Proceedings of the ASME 2022 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference
IDETC/CIE 2022
August 14-17, 2022, St. Louis, Missouri, USA

IDETC/CIE 2022-90696

DEEP LEARNING OF CROSS-MODAL TASKS FOR CONCEPTUAL DESIGN OF ENGINEERED PRODUCTS: A REVIEW

Xingang Li
Walker Department of Mechanical Engineering
University of Texas at Austin
Austin, Texas 78712
Email: xingang.li@utexas.edu

Ye Wang
Autodesk Research
Pier 9
San Francisco, California 94611
Email: ye.wang@autodesk.com

Zhenghui Sha*
Walker Department of Mechanical Engineering
University of Texas at Austin
Austin, Texas 78712
Email: zsha@austin.utexas.edu

ABSTRACT

Conceptual design is the foundational stage of a design process, translating ill-defined design problems to low-fidelity design concepts and prototypes. While deep learning approaches are widely applied in later design stages for design automation, we see fewer attempts in conceptual design for three reasons: 1) the data in this stage exhibit multiple modalities: natural language, sketches, and 3D shapes, and these modalities are challenging to represent in deep learning methods; 2) it requires knowledge from a larger source of inspiration instead of focusing on a single design task; and 3) it requires translating designers' intent and feedback, and hence needs more interaction with designers and/or users. With recent advances in deep learning of cross-modal tasks (DLCMT) and the availability of large cross-modal datasets, we see opportunities to apply these learning methods to the conceptual design of product shapes. In this paper, we review 30 recent journal articles and conference papers across computer graphics, computer vision, and engineering design fields that involve DLCMT of three modalities: natural language, sketches, and 3D shapes. Based on the review, we identify the challenges and opportunities of utilizing DLCMT in 3D shape concepts generation, from which we propose a list of research questions pointing to future research directions.

*Corresponding author.

1 INTRODUCTION

Product shape is essential in the conceptual design of engineered products, because it could affect both the aesthetics and engineering performance of a product [1]. Figure 1 shows the overall information flow and the key steps in conceptual design, where the information can be categorized into three modalities: natural language (e.g., text), sketches (e.g., 2D silhouette), and 3D shapes (e.g., meshes). We call them design modalities. Generally, documents of customer needs and engineering requirements are in the form of natural languages. Design sketches and drawings are effective ways for brainstorming and expressing designers' preferences. Low-fidelity design concepts and prototypes from the conceptual design stage are often represented by 3D shapes in digital format. Design Search and Design Creation are two important steps in conceptual design to gather information of existing design solutions for inspiration and explore the design space for novel design concepts.

Early design automation methods, such as grammar- and rule-based methods, primarily rely on human design experience and knowledge to generate design alternatives [2]. On the contrary, driven by data, deep learning methods have been primarily applied to the later stages of engineering design for design automation [3]. It is challenging to apply deep learning methods to the conceptual design stage (i.e., the early design stage) for several reasons. For example, data in the conceptual design stage exhibit multiple modalities, but deep learning methods are usually applied to handle a single design modality. Moreover, in conceptual design, designers often gather a large set of infor-

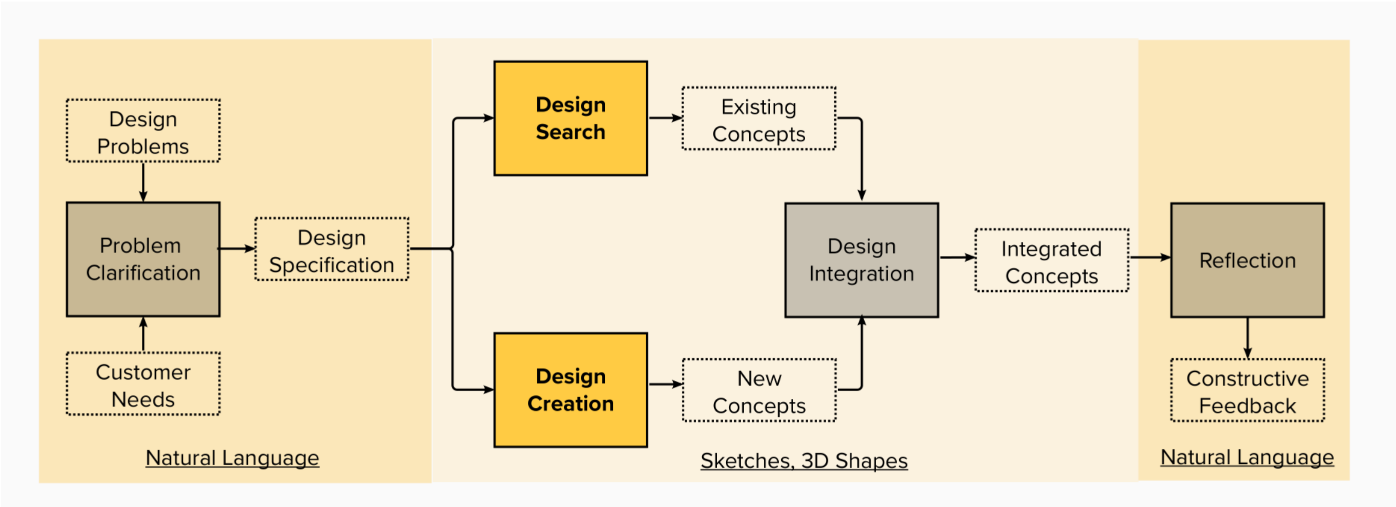

FIGURE 1. Conceptual design stage in the development of engineered products

mation for design inspiration, but deep learning methods tend to focus on a single design task. Finally, human (either user or designer) inputs and interactions are desired in conceptual design to improve design creativity and human-centered design, but deep learning-based design methods do not directly interact with human data but only implicitly capture human preferences from training datasets.

With recent development in deep learning of cross-modal tasks (DLCMT) [1], we see the opportunities of applying those methods to address these challenges, particularly in product shape design, such as car body and plane fuselage [5, 6]. DLCMT methods can allow human input of one design modality and translate them to another design modality, e.g., from natural language to 3D shapes. In DLCMT, there are cross-modal retrieval and generation methods. Cross-modal retrieval can be used to search an existing design repository for inspiring design ideas, while cross-modal generation can be used to explore a design space to create novel design concepts. They can be used in both the Design Search step and the Design Creation step (see Figure 1).

In this paper, we review 30 recent journal articles and conference papers from computer graphics, computer vision, and engineering design fields on DLCMT. We focus on text, sketches, and 3D shapes because they are the major design modalities in conceptual design. Specifically, we reviewed deep learning methods on three cross-modal tasks: text-to-sketch, text-to-3D, and sketch-to-3D. We found that most of the literature comes from computer graphics and computer vision, with few attempts in engineering design applications. This poses new opportunities to adapt the developed models and techniques to solve engineering design problems, and particularly, to bridge human inputs and interactions with deep learning methods in the conceptual design of product shapes.

The remainder of this paper is organized as follows. Section 2 introduces background knowledge in conceptual design and design modalities. Section 3 presents the scope and criteria of our review. We show and summarize the review results in Section 4. In Section 5, we discuss the challenges and research questions when applying DLCMT in the conceptual design of engineered products. Section 6 concludes our work with closing remarks and a brief discussion on future research directions.

2 BACKGROUND

2.1 CONCEPTUAL DESIGN IN PRODUCT ENGINEERING

In the conceptual design stage of a design process, product shape is a key consideration that is highly related to a product's engineering performance and aesthetics [1, 7]. Therefore, in this paper we focus on 3D shape design. As shown in Figure 1, we adapt the five-step concept generation method [1] to facilitate the review process. The five steps are Problem Clarification, Design Search, Design Creation, Design Integration, and Reflection. Through these five steps, the method transfers information, such as customer needs, engineering requirements, and design ideas, to design concepts in the form of sketches and 3D shapes. The corresponding input and output of each step are represented

[1]DLCMT is a class of problems, aiming to translate one modality of data to another, e.g., from text to 3D shapes. To solve this problem, there is a large body of literature on cross-modal representation learning (CMRL). CMRL aims to build embeddings using information from multiple modalities (e.g., texts, audio, and images) in a common semantic space, which allows the model to compute cross-modal similarity [4]. In this paper, our review is not limited to reviewing CMRL methods but also include other deep learning methods that can solve cross-modal problems.

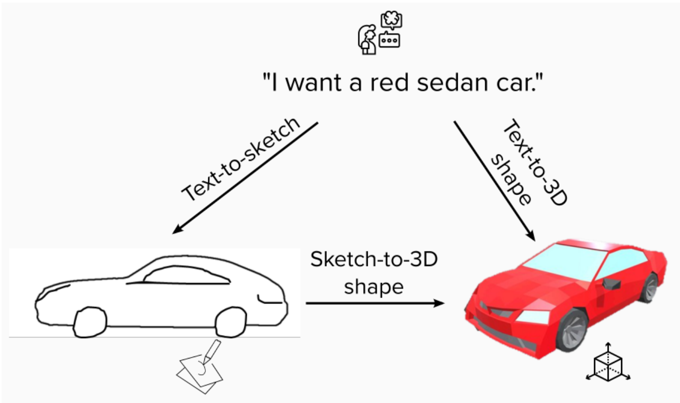

FIGURE 2. Cross-modal tasks in conceptual design

by dotted rectangles. The process is linear in sequence from left to right but almost always iterative. For example, feedback from Reflection could influence Problem Clarification and its subsequent steps. In this paper, we primarily focus on the review of DLCMT that can be applied in the steps of Design Search and Design Creation because they are the important steps for the inspiration of novel design concepts.

2.2 DESIGN SEARCH

Design Search is the step of gathering information of existing design solutions to a design problem. In practice, several ways, such as patents, literature, and benchmarking, can be used to gather useful information [1]. By analyzing those existing products, designers can summarize their advantages and disadvantages, so that they can make necessary and customized changes to existing designs to create satisfying ones. However, the repository of existing design options could be huge, so the search process would be time consuming and cumbersome, placing significant cognitive and physical burdens on designers. One possible solution to this issue is to use an AI-assisted search process, where designers can pre-define search criteria and utilize computers to search for relevant design solutions. For 3D shape design, shape retrieval methods are representative examples.

2.3 DESIGN CREATION

Design Creation emphasizes exploring novel design concepts. Designers brainstorm ideas and explore the design space to create novel design concepts based on the knowledge of designers [1]. Sketches are often used for brainstorming design concepts, and low-fidelity 3D models are then created for better visualization and further development. However, creating 3D models involves lots of manual work and could be time-consuming. DLCMT can be used to automate the sketch-to-3D shape process, thus facilitating the creation of novel 3D shapes.

TABLE 1. Statistics of the papers reviewed

Category	Method	Number of Papers
Text to Sketch	Retrieval	None
	Generation	5
Text to 3D Shape	Retrieval	4
	Generation	6
Sketch to 3D Shape	Retrieval	6
	Generation	9

2.4 MODALITIES IN CONCEPTUAL DESIGN

As shown in Figure 1, there are three major design modalities: natural language (NL), sketches, and 3D shapes, in the conceptual design process. In the example of a car body design as shown in Figure 2, the three modalities could be ”I want a red sedan car” (NL), hand-sketching a car with desired features (sketch), and then creating a CAD model of the car (3D shape). Natural language allows people to convey and communicate ideas and thoughts. It is also the primary means for documentation, such as documentation of customer needs and engineering requirements. Sketches are often used to brainstorm design concepts because sketching can stimulate designers’ creative imagination [8]. Then, a 3D shape is often built to provide better visualization and a low-fidelity prototype model for further evaluation and development of a concept.

Natural language data are often in the format of text. Sketches are in 2D imaginary format consisting of strokes and lines. 3D shapes are typically built as boundary representation (B-rep) models (e.g., NURBS) using CAD software in engineering design. However, in the computer graphics field and the 3D deep learning field, 3D shapes are usually parameterized as meshes, point clouds, and voxel grids. Compared to CAD models, these 3D representations typically have lower fidelity with fewer geometric details and structural information because of the limitations of computational resources.

3 REVIEW SCOPE AND METHODOLOGY

We confined the scope of our review to the following three aspects: 1) 3D shape design of discrete, physical, and engineered products. 2) Design Search and Design Creation steps (see Figure 1) in conceptual design, and 3) deep learning of cross-modal tasks for the three design modalities: text-to-sketch, text-to-3D, and sketch-to-3D.

To identify relevant studies and publications, we searched in Google Scholar by keywords: ”text-to-sketch retrieval”, ”text-to-sketch generation”, ”sketch-based 3D shape retrieval”, ”sketch-based 3D shape generation”, ”text-to-shape retrieval”, and ”text-to-shape generation”. These keywords are identified from the existing publications. For "sketch-based 3D shape generation", it has another three commonly used names: "sketch-based 3D shape reconstruction", "sketch-based 3D shape synthesis", and "3D shape reconstruction from sketches", so we also used them in our search. The initial search yielded a large set of seed papers (more than 600). The majority of those papers are related to sketch-based 3D shape retrieval and generation, from which we picked representative papers with high citations in Google Scholar and recently published papers after the year 2020. We didn't find any work related to text-to-sketch retrieval, possibly due to the lack of interests in practical applications. For the other categories, we filtered some irrelevant papers based on the three criteria defined in the scope of the review. The statistics of the papers reviewed are summarized in Table 1. We acknowledge that the inclusion of relevant work regarding the sektch-to-3D shape may not cover all areas, so it leaves a more comprehensive review task for our future work. However, this paper initiates the first step towards a comprehensive review of all relevant research on DLCMT that could be beneficial to the conceptual design of product shapes.

4 REVIEW RESULTS

4.1 METHODS FOR DESIGN SEARCH

In this section, we summarize our review related to text-to-3D and sketch-to-3D shape retrieval.

4.1.1 TEXT-TO-3D SHAPE RETRIEVAL

There has been only a little research on text-driven 3D content search. For the state-of-the-art methods, learning a joint embedding for text and 3D shapes is a common strategy.

Chen et al. [9] constructed a joint embedding of text and 3D shapes by using a convolution neural network (CNN) + recurrent neural network (RNN) encoder on text and a 3D-CNN encoder on 3D voxel shapes. A triplet loss was applied and learning by association was used to align the embedded representations of text and 3D shapes. They also introduced two datasets: 1) ShapeNet [10] (chairs and tables only) with a natural language description and 2) geometric primitives with synthetic text descriptions. However, the computational cost caused by the cubic complexity of 3D voxels limits this method to the machine learning of low-resolution voxels. Consequently, the learned joint representations will have a low discriminative ability. Han et al. [11] built a $Y^2Seq2Seq$ network architecture using a Gated Recurrent Unit (GRU) to encode features of multiple-view images to represent the shape. To obtain the joint embedding of text and sketches, they trained the network using both intermodality and intramodality reconstruction losses, in addition to the triplet loss and classification loss. Therefore, the proposed network could learn more discriminative representations than [9]. Tang et al. [12] proposed to incorporate part-level information of 3D shapes represented by point clouds. They applied a shape encoder with a pre-trained point-based segmentation network [13] to learn part embedding and a bidirectional GRU text encoder to learn word embedding. An alignment-based cross-attention module was then used to predict a pair of symmetrical formulations (i.e., shape-text and text-shape) to achieve the matching of parts with words.

The methods mentioned above need to use complex attention mechanisms or losses to learn the joint embedding. To avoid these complexities, using large batch contrasive learning, Ruan et al. [14] proposed a tri-modal learning scheme, which achieved state-of-the-art text-to-shape retrieval. The scheme consists of a 3D CNN voxel shape encoder, a bidirectional GRU (Bi-GRU) text encoder, and a multi-view CNN image encoder. The embedding vectors of the voxel shape, text, and image are aligned in a latent space. Then, a bidirectional contrastive loss was minimized to learn effective representations of shape, text, and image.

4.1.2 SKETCH-TO-3D SHAPE RETRIEVAL

Wang et al. [15] proposed to learn feature representations for sketch-to-3D shape retrieval, which avoided computing multiple views of a 3D model. They applied two Siamese CNNs for views of 3D shapes and sketches, respectively, and a loss function defined on the within-domain and cross-domain similarities. To reduce the discrepancies between the sketch features and the 3D shape features, Zhu et al. [16] built a pyramid cross-domain neural network of sketches and 3D shapes. They used the network to establish a many-to-one relationship between the sketch features and a 3D shape feature. The learned features of 3D shapes and sketches were used for retrieval. Dai et al. [17] proposed a novel deep correlated holistic metric learning method with two distinct neural networks for sketch and 3D shape. Such a deep learning method mapped features from both domains into one feature space. In the construction of its loss function, both discriminative loss and correlation loss were used to increase the discrimination of features within each domain and the correlation between domains.

In the methods mentioned above, deep metric learning [18] was applied to mitigate the modality discrepancy between the sketch and the 3D shape. There are also methods studying how to represent 3D shapes more comprehensively so that 3D shapes can better correspond to sketches. Xie et al. [19] proposed to learn a Wasserstein barycenter of CNN features extracted from the 2D projections of a 3D shape. They constructed the metric network to map sketches and the Wasserstein barycenters of 3D shapes to a common deep feature space. A discriminative loss was then formulated to learn the deep features. The deep features learned could then be used for the sketch-to-3D shape retrieval.

SHREC 2014 sketch-based 3D shape retrieval benchmark [20] was commonly used by all methods introduced above. These methods aimed to retrieve objects by coarse category-level retrieval of 3D shapes given an input sketch. Qi et al. [21] introduced a novel task of fine-grained instance-level sketch-to-3D shape retrieval, with the aim of retrieving one specific 3D shape that best matches the input sketch. They created a set of paired sketch-to-3D shape data of chairs and lamps from ShapeNet [10]. Then, they built a deep joint embedding learning-based model with a novel cross-modal view attention module to learn the features of sketches and 3D shapes. There is also an interest in using CAD data in 3D shape retrieval. Qin et al. [22] developed a sketch-to-3D CAD shape retrieval approach using variational auto encoder (VAE) and structural semantics. They created their training dataset by collecting 3D CAD models from local companies and obtained their six-view projections as sketch data.

4.2 METHODS FOR DESIGN CREATION

In this section, we summarize our review relevant to text-to-3D shape, text-to-sketch, and sketch-to-3D shape generation methods.

4.2.1 TEXT-TO-3D SHAPE GENERATION

Chen et al. [9] proposed a challenging task of text-to-3D shape generation. As introduced in Subsection 4.1.1, they constructed a joint embedding of text and 3D shape and used it for text-to-3D shape retrieval task. Furthermore, they combined the joint embedding model with a conditional Wasserstein GAN (WGAN) [23] framework, which enables the generation of colored voxel shapes in low resolution (32^3). To resolve this low-resolution issue, Fukamizu et al. [24] proposed a two-stage method that can first generate a low-resolution shape (32^3), which roughly reflected a given text (Stage I), and then generated a high-resolution (64^3) shape reflecting the detail of the text (Stage II). Stage I was built based on [9] and in Stage II, a new network model was built based on StackGAN [25]. Li et al. [26] proposed to use class labels to guide the generation of 3D voxel shapes with the assumption that shapes with different labels (e.g., chairs and tables) have different characteristics. They added an independent classifier to the WGAN framework to guide the training process. The classifier could be trained together with the generator to enable more distinctive class features in the generated 3D shapes.

The methods introduced above only support the generation of 3D shapes in individual ShapeNet [10] categories (e.g., the chair category or the table category). Generalization of these methods remains challenging due to the unavailability and limited size of paired data of 3D shapes and text description. To improve the ability of generalization, some methods try to utilize some pretrained models (e.g., Contrastive Language-Image Pre-Training (CLIP) [27]) and zero-shot learning techniques. Sanghi et al. [28] proposed a method called CLIP-forge, which could generate 3D voxel shapes from text descriptions for ShapeNet objects. It required training data (i.e., rendered images, voxel shapes, query points, and occupancies) obtained from ShapeNet 3D shapes without text labels. They first learned an encoding vector of a 3D geometry and then learned a normalizing flow model [29] of that encoding vector conditioned on a CLIP feature embedding. However, CLIP-Forge cannot be well generalized outside ShapeNet categories. Chu [30] proposed a method consisting of a generator, a scorer, and an optimization loop using a genetic algorithm to optimize 3D geometry. The objective function of the optimization was to maximize the CLIP score obtained from the rendered images of 3D surfaces. This method could generate 3D shapes and 2D renders at different view angles for a wide range of categories beyond ShapeNet data. Jain et al. [31] combined Neural Radiance Field (NeRF) [32] with an image-text loss from CLIP to form Dream Fields. A Dream Field is a neural 3D representation that can return a rendered 2D image given a desired viewpoint. After training, the method could generate colored 3D neural geometry from text prompts without using 3D shape data, resulting in a better generalization ability.

4.3 TEXT-TO-SKETCH GENERATION

Sketches could inspire design ideas and text-to-sketch tools could help designers efficiently capture fleeting design inspirations. The generation of images from text descriptions (i.e., text-to-image synthesis/generation) has seen greatly progressed recently [33]. Unlike text-to-image generation, text-to-sketch synthesis is more challenging, and can only rely on rigid edge/stroke information without color features (i.e., pixel values) in an image [34].

Text2Sketch [35] applied a Stagewise-GAN to encode human face attributes identified from text descriptions and transforms those attributes into sketches, which was trained on a manually annotated dataset of text-face sketches. Although the method was applied in face recognition instead of product design, it is worth being introduced here because the method is inspiring and could be applied to the design domain if a different dataset is used. Yuan et al. [34] constructed a bird sketch dataset by modifying the Caltech-UCSD Birds (CUB) dataset [36], based on which they trained a novel GAN-based model, called T2SGAN. The model featured a Conditional Layer-Instance Normalization module that could fuse the image features and sentence vectors, thus efficiently guiding the generation of sketches.

The methods mentioned above were developed for single-object sketch synthesis, and there are also methods for multi-object generation, which could be useful for generating designs part by part. Huang et al. [37] adopted a two-step neural network: 1) a transformer-based mixture density network for the scene composer to generate high-level layouts of sketches, and 2) a sketch-RNN [38] based object sketcher to generate individual object sketches. The scene composer and the object sketcher were trained using the Visual Genome dataset [39] and the "Quick, Draw!" dataset [40], respectively. Since different datasets of text and sketches can be used, this method helped avoid the requirement for paired data of text description and sketches of an object. Based on their previous work, Huang et al. [41] took a further step and proposed an interactive sketch generation system called Scones. It used a Composition Proposer to propose a scene-level composition layout of objects and an Object Generator to generate individual object sketches. Bhunia et al. [42] introduced a network architecture consisting of a part locator network and a part sketcher network to generate realistic creative sketches. The part locator networks, composed of two graph-aware transformer encoders and a Gaussian mixture model decoder, were used to capture the coarse structure of a sketch by predicting the position of the box including one part of the whole sketch. The part sketcher network then took the predicted box locations and generated the final sketch using a standard GAN architecture. The network was trained on a dataset [43] that contains birds and creatures.

4.3.1 SKETCH-TO-3D SHAPE GENERATION

There are mainly two paradigms for 3D shape reconstruction from 2D sketches: the geometric-based method and the learning-based method. Sketch-based interfaces for modeling (SBIM) is the major branch of geometric-based methods [44] and we do not review this line of work in light of the review scope. We also excluded some methods that apply deep learning techniques but require predefined geometric models to guide 3D reconstruction, such as the methods presented in [45, 46]. We focus on reviewing the end-to-end deep learning-based methods.

Sketch-to-3D shape generation without any predefined geometric models was initialized by Lun et al. [47]. They proposed an encoder-multiview-decoder architecture that can extract the depth and normal maps from a single sketch or multiple sketches, and output a 3D shape in point clouds. The resulting point-cloud representation can be converted to a polygon mesh for better visualization. Similarly, Nozawa et al. [48] extracted the depth and mask information from a single input sketch by an encoder-decoder network. Then a lazy learning [49] method was performed to find similar samples in the dataset to synthesize a 3D shape represented by point clouds. Later, Nozawa et al. [50] extended that method by changing the architecture with a combination of GAN and lazy learning. Delanoy et al. [51] proposed an interactive sketch-to-3D generations system. They used a CNN to transform sketches to 3D voxel shapes, and another CNN as an updater to update the predicted 3D shape while the user is providing more sketches. Su et al. [52] applied an encoder-decoder network to predict normal maps from sketches and optional input of user-specified normal samples. Those methods can only deal with 3D shape generation from sketches within a specific category. To improve generality, Li et al. [53] introduced an intermediate CNN layer to model the direction of dense curvature. They also used an additional output confidence map along with the depth and normal maps extracted using CNNs. They trained the network based on a variety of categories (e.g., fishes, birds, and human characters), which can then output 3D mesh shapes. Using meshes as the 3D representation for deep learning methods is challenging, but the quality of the resulting 3D shapes is better than that of using point clouds and voxels.

The methods introduced above have to be trained using supervised learning, which means that the training data must be pairs of sketches and 3D shapes (i.e., labeled data). Wang et al. [54] proposed an unsupervised learning method for sketch-to-3D shape reconstruction. They embedded unpaired sketches and rendered images from 3D shapes to a common latent space by training an adaption network via autoencoder with adversarial loss. During the inference of 3D shapes from sketches, they retrieved several nearest-neighboring 3D shapes from the training dataset as prior knowledge for a 3D GAN to generate new 3D shapes that best match the input sketch. This method can only output very coarse 3D voxel shapes but provides an interesting idea based on unsupervised learning for sketch-to-3D shape generation.

Besides the usage of general 3D shape representations (i.e., point clouds, voxels, and meshes), there are other 3D representations used in this application. For example, Smirnov et al. [5] proposed a novel deformable parametric template composed of Coon patches that can be easily fitted into a conventional CAD modeling pipeline. The resulting 3D shapes can be easily converted to NURBS representation, allowing edits in CAD software. Guillard et al. [6] proposed a pipeline for reconstructing and editing 3D shapes in DeepSDF [55] format from 2D sketches using an encoder-decoder architecture, which can output high-quality mesh shapes.

5 DISCUSSION

In this section, we start our discussion on 3D representations in DLCMT methods and their ability of generalization. We then discuss how DLCMT can facilitate human-centered design and how to apply DLCMT methods in engineering conceptual design. Finally, we summarize the challenges and propose research questions to be answered in this field.

5.1 3D REPRESENTATIONS IN DLCMT METHODS

3D shapes with high visual quality and rich geometric details can help designers better understand a design concept. Voxel grids, point clouds, and meshes are the most commonly used representations for 3D geometry. Similar to the pixels of images, voxel grids are naturally adapted to convolutional neural network (CNN) model, which is the major reason for its prevalence in 3D geometry learning research. The majority of the studies we reviewed uses voxels for 3D shape representation [9,11,14,23,24,26,28,51,54]. Voxel shapes are usually needed to

TABLE 2. Comparison of pros and cons of the three representations to deep learning methods

3D Representation	Pros	Cons
Voxel grids	• The data structure in fourth-order tensor makes it easy to be adapted in 3D convolution operations	• Low visual quality • High computational cost because the number of the 3D representation parameters scale with the increase of spatial resolution in cubes • Cannot be directly used in engineering analyses (e.g., the FEA analyses) for performance evaluation
Point clouds	• Compatible with the output data format of common scanning software • Compact for data storage and management	• Low visual quality • No detailed geometric information about relationships between points making it hard to convert to meshes • Cannot be directly used in engineering analyses (e.g., the FEA analyses) for performance evaluation
Meshes	• High visual quality • Compact for data storage and management • Widely-accepted 3D representation in computer graphics • Compatible with downstream engineering software requirements such as the FEA and CFD tools	• Discrete and disordered elements make it challenging to be processed by deep learning methods

converted to mesh shapes for better visualization. However, the transformed mesh shapes will look coarse if the resolution of the voxel shapes is low. This could negatively influence the subjective evaluation of the shape of a design concept, and the design concept might be overlooked by designers. An intuitive way to improve the resolution of the resulting 3D voxel shapes is to use high-resolution training data, but this may not be feasible due to the limited computing resources for training the neural network. Fukamizu et al. [24] provided a two-stage strategy to synthesize high resolution 3D voxel shapes from natural language, which could be an inspiring method for dealing with low resolution issues. Point clouds [12, 13, 47, 48, 50] are more efficient in representing 3D objects, but do not cover geometric details. For example, it does not encode the relationship between points and the resulting topology of an object, leading to a challenging conversion to meshes. Using meshes [53] for 3D representation could generally alleviate the low visual quality and data storage problems, but, in the meantime, it is challenging to prepare meshes for deep learning methods due to their discrete face structures and unordered elements. Please see Table 2 for the pros and cons of applying those three representations to deep learning methods.

In addition to the above three representations, there are a few new 3D representations that are promising to handle the trade-off between the effectiveness of training neural networks and the quality of the resulting 3D shapes. Neural Radiance Field (NeRF) [31, 32] is a method for generating novel views of scenes or objects. NeRF can take a set of input images of an object and render the complete object by interpolating between the images. NeRF is topology-free and can be sampled at high spatial resolutions. However, 3D shapes represented by NeRF are "hidden in the black box" and we can only observe it by images rendered from different viewpoints. The deep implicit field [6, 55] represents a shape's surface by a continuous volumetric field, encoding a shape's boundary as the zero-level set of the learned implicit 3D shape function. It is a promising representation for high-resolution 3D shapes, but requires fewer data storage. However, all the 3D representations mentioned above (i.e., voxels, point clouds, meshes, NeRF, and deep implicit field) are generally not adapted to CAD software. This often brings about compatibility issues that could impede downstream editing and engineering analyses of the generated 3D shapes. To solve these issues, there are typically two ways. One way is to convert them to CAD models (e.g., converting STL/OBJ meshes to B-Rep solids). Another way is to handle the CAD shape data directly in deep learning models. CAD datasets [22, 56–58] and deep learning methods based on uni-modal CAD data [59–63] have recently been introduced. Deep learning of uni-modal CAD data is still an underexplored field, and DLCMT using such a data format [5] turns out to be a promising research direction.

5.2 GENERALIZATION OF DLCMT METHODS

In engineering design applications, a deep learning method can be effective in one design object but may not be easily applied to other design objects, because many deep learning methods are developed based on available training data in one design domain. It could be challenging for deep learning methods to generalize to multiple design problems. Seeking a generalization of applying DLCMT in engineering design could be even challenging due to the unavailability of data pairs between 3D shapes and their corresponding text descriptions or sketches. Some methods utilize transfer learning techniques (e.g. zero shot learning) [28, 30, 31] or specially designed neural network architectures to improve generalization, which could be good references for the design community.

5.3 INCORPORATING HUMAN INPUTS INTO DEEP LEARNING-AIDED DESIGN PROCESSES

In general, deep learning methods (e.g., VAEs and GANs) could generate new data that are not seen in the training data set, but are still based on interpolation within the boundary of the training data. Therefore, the new designs generated still share great similarities with the existing ones. Recently, efforts have been spent on developing new network architectures to enable deep-learning models' extrapolating capabilities, and thus can generate true novel designs [64, 65]. However, despite advances in model development, one observation is that human input and interaction are not much emphasized in the deep learning-aided design process [3]. Burnap et al. [66] pointed out that a human's perception of the quality of the generated design concepts is often not in correspondence with their numerical performance measures. The reason could be that in most deep learning-aided design processes, designers can only passively select the preferred design concepts from a set of computer-generated design options. There is a need to actively involve designers and/or users in the data-driven design process [3, 7]. There have been some efforts on this recently. For example, the method shown in [67] allows users to manipulate the latent space vectors learned by a GAN model to create preferred design options. We argue that a more intuitive and natural way to actively interact with humans is through natural language or sketches, and DLCMT could play a role in this regard in transferring one modality to another (e.g., text to 3D shape and sketch to 3D shapes). As computers can propose novel designs beyond the human's imagination, designers could learn from computers. On the other hand, designers can continuously supplement new design ideas during the interaction to guide computers to generate creative and feasible design concepts that meet the requirements of aesthetics, functionality, and manufacturability.

5.4 APPLYING DLCMT TO ENGINEERING DESIGN

The majority of the DLCMT literature is from computer graphics or computer vision communities, with only one sketch-to-3D retrieval work focusing on design research [22]. Similarly, there are only a few studies (e.g., [68–71]) on unimodal 3D shape synthesis from the engineering design community. Regenwetter et al. [3] state that 3D synthesis works are less relevant to engineering design because they focus more on visual appearance, instead of functional performance or manufacturability. We agree that engineering performance and manufacturability should be considered in engineering design research. But, we argue that 3D shape synthesis methods could be beneficial to the product shape design in conceptual design and should receive more attentions from the design community.

On the other hand, focusing on design synthesis or retrieval, DLCMT could be applied to engineered product shape design, but they still have a less emphasis on engineering performance of the product. For example, using text-to-3D shape methods, a user said that "I want an SUV with low fuel consumption". The method could generate an SUV car shape for the user, but no engineering performance is considered. We might ask a question: how can we make the computer understand that text description and translate it to a primitive SUV car shape with account for the engineering performance, such as drag coefficient? Therefore, how to integrate engineering performance evaluation, e.g. surrogate modeling, into DLCMT to enable shape design taking into account the engineering performance of the product could be an interesting research direction. Similarly, it is also worth exploring how other aspects (e.g., manufacturability) of engineering design can be counted when applying DLCMT to engineering design. To facilitate such applications, we encourage proposals of cross-modal datasets with corresponding engineering performance and manufacturability analysis of the designs (e.g., sketch-to-3D shape dataset with engineering performance of the 3D shapes).

5.5 CHALLENGES AND RESEARCH QUESTIONS

As discussed above, our review of the current state-of-the-art DLCMT reveals opportunities to use these methods for human-centered product shape design in the conceptual design stage. In addition, these methods have the potential to accelerate the democratization of the product design process, which allows ordinary people to get involved in personalized design [72]. This could also be a good opportunity to develop education tools for design education and training junior design engineers.

Challenges coexist with opportunities. First, there is a deficiency of multi-modality data of engineered products. Data is the fuel for deep learning-based design methods. We encourage the sharing and publishing of 3D shape datasets of engineered products paired with the data collected from customer needs analysis, engineering performance, and manufacturability. Such datasets could greatly promote the verification and validation of existing DLCMT methods and the development of new DLCMT methods for engineering design. Second, choosing the most appropriate 3D representation compatible with the adopted deep learning technique is still a challenging task. It involves considerations of data availability, data preprocessing, computational cost, visual quality of the resulting 3D shapes, data postprocessing, and the ability to adapt to later design stages. Third, integrating DLCMT with engineering performance evaluation and the analysis of manufacturability of the resulting designs poses new challenges. Lastly, the challenge in generalizing DLCMT methods couples with other challenges, and requires a community-wide effort to share data sets, create data repositories, define benchmark problems, and develop testing standards. Based on these challenges, we propose the following research questions (RQs) pointing to future research directions, in the hope of arousing a greater discussion within the engineering design community.

RQ1: DLCMT has the potential to facilitate human-centered design, but what are the possible ways to maximize human involvement and stimulate their design creativity at the human-AI interface?
RQ2: With the establishment of human-AI interaction in conceptual design based on DLCMT, how could the co-evolution between human and AI look like?
RQ3: Since DLCMT can shorten the cycle of generating 3D shapes and even connect to the downstream engineering analyses and manufacturing requirements, then how could the information coming from the later design stages influence the regeneration of 3D shape concepts, and thereby a designer's decisions?
RQ4: What are the guidelines for selecting the most appropriate data representations in DLCMT?
RQ5: To what extent can the generalizability and transferability of the latent representation of multi-modality data learned from DLCMT be across different product shape categories?

6 CLOSING REMARKS

In this paper, we reviewed deep learning of cross-modal tasks (DLCMT), including text-to-sketch, text-to-3D shape, and sketch-to-3D shape retrieval and generation methods, for the conceptual design of product shapes. Those methods could be applied in the Design Search and Design Creation steps of the conceptual design. Different from other deep learning methods applied in engineering design, DLCMT allows human inputs of texts and sketches, which reflect designers' and/or user's preferences. As designers can be more actively involved in such a design process, human-computer interaction and collaboration are promoted, thereby it has a great potential to improve human-centered conceptual design of products compared to traditional design automation methods and computer-aided design methods. DLCMT could also facilitate the engineering design education and democratization of product development by allowing intuitive inputs (e.g., texts descriptions and sketches).

With the attempt of applying new 3D data representations in DLCMT and the availability of more public datasets, opportunities open up for the development of new DLCMT methods. However, the deficiency of training datasets, trade-off in the choice of representations of 3D shapes, lack of consideration of engineering performance and manufacturability, and the ability of generalization still challenge the design community to apply DLCMT to engineering design. We would like to encourage attentions and efforts from the design community.

In our future work, we will continue the review and conduct a more comprehensive analysis of the relevant work on DLCMT. We hope that this review effort could facilitate the discussion and attract more attention of design researchers in utilizing DLCMT to improve the human-centered conceptual design of product shapes and beyond.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support from the National Science Foundation through the award 2207408.

REFERENCES

[1] Ulrich, K. T., 2003. *Product design and development*. Tata McGraw-Hill Education.
[2] Chakrabarti, A., Shea, K., Stone, R., Cagan, J., Campbell, M., Hernandez, N. V., and Wood, K. L., 2011. "Computer-based design synthesis research: an overview". *Journal of Computing and Information Science in Engineering,* ***11***(2).
[3] Regenwetter, L., Nobari, A. H., and Ahmed, F., 2021. "Deep generative models in engineering design: A review". *arXiv preprint arXiv:2110.10863*.
[4] Liu, Z., Lin, Y., and Sun, M., 2020. *Cross-Modal Representation*. Springer Singapore, Singapore, pp. 285–317.
[5] Smirnov, D., Bessmeltsev, M., and Solomon, J., 2019. "Learning manifold patch-based representations of man-made shapes". *arXiv preprint arXiv:1906.12337*.
[6] Guillard, B., Remelli, E., Yvernay, P., and Fua, P., 2021. "Sketch2mesh: Reconstructing and editing 3d shapes from sketches". *arXiv preprint arXiv:2104.00482*.
[7] Mountstephens, J., and Teo, J., 2020. "Progress and challenges in generative product design: A review of systems". *Computers,* ***9***(4), p. 80.
[8] Menezes, A., and Lawson, B., 2006. "How designers perceive sketches". *Design studies,* ***27***(5), pp. 571–585.
[9] Chen, K., Choy, C. B., Savva, M., Chang, A. X., Funkhouser, T., and Savarese, S., 2018. "Text2shape: Generating shapes from natural language by learning joint embeddings". In Asian conference on computer vision, Springer, pp. 100–116.
[10] Chang, A. X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., and Su, H., 2015. "Shapenet: An information-rich 3d model repository". *arXiv preprint arXiv:1512.03012*.
[11] Han, Z., Shang, M., Wang, X., Liu, Y.-S., and Zwicker, M., 2019. "Y2seq2seq: Cross-modal representation learning for 3d shape and text by joint reconstruction and prediction of view and word sequences". In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, pp. 126–133.
[12] Tang, C., Yang, X., Wu, B., Han, Z., and Chang, Y., 2021. "Part2word: Learning joint embedding of point clouds and text by matching parts to words". *arXiv preprint arXiv:2107.01872*.
[13] Qi, C. R., Su, H., Mo, K., and Guibas, L. J., 2017. "Pointnet: Deep learning on point sets for 3d classification and segmentation". In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 652–660.
[14] Ruan, Y., Lee, H.-H., Zhang, K., and Chang, A. X., 2022.

"Tricolo: Trimodal contrastive loss for fine-grained text to shape retrieval". *arXiv preprint arXiv:2201.07366*.
[15] Wang, F., Kang, L., and Li, Y., 2015. "Sketch-based 3d shape retrieval using convolutional neural networks". In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1875–1883.
[16] Zhu, F., Xie, J., and Fang, Y., 2016. "Learning cross-domain neural networks for sketch-based 3d shape retrieval". In Proceedings of the AAAI conference on artificial intelligence, Vol. 30.
[17] Dai, G., Xie, J., and Fang, Y., 2018. "Deep correlated holistic metric learning for sketch-based 3d shape retrieval". *IEEE Transactions on Image Processing,* ***27***(7), pp. 3374–3386.
[18] Kaya, M., and Bilge, H. Ş., 2019. "Deep metric learning: A survey". *Symmetry,* ***11***(9), p. 1066.
[19] Xie, J., Dai, G., Zhu, F., and Fang, Y., 2017. "Learning barycentric representations of 3d shapes for sketch-based 3d shape retrieval". In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5068–5076.
[20] Li, B., Lu, Y., Li, C., Godil, A., Schreck, T., Aono, M., Burtscher, M., Fu, H., Furuya, T., Johan, H., et al., 2014. "Shrec'14 track: Extended large scale sketch-based 3d shape retrieval". In Eurographics workshop on 3D object retrieval, Vol. 2014, pp. 121–130.
[21] Qi, A., Gryaditskaya, Y., Song, J., Yang, Y., Qi, Y., Hospedales, T. M., Xiang, T., and Song, Y.-Z., 2021. "Toward fine-grained sketch-based 3d shape retrieval". *IEEE Transactions on Image Processing,* ***30***, pp. 8595–8606.
[22] Qin, F., Qiu, S., Gao, S., and Bai, J., 2022. "3d cad model retrieval based on sketch and unsupervised variational autoencoder". *Advanced Engineering Informatics,* ***51***, p. 101427.
[23] Arjovsky, M., Chintala, S., and Bottou, L., 2017. "Wasserstein generative adversarial networks". In International conference on machine learning, PMLR, pp. 214–223.
[24] Fukamizu, K., Kondo, M., and Sakamoto, R., 2019. "Generation high resolution 3d model from natural language by generative adversarial network". *arXiv preprint arXiv:1901.07165*.
[25] Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., and Metaxas, D. N., 2017. "Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks". In Proceedings of the IEEE international conference on computer vision, pp. 5907–5915.
[26] Li, B., Yu, Y., and Li, Y., 2020. "Lbwgan: Label based shape synthesis from text with wgans". In 2020 International Conference on Virtual Reality and Visualization (ICVRV), IEEE, pp. 47–52.
[27] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al., 2021. "Learning transferable visual models from natural language supervision". In International Conference on Machine Learning, PMLR, pp. 8748–8763.
[28] Sanghi, A., Chu, H., Lambourne, J. G., Wang, Y., Cheng, C.-Y., and Fumero, M., 2021. "Clip-forge: Towards zero-shot text-to-shape generation". *arXiv preprint arXiv:2110.02624*.
[29] Dinh, L., Sohl-Dickstein, J., and Bengio, S., 2016. "Density estimation using real nvp". *arXiv preprint arXiv:1605.08803*.
[30] Chu, E., 2021. "Evolving evocative 2d views of generated 3d objects". *arXiv preprint arXiv:2111.04839*.
[31] Jain, A., Mildenhall, B., Barron, J. T., Abbeel, P., and Poole, B., 2021. "Zero-shot text-guided object generation with dream fields". *arXiv preprint arXiv:2112.01455*.
[32] Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R., 2020. "Nerf: Representing scenes as neural radiance fields for view synthesis". In European conference on computer vision, Springer, pp. 405–421.
[33] Frolov, S., Hinz, T., Raue, F., Hees, J., and Dengel, A., 2021. "Adversarial text-to-image synthesis: A review". *Neural Networks,* ***144***, pp. 187–209.
[34] Yuan, S., Dai, A., Yan, Z., Guo, Z., Liu, R., and Chen, M., 2021. "Sketchbird: Learning to generate bird sketches from text". In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2443–2452.
[35] Wang, Y., Chang, L., Cheng, Y., Jin, L., Cheng, Z., Deng, X., and Duan, F., 2018. "Text2sketch: Learning face sketch from facial attribute text". In 2018 25th IEEE International Conference on Image Processing (ICIP), IEEE, pp. 669–673.
[36] Wah, C., Branson, S., Welinder, P., Perona, P., and Belongie, S., 2011. "The caltech-ucsd birds-200-2011 dataset".
[37] Huang, F., and Canny, J. F., 2019. "Sketchforme: Composing sketched scenes from text descriptions for interactive applications". In Proceedings of the 32nd annual ACM symposium on user interface software and technology, pp. 209–220.
[38] Ha, D., and Eck, D., 2017. "A neural representation of sketch drawings". *arXiv preprint arXiv:1704.03477*.
[39] Krishna, R., Zhu, Y., Groth, O., Johnson, J., Hata, K., Kravitz, J., Chen, S., Kalantidis, Y., Li, L.-J., Shamma, D. A., et al., 2017. "Visual genome: Connecting language and vision using crowdsourced dense image annotations". *International journal of computer vision,* ***123***(1), pp. 32–73.
[40] Jongejan, J., Rowley, H., Kawashima, T., Kim, J., and Fox-Gieg, N., 2016. "The quick, draw!-ai experiment". *Mount View, CA, accessed Feb,* ***17***(2018), p. 4.
[41] Huang, F., Schoop, E., Ha, D., and Canny, J., 2020.

"Scones: towards conversational authoring of sketches". In Proceedings of the 25th International Conference on Intelligent User Interfaces, pp. 313–323.
[42] Bhunia, A. K., Khan, S., Cholakkal, H., Anwer, R. M., Khan, F. S., Laaksonen, J., and Felsberg, M., 2021. "Doodleformer: Creative sketch drawing with transformers". *arXiv preprint arXiv:2112.03258*.
[43] Ge, S., Goswami, V., Zitnick, C. L., and Parikh, D., 2020. "Creative sketch generation". *arXiv preprint arXiv:2011.10039*.
[44] Olsen, L., Samavati, F. F., Sousa, M. C., and Jorge, J. A., 2009. "Sketch-based modeling: A survey". *Computers & Graphics,* ***33***(1), pp. 85–103.
[45] Nishida, G., Garcia-Dorado, I., Aliaga, D. G., Benes, B., and Bousseau, A., 2016. "Interactive sketching of urban procedural models". *ACM Transactions on Graphics (TOG),* ***35***(4), pp. 1–11.
[46] Han, X., Gao, C., and Yu, Y., 2017. "Deepsketch2face: A deep learning based sketching system for 3d face and caricature modeling". *ACM Transactions on graphics (TOG),* ***36***(4), pp. 1–12.
[47] Lun, Z., Gadelha, M., Kalogerakis, E., Maji, S., and Wang, R., 2017. "3d shape reconstruction from sketches via multi-view convolutional networks". In 2017 International Conference on 3D Vision (3DV), IEEE, pp. 67–77.
[48] Nozawa, N., Shum, H. P., Ho, E. S., and Morishima, S., 2020. "Single sketch image based 3d car shape reconstruction with deep learning and lazy learning.". In VISIGRAPP (1: GRAPP), pp. 179–190.
[49] Aha, D. W., 2013. *Lazy learning*. Springer Science & Business Media.
[50] Nozawa, N., Shum, H. P., Feng, Q., Ho, E. S., and Morishima, S., 2021. "3d car shape reconstruction from a contour sketch using gan and lazy learning". *The Visual Computer*, pp. 1–14.
[51] Delanoy, J., Aubry, M., Isola, P., Efros, A. A., and Bousseau, A., 2018. "3d sketching using multi-view deep volumetric prediction". *Proceedings of the ACM on Computer Graphics and Interactive Techniques,* ***1***(1), pp. 1–22.
[52] Su, W., Du, D., Yang, X., Zhou, S., and Fu, H., 2018. "Interactive sketch-based normal map generation with deep neural networks". *Proceedings of the ACM on Computer Graphics and Interactive Techniques,* ***1***(1), pp. 1–17.
[53] Li, C., Pan, H., Liu, Y., Tong, X., Sheffer, A., and Wang, W., 2018. "Robust flow-guided neural prediction for sketch-based freeform surface modeling". *ACM Transactions on Graphics (TOG),* ***37***(6), pp. 1–12.
[54] Wang, L., Qian, C., Wang, J., and Fang, Y., 2018. "Unsupervised learning of 3d model reconstruction from hand-drawn sketches". In Proceedings of the 26th ACM international conference on Multimedia, pp. 1820–1828.
[55] Park, J. J., Florence, P., Straub, J., Newcombe, R., and Lovegrove, S., 2019. "Deepsdf: Learning continuous signed distance functions for shape representation". In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 165–174.
[56] Willis, K. D., Pu, Y., Luo, J., Chu, H., Du, T., Lambourne, J. G., Solar-Lezama, A., and Matusik, W., 2021. "Fusion 360 gallery: A dataset and environment for programmatic cad construction from human design sequences". *ACM Transactions on Graphics (TOG),* ***40***(4), pp. 1–24.
[57] Regenwetter, L., Curry, B., and Ahmed, F., 2021. "Biked: A dataset and machine learning benchmarks for data-driven bicycle design". In International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Vol. 85383, American Society of Mechanical Engineers, p. V03AT03A019.
[58] Manda, B., Dhayarkar, S., Mitheran, S., Viekash, V., and Muthuganapathy, R., 2021. "'cadsketchnet'-an annotated sketch dataset for 3d cad model retrieval with deep neural networks". *Computers & Graphics,* ***99***, pp. 100–113.
[59] Wu, R., Xiao, C., and Zheng, C., 2021. "Deepcad: A deep generative network for computer-aided design models". In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 6772–6782.
[60] Para, W., Bhat, S., Guerrero, P., Kelly, T., Mitra, N., Guibas, L. J., and Wonka, P., 2021. "Sketchgen: Generating constrained cad sketches". *Advances in Neural Information Processing Systems,* ***34***.
[61] Ganin, Y., Bartunov, S., Li, Y., Keller, E., and Saliceti, S., 2021. "Computer-aided design as language". *Advances in Neural Information Processing Systems,* ***34***.
[62] Willis, K. D., Jayaraman, P. K., Lambourne, J. G., Chu, H., and Pu, Y., 2021. "Engineering sketch generation for computer-aided design". In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2105–2114.
[63] Jayaraman, P. K., Sanghi, A., Lambourne, J. G., Willis, K. D., Davies, T., Shayani, H., and Morris, N., 2021. "Uv-net: Learning from boundary representations". In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11703–11712.
[64] Elgammal, A., Liu, B., Elhoseiny, M., and Mazzone, M., 2017. "Can: Creative adversarial networks, generating" art" by learning about styles and deviating from style norms". *arXiv preprint arXiv:1706.07068*.
[65] Chen, W., and Ahmed, F., 2021. "Padgan: Learning to generate high-quality novel designs". *Journal of Mechanical Design,* ***143***(3).
[66] Burnap, A., Liu, Y., Pan, Y., Lee, H., Gonzalez, R., and Papalambros, P. Y., 2016. "Estimating and exploring the product form design space using deep generative models". In International Design Engineering Technical Conferences and Computers and Information in Engineering Confer-

ence, Vol. 50107, American Society of Mechanical Engineers, p. V02AT03A013.

[67] Valdez, S., Seepersad, C., and Kambampati, S., 2021. “A framework for interactive structural design exploration”. In International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Vol. 85390, American Society of Mechanical Engineers, p. V03BT03A006.

[68] Shu, D., Cunningham, J., Stump, G., Miller, S. W., Yukish, M. A., Simpson, T. W., and Tucker, C. S., 2020. “3d design using generative adversarial networks and physics-based validation”. *Journal of Mechanical Design,* ***142***(7).

[69] Zhang, W., Yang, Z., Jiang, H., Nigam, S., Yamakawa, S., Furuhata, T., Shimada, K., and Kara, L. B., 2019. “3d shape synthesis for conceptual design and optimization using variational autoencoders”. In International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Vol. 59186, American Society of Mechanical Engineers, p. V02AT03A017.

[70] Li, X., Xie, C., and Sha, Z., 2021. “Part-aware product design agent using deep generative network and local linear embedding”. In Proceedings of the 54th Hawaii International Conference on System Sciences, p. 5250.

[71] Brock, A., Lim, T., Ritchie, J. M., and Weston, N., 2016. “Context-aware content generation for virtual environments”. In International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Vol. 50084, American Society of Mechanical Engineers, p. V01BT02A045.

[72] Starly, B., Angrish, A., and Cohen, P., 2019. “Research directions in democratizing innovation through design automation, one-click manufacturing services and intelligent machines”. *arXiv preprint arXiv:1909.10476*.