ACM International Conference on Multimedia (New York, NY, USA, 2018), MM '18, Association for Computing Machinery, p. 1820–1828. URL: https://doi.org/10.1145/3240508.3240699, doi:10.1145/3240508.3240699. 4

[WS19] WORTMANN T., SCHROEPFER T.: From Optimization to Performance-Informed Design. Proceedings of the Symposium on Simulation for Architecture and Urban Design (SIMAUD) (2019). 15

[WWF*23] WU Z., WANG Y., FENG M., XIE H., MIAN A.: Sketch and Text Guided Diffusion Model for Colored Point Cloud Generation. In Proceedings of the IEEE International Conference on Computer Vision (2023), p. 8895 – 8905. doi:10.1109/ICCV51070.2023.00820. 5, 9, 10, 14, 15, 16, 17, 19, 32

[WWH*25a] WANG T., WU Z., HE Q., CHU J., QIAN L., CHENG Y., XING J., ZHAO J., JIN L.: StickMotion: Generating 3D Human Motions by Drawing a Stickman, 2025. URL: https://arxiv.org/abs/2503.04829, arXiv:2503.04829. 3

[WWH*25b] WANG Z., WANG T., HE Z., HANCKE G. P., LIU Z., LAU R. W. H.: Phidias: A generative model for creating 3d content from text, image, and 3d conditions with reference-augmented diffusion. In The Thirteenth International Conference on Learning Representations (2025). URL: https://openreview.net/forum?id=TEkoMEjf7E. 19

[WWX*17] WU J., WANG Y., XUE T., SUN X., FREEMAN W. T., TENENBAUM J. B.: MarrNet: 3D Shape Reconstruction via 2.5D Sketches. Advances in Neural Information Processing Systems (NeurIPS) (2017). 9

[WWY25] WU Z., WANG Q., YANG J.: SketchTriplet: Self-Supervised Scenarized Sketch-Text-Image Triplet Generation. IEEE Internet of Things Journal (2025). doi:10.1109/JIOT.2024.3523382. 16

[WWZ*24] WU Z., WANG Q., ZHENG X., YE J., YANG P., WANG Y., WANG Y.: Doodle Your Motion: Sketch-Guided Human Motion Generation. IEEE Transactions on Visualization and Computer Graphics (2024), 1–11. doi:10.1109/TVCG.2024.3521333. 3

[WYZ*24] WU T., YUAN Y.-J., ZHANG L.-X., YANG J., CAO Y.-P., YAN L.-Q., GAO L.: Recent Advances in 3D Gaussian Splatting, 2024. URL: https://arxiv.org/abs/2403.11134, arXiv:2403.11134. 13

[WZF*23] WU T., ZHANG J., FU X., WANG Y., JIAWEI REN L. P., WU W., YANG L., WANG J., QIAN C., LIN D., LIU Z.: OmniObject3D: Large-Vocabulary 3D Object Dataset for Realistic Perception, Reconstruction and Generation. In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2023). 8, 13, 33

[WZW*24] WANG H., ZHAO M., WANG Y., QUAN W., YAN D.-M.: VQ-CAD: Computer-Aided Design model generation with vector quantized diffusion. Computer Aided Geometric Design 111 (2024), 102327. URL: https://www.sciencedirect.com/science/article/pii/S016783962400061X, doi:https://doi.org/10.1016/j.cagd.2024.102327. 3

[WZY*24] WANG Z., ZHU X., YU J., ZHANG T., LEI Z.: S2TD-Face: Reconstruct a Detailed 3D Face with Controllable Texture from a Single Sketch. MM 2024 - Proceedings of the 32nd ACM International Conference on Multimedia (2024), 6453 – 6462. doi:10.1145/3664647.3681159. 3

[XABP24] XIE T., AIGERMAN N., BELILOVSKY E., POPA T.: Sketch-guided Cage-based 3D Gaussian Splatting Deformation, 2024. URL: https://arxiv.org/abs/2411.12168, arXiv:2411.12168. 14, 16

[XCS*14] XU B., CHANG W., SHEFFER A., BOUSSEAU A., MCCRAE J., SINGH K.: True2Form: 3D curve networks from 2D sketches via selective regularization. ACM Trans. Graph. 33, 4 (jul 2014). URL: https://doi.org/10.1145/2601097.2601128, doi:10.1145/2601097.2601128. 3

[XHH*22] XU R., HAN Z., HUI L., QIAN J., XIE J.: Domain Disentangled Generative Adversarial Network for Zero-Shot Sketch-Based 3D Shape Retrieval, 2022. URL: https://arxiv.org/abs/2202.11948, arXiv:2202.11948. 4

[XHY*23] XU P., HOSPEDALES T. M., YIN Q., SONG Y.-Z., XIANG T., WANG L.: Deep Learning for Free-Hand Sketch: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 45, 1 (2023), 285–312. URL: http://dx.doi.org/10.1109/TPAMI.2022.3148853, doi:10.1109/tpami.2022.3148853. 7

[XKK*24] XIAO T., KIM K. G., KRUKAR J., SUBRAMANIYAN R., KIEFER P., SCHWERING A., RAUBAL M.: VResin: Externalizing spatial memory into 3D sketch maps. International Journal of Human-Computer Studies 190 (2024), 103322. URL: https://www.sciencedirect.com/science/article/pii/S107158192400106X, doi:https://doi.org/10.1016/j.ijhcs.2024.103322. 2, 3

[XNW*24] XU Y., NG Y., WANG Y., SA I., DUAN Y., LI Y., JI P., LI H.: Sketch2Scene: Automatic Generation of Interactive 3D Game Scenes from User's Casual Sketches, 2024. URL: https://arxiv.org/abs/2408.04567, arXiv:2408.04567. 3, 19

[XSL*22] XIAO C., SU W., LIAO J., LIAN Z., SONG Y.-Z., FU H.: DifferSketching: How Differently Do People Sketch 3D Objects? ACM Transactions on Graphics 41, 6 (2022). doi:10.1145/3550454.3555493. 6, 7, 8

[XT15] XIE S., TU Z.: Holistically-nested edge detection. In Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2015), pp. 1395–1403. 7

[XWJ*20] XIANG N., WANG R., JIANG T., WANG L., LI Y., YANG X., ZHANG J.: Sketch-based modeling with a differentiable renderer. Computer Animation and Virtual Worlds 31, 4-5 (2020), pp.e1939. doi:10.1002/cav.1939. 5, 9, 15, 32

[XWL*22] XU X., WILLIS K. D., LAMBOURNE J. G., CHENG C.-Y., JAYARAMAN P. K., FURUKAWA Y.: SkexGen: Autoregressive Generation of CAD Construction Sequences with Disentangled Codebooks. In International Conference on Machine Learning (2022), PMLR, pp. 24698–24724. 11

[XX23] XIA W., XUE J.-H.: A Survey on Deep Generative 3D-aware Image Synthesis. ACM Computing Surveys 56, 4 (November 2023), 1–34. URL: http://dx.doi.org/10.1145/3626193, doi:10.1145/3626193. 3, 12, 14

[YAB*22] YU E., ARORA R., BÆRENTZEN J. A., SINGH K., BOUSSEAU A.: Piecewise-smooth surface fitting onto unstructured 3D sketches. ACM Trans. Graph. 41, 4 (July 2022). URL: https://doi.org/10.1145/3528223.3530100, doi:10.1145/3528223.3530100. 3

[YBP*24] YUAN H., BOUSSEAU A., PAN H., ZHANG C., MITRA N. J., LI C.: DiffCSG: Differentiable CSG via Rasterization. Proceedings - SIGGRAPH Asia 2024 Conference Papers, SA 2024 (2024). arXiv:2409.01421, doi:10.1145/3680528.3687608. 8

[YCYW20] YAN G., CHEN Z., YANG J., WANG H.: Interactive liquid splash modeling by user sketches. ACM Trans. Graph. 39, 6 (November 2020). URL: https://doi.org/10.1145/3414685.3417832, doi:10.1145/3414685.3417832. 3

[YJK*23] YANG H.-B., JOHANES M., KIM F. C., BERNHARD M., HUANG J.: Architectural Sketch to 3D Model: An Experiment on Simple-Form Houses. In Computer-Aided Architectural Design. INTERCONNECTIONS: Co-computing Beyond Boundaries (Cham, 2023), Turrin M., Andriotis C., Rafiee A., (Eds.), Springer Nature Switzerland, pp. 53–67. 3, 7

[YSR*20] YAO Y., SCHERTLER N., ROSALES E., RHODIN H., SIGAL L., SHEFFER A.: Front2Back: Single View 3D Shape Reconstruction via Front to Back Prediction . In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (Los Alamitos, CA, USA, June 2020), IEEE Computer Society, pp. 528–537. URL: https://doi.ieeecomputersociety.org/10.1109/CVPR42600.2020.00061, doi:10.1109/CVPR42600.2020.00061. 2, 11

[YYW25] YAO J., YANG B., WANG X.: Reconstruction vs. Generation: Taming Optimization Dilemma in Latent Diffusion Models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2025), pp. 15703–15712. 16

[YZF*21] YANG L., ZHUANG J., FU H., WEI X., ZHOU K., ZHENG Y.: SketchGNN: Semantic Sketch Segmentation with Graph Neural Networks. ACM Trans. Graph. 40, 3 (August 2021). URL: https://doi.org/10.1145/3450284, doi:10.1145/3450284. 11

[YZM*23] YAN H., ZHANG H., MU X., FAN J., ZHANG Z.: FashionDiff: A Controllable Diffusion Model Using Pairwise Fashion Elements for Intelligent Design. In Proceedings of the 31st ACM International Conference on Multimedia (New York, NY, USA, 2023), MM '23, Association for Computing Machinery, p. 1401–1411. URL: https://doi.org/10.1145/3581783.3612127, doi:10.1145/3581783.3612127. 3

[ZCM*24] ZHAN Z., CHEN D., MEI J.-P., ZHAO Z., CHEN J., CHEN C., LYU S., WANG C.: Conditional Image Synthesis with Diffusion Models: A Survey, 2024. URL: https://arxiv.org/abs/2409.19365, arXiv:2409.19365. 4, 10

[ZGG21] ZHANG S., GUO Y., GU Q.: Sketch2Model: View-Aware 3D Modeling from Single Free-Hand Sketches. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR abs/2105.06663 (2021). doi:10.1109/CVPR46437.2021.00595. 5, 6, 7, 8, 9, 10, 12, 15, 16, 31, 32, 33

[ZGZS20] ZHONG Y., GRYADITSKAYA Y., ZHANG H., SONG Y.-Z.: Deep Sketch-Based Modeling: Tips and Tricks. 2020 International Conference on 3D Vision (3DV) (2020), 543 – 552. doi:10.1109/3DV50981.2020.00064. 3, 8, 9, 16, 18, 33

[ZGZS22] ZHONG Y., GRYADITSKAYA Y., ZHANG H., SONG Y.-Z.: A study of deep single sketch-based modeling: View/style invariance, sparsity and latent space disentanglement. Computers & Graphics 106 (2022), 237–247. URL: https://www.sciencedirect.com/science/article/pii/S0097849322001078, doi:https://doi.org/10.1016/j.cag.2022.06.005. 6

[ZHD*24] ZANG Y., HAN Y., DING C., ZHANG J., CHEN T.: Magic3DSketch: Create Colorful 3D Models From Sketch-Based 3D Modeling Guided by Text and Language-Image Pre-Training. Neurocomputing 661 (2024). URL: https://arxiv.org/abs/2407.19225, doi:10.1016/j.neucom.2025.131925. 5, 9, 14, 15, 31, 32

[ZHH96] ZELEZNIK R. C., HERNDON K. P., HUGHES J. F.: SKETCH: an interface for sketching 3D scenes. In Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques (New York, NY, USA, 1996), SIGGRAPH '96, Association for Computing Machinery, p. 163–170. URL: https://doi.org/10.1145/237170.237238, doi:10.1145/237170.237238. 2

[ZIE*18] ZHANG R., ISOLA P., EFROS A. A., SHECHTMAN E., WANG O.: The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In European Conference on Computer Vision (2018), p. 586 – 595. doi:10.1109/CVPR.2018.00068. 13

[ZLD16] ZHENG Y., LIU H., DORSEY J.: S mart C anvas : Context-inferred Interpretation of Sketches for Preparatory Design Studies. Computer Graphics Forum 35 (05 2016), 37–48. doi:10.1111/cgf.12809. 19

[ZLX*25] ZHANG N., LI M., XU C., FENG H., HU X., ZHANG J., CAO W., WANG C., FU Y.: StrandDesigner: Towards Practical Strand Generation with Sketch Guidance. MM 2025 - Proceedings of the 33rd ACM International Conference on Multimedia, Co-Located with MM 2025 (2025), 10296 – 10304. doi:10.1145/3746027.3755529. 3

[ZLY*23] ZHOU J., LUO Z., YU Q., HAN X., FU H.: GA-Sketching: Shape Modeling from Multi-View Sketching with Geometry-Aligned Deep Implicit Functions. Computer Graphics Forum 42, 7 (2023). doi:10.1111/cgf.14948. 5, 6, 9, 15, 16, 17, 32

[ZLZ*25] ZHOU Y., LI M., ZENG X., LIN J., ZHOU Y.: Sketch2Symm: Symmetry-aware sketch-to-shape generation via semantic bridging, 2025. URL: https://arxiv.org/abs/2510.11303, arXiv:2510.11303. 19

[ZPW*23] ZHENG X.-Y., PAN H., WANG P.-S., TONG X., LIU Y., SHUM H.-Y.: Locally Attentional SDF Diffusion for Controllable 3D Shape Generation. ACM Transactions on Graphics (SIGGRAPH) 42, 4 (2023). 5, 6, 9, 10, 15, 16, 17, 32, 33

[ZQG*20] ZHONG Y., QI Y., GRYADITSKAYA Y., ZHANG H., SONG Y.-Z.: Towards Practical Sketch-based 3D Shape Generation: The Role of Professional Sketches. IEEE Transactions on Circuits and Systems for Video Technology (2020). 2, 3, 5, 6, 7, 8, 9, 11, 12, 15, 16, 32

[ZRA23] ZHANG L., RAO A., AGRAWALA M.: Adding conditional control to text-to-image diffusion models. In Proceedings of the IEEE/CVF International Conference on Computer Vision (2023), pp. 3836–3847. 10

[ZXC*24] ZHENG W., XIA H., CHEN R., SUN L., SHAO M., XIA S., DING Z.: Sketch3d: Style-consistent guidance for sketch-to-3d generation. In Proceedings of the 32nd ACM International Conference on Multimedia (New York, NY, USA, 2024), MM '24, Association for Computing Machinery, p. 3617–3626. URL: https://doi.org/10.1145/3664647.3680641, doi:10.1145/3664647.3680641. 2, 5, 7, 8, 9, 13, 14, 15, 16, 17, 19, 32, 33

[ZYG*24] ZOU Z.-X., YU Z., GUO Y.-C., LI Y., LIANG D., CAO Y.-P., ZHANG S.-H.: Triplane Meets Gaussian Splatting: Fast and Generalizable Single-View 3D Reconstruction with Transformers. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2024), 10324 – 10335. doi:10.1109/CVPR52733.2024.00983. 13

[ZZX24] ZHANG Y., ZHANG T., XIE H.: TexControl: Sketch-Based Two-Stage Fashion Image Generation Using Diffusion Model. Proceedings - 2024 Nicograph International, NICOInt 2024 (2024), 64 – 68. doi:10.1109/NICOInt62634.2024.00021. 3

11. Supplementary

In the supplementary, we summarize the different datasets used in the literature for DS-3DM in Table 8 for our MORPHEUS design space.

Disclosure Statement: None declared

11.1. Architectures

In this section, we display the different model architectures highlighted in Table 7 through different diagrams; see Figures 12 and 13

11.2. Questionnaire

Free2CAD evaluated expressiveness and accessibility by asking five novice users to watch a 20-minute tutorial about modeling and then a 15-minute time range to draw from a single isometric view of the object.

Sketch-a-Shape uses the Amazon SageMaker Ground Truth and crowd workers from the Mechanical Turk workforce. It generates 3D shapes with sketches from different datasets [EHA12, ZGG21, WGXL19, HE17a].

GA-Sketching performed both a usability study and a perceptive study. The usability study was conducted on eight novice 3D modelers aged 18 to 28 years. They watched an 8-minute video and had 15 minutes to familiarize themselves with the system. They had to create two models (a chair and an airplane). After the session, they had to fill in a System Usability Scale (SUS) questionnaire (5-point scale) and a NASA Task Load Index (NASA-TLX) questionnaire to evaluate the mental demand, physical demand, temporal demand, performance, effort, frustration.

The perceptive study was conducted through an online questionnaire with 40 novices.

In Table 11.2, we presented evaluation metrics related to the Quality, Similarity, Alignment, and Usability of the 3D generation methods. Some other methods are also focused on the reconstruction quality. Therefore, they evaluate the Fidelity and Accuracy. Reconstruction-based evaluations focus more on retrieval problems; therefore, the setting and goal differ.

- **Fidelity/Accuracy:**
  - How well does the output 3D model match the input sketch? [ZHD*24]
  - Does the generated 3D model represent the sketch? (Y/N) [BBD22]
  - Are there any features of the sketch missing in the generated 3D model? (Y/N) [BBD22]
  - Which of the 3D models on the right-hand side best matches the sketch on the left-hand side?
- **Quality:**
  - How do you think of the quality of the output 3D model? [ZHD*24]
  - What is the quality of the generated 3D model? (on a scale of 1-5) (5-best) [BBD22]
  - How realistic is the shape? [BHSH*24]
- **Similarity:**
  - How close is the output to the input sketch (Likert)? [BBD22]
  - How close to the input sketch does the resulting chair look? [BHSH*24]
- **Alignment with Text Prompts:**
  - The alignment to the text prompt [CPL*23]
  - Text faithfulness [LFLG24]
- **User Experience and Usability:**
  - On a Likert scale from 1 (strongly disagree) to 5 (strongly agree), do you agree that our system allowed them to achieve the buildings they wanted? [NGDA*16]
  - Do you agree that the system interpreted well the shapes they drew? [NGDA*16]
  - Choose the better one by jointly considering the following three aspects: (1) the alignment to the text prompt, (2) the fidelity of the visual appearance, and (3) the accuracy of the geometry. [CPL*23]
  - Rate the generated shapes based on how they match their expectation using Likert.
  - Rate the satisfaction score based on generation speed, shape quality, consistency, and resolution, using Likert.
  - Rate "system functional integrity", "user interface convenience", "generated results satisfaction", and "conformity to expectations" by using Likert .
  - Answer questions to evaluate the system based on the System Usability Scale (SUS).
  - Was this approach consistent with your drawing habits? [DZX24]
  - Was it reasonable to use the generation method for car shells?
  - Did the generated car shell models basically meet your expectations? [DZX24]
  - Are you satisfied with the retrieved models? [DZX24]
  - Answer a System Usability Scale (SUS) questionnaire and a NASA Task Load Index (NASA-TLX) questionnaire to evaluate the usability and workload of our system.

Paper	Input									Models						Output								
	Amount			View			Style										Part Semantic			Options			Geometry		
Nishida et al. [NGDA*16]			■				■			■									■			■			
Delanoy et al. [DBA*17]	■			■						■								■				■			
ShapeMVD [LGK*17]	■	■		■			■			■										■		■			
Contour3D [JFD20]		■		■			■			■										■			■		
DeepSketch [ZQG*20]	■				■			■		■			■				■			■				■	
Sketch2CAD [LPBM20]	■	■		■			■			■								■		■			■		
SketchDiff [XWJ*20]			■	■			■						■							■			■		
FreeHandRec [WLY*20]			■		■				■		■									■				■	
Sketch2Model [ZGG21]			■		■			■		■	■			■						■			■		
Sketch2Mesh [GRYF21]			■		■		■			■				■			■				■		■		
Free2CAD [LPBM22]	■	■		■			■			■			■				■			■					■
SS2Mesh [BBD22]			■	■				■		■	■									■			■		
GeoCode [PLH*22]	■			■			■			■	■							■			■		■		
SketchSampler [GYS*22]	■			■					■	■										■			■		
LAS-Diffusion [ZPW*23]			■		■			■			■	■									■		■		
Sketch-A-Shape [SJR*23]			■			■			■		■										■		■		
SKED [MPS*23]		■		■					■					■	■		■			■					■
CLIPXPlore [HHL*23]			■			■		■				■	■	■			■				■				■
D3DSketch+ [CFZ*23]			■	■			■					■	■							■			■		
Control3D [CPL*23]			■			■			■		■	■	■	■						■			■		
Re3DSketch [CDZ*24]			■		■		■				■		■					■			■				
Sketch2Point [KWQ23]			■			■		■				■								■			■		
GA-Sketching [ZLY*23]			■			■		■			■			■						■			■		
S2PointCol [WWF*23]			■			■			■		■	■						■		■					■
Sketch2Vox [Wan24]	■				■			■					■				■			■			■		
SketchDream [LFLG24]			■		■		■				■	■		■	■		■				■				■
SENS [BHSH*24]			■			■			■		■	■	■					■			■		■		
DY3D [BKD*24]			■			■			■		■	■	■					■			■		■		
Vitruvio [THAF24]	■				■			■			■									■					■
MVControl [LCZL25]			■		■			■			■	■	■	■	■					■			■		
SHLine [FQS*24]		■			■			■			■	■	■							■				■	
M3DSketch [ZHD*24]			■		■		■							■			■			■			■		
Sketch2NeRF [CYW*24]			■			■			■		■	■		■	■		■				■				■
DualShape [DZX24]		■		■			■				■									■			■		
Sketch3D [ZXC*24]			■			■			■		■			■	■		■			■				■	

Table 7: *Unified categorization of* DS-3DM *methods (Section 5). This table integrates input characteristics (Amount, View, Style), Model architectures, and Output capabilities (Part semantic, Options, Geometry). The Input and Output sections are highlighted to distinguish the processing stages. Filled cells indicate the presence of a feature or architecture component.*

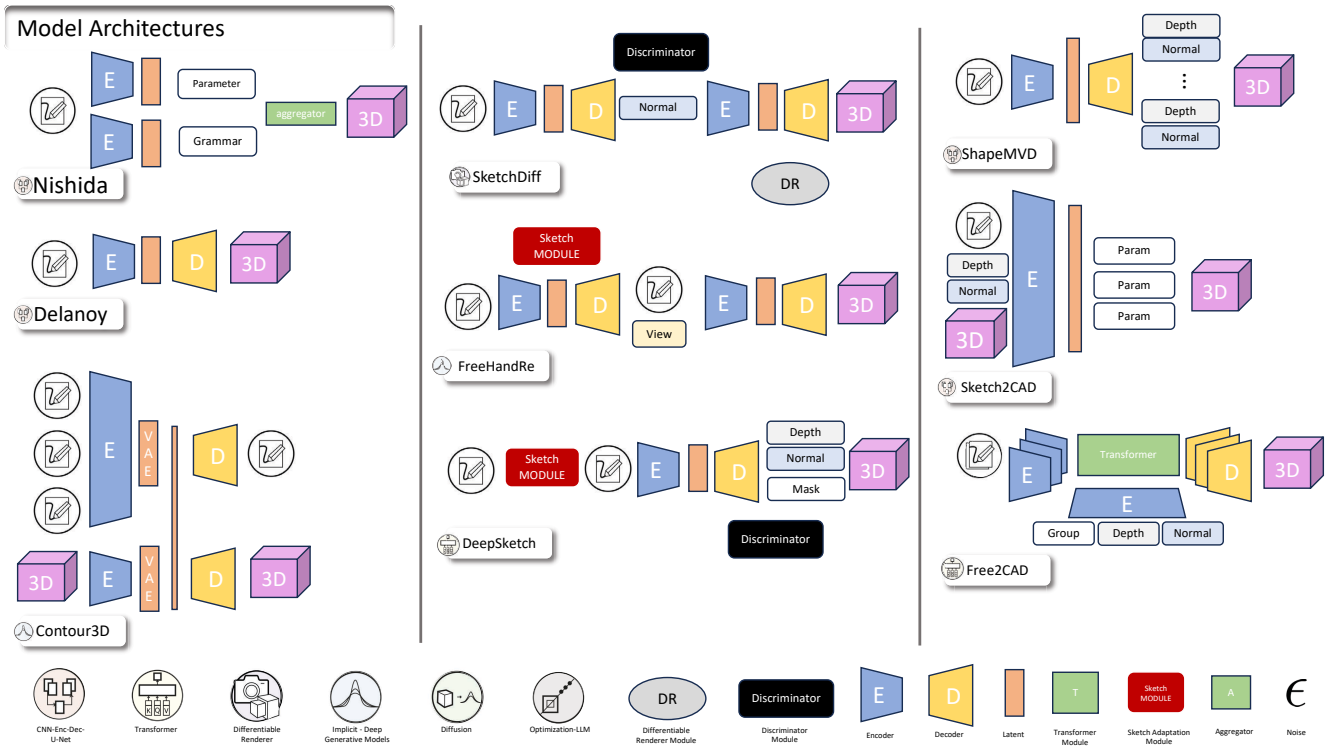


Figure 12: *Part 1 illustrates the architectures of AI models for* DS-3DM *methods.*

Paper	Dataset	Category	Shapes	Sketches	Sketch/Shape	Viewpoints	Usage	Format	Style
Nishida 2016 *[NGDA*16]*	Parametric Primitives	Parametric Primitives	procedural	grammar	-	1	TE	PNG	SSAO
Delanoy 2020 *[DBA*17]*	ShapeCOSEG	Chair /vase	700	5.600	–	8	TE	PNG	SC
	Primitives	Primitives	–	–	–	8	TE	PNG	SC
Gryaditskaya 2019 *[GSH*19]*	OpenSketch	Product Design	12	400	400 / 12	3	–	PNG SVG	H 15 PV
Wang 2020 *[WLY*20]*	ShapeNet	13 categories	43.783	390	390 / 130	3	E	PNG	H 15 wDS
Zhong 2020 *[ZGZS20]*	SNCore	Chair plane lamp	13.141	39.423	3 / 1	48	TE	PNG	Naive
	SNCore	Chair plane lamp	13.141	39.423	13 / 1	1	TE	SVG	Stylize
	SNCore	Chair plane lamp	13.141	39.423	13 / 1	1	TE	PNG SVG	Style-unified
	PS3d	Chair plane lamp	13.141	PS3d	PS3d	5	TE	PNG SVG	PS3d
Zhang 2021 *[ZGG21]*	SN-Sketch	13 categories	–	1.300	100 / 1	20	E	PNG	H copy wDS
	SN-Sketch	13 categories	–	1.300	100 / 1	21	TE	PNG	Canny
Guillard 2021 *[GRYF21]*	ShapeNet	Car chair	6.819	–	–	16	TE	PNG	Canny
	ShapeNet	Car chair	6.819	–	–	1	TE	PNG	SketchFD
	ShapeNet	Car chair	6.819	–	–	1	E	PNG	SC
	ShapeNet	Car	–	113	3 / 1	1	E	PNG	Students 5 tracing
	PS3d	Chair	–	177	3 / 1	1	E	PNG	P wDS
Vitruvio *[THAF24]*	Manhattan 1k	Building	1.000	24k	24 / 1	24	TE	PNG	Canny
Sketch3D *[ZXC*24]*	ShapeNet-Sketch3D	10 ShapeNet+Text	11k	220k	20 / 1	20		PNG	Canny
SketchDream, MVControl	Objaverse	3D Obj	400k		4 / 1	30	TE	PNG	Canny
Sketch2NeRF [CYW*24]	OmniObject3D-Sketch [WZF*23]	3D Obj - 20 categories	-	-	24 / 1	-	TE	PNG	HED
DoodleYour3D [BKD*24]	ShapeNet [CFG*15]	Chair	6,755	81,060	12 / 1	6	TE	PNG SVG	CLIPasso + InfDraw [VPB*22, CDI22]
SENS [BHSH*24]	ShapeNet [CFG*15]	Chair plane lamp	9,363	224,712	24 / 1	6	TE	PNG SVG	CLIPasso + others [BHSH*24]
LAS-Diffusion [ZPW*23]	ShapeNet [CFG*15]	5 Cat.	-	-	50 / 1	10	TE	PNG	Canny

Table 8: *Datasets for monocular sketch reconstruction. Few of these works involved professionals (P) or students in sketching on different digital surfaces (wDS), such as ISKN Slate 2 and iPad Pro. Some of them were also capable of tracking strokes. H indicates that a heterogeneous pool of people have been asked to sketch: students, professionals, and others. Ps3d: ProSketch3d [ZGZS20], SC: suggestive contours, T: Test , TE: Test and Evaluation.*

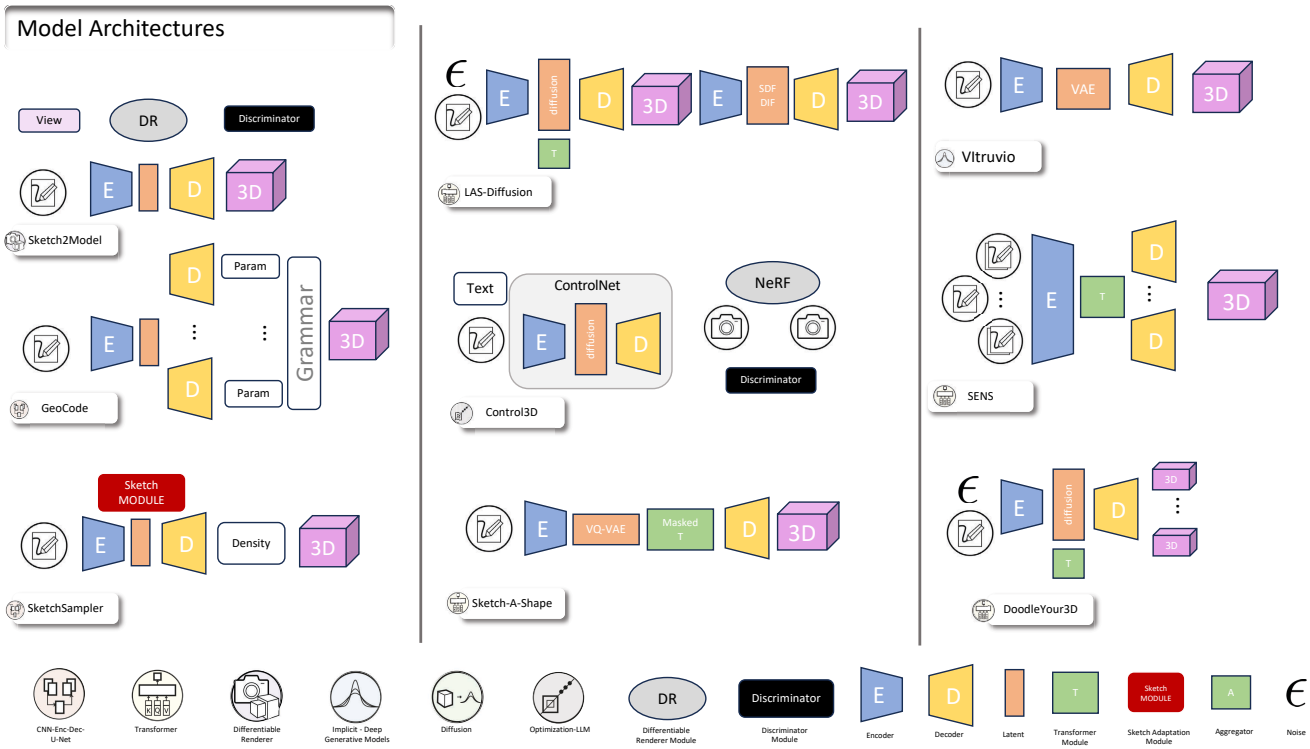


Figure 13: *Part 2 illustrates the architectures of AI models for* DS-3DM *methods.*