CSE499A Section 10

Group 3

Md. Rakib Hasan Bhuiyan

Ashikur Rahman Ashik

Md. Rafid Hasan

Masud Shahriar

Project Overview

**CAD Generation from Text Prompts using Multi-modal LLM**

# Introduction

Computer-Aided Design (CAD) is essential in engineering, manufacturing, and architecture, but traditional modeling requires expertise in complex software. Recent advancements in Multimodal Large Language Models (MLLMs) enable automated 2D and 3D CAD generation from text prompts, bridging the gap between human intent and design execution.

This project introduces a hybrid procedural + neural approach, where MLLMs generate parametric CAD scripts (e.g., OpenSCAD, Fusion 360 API) instead of raw meshes. It also features interactive text-based refinement, constraint extraction from sketches, and reinforcement learning for error correction, making CAD generation more precise, editable, and aligned with real-world workflows.

# Datasets

[ABC Dataset](https://deep-geometry.github.io/abc-dataset/) (A Big CAD Dataset)

Contains 1 million parametric CAD models.

Format: STEP, STL

[ShapeNet](https://shapenet.org/)

Large-scale repository of 3D objects with text annotations.

Format: OBJ, STL, PLY

[Fusion 360 Gallery Dataset](https://fusion360gallery.autodesk.com/)

Parametric CAD models with design history and sketches.

Format: Fusion 360 Native, STEP, IGES

[Text-to-Shape Dataset](https://github.com/autonomousvision/text2shape)

A dataset mapping text descriptions to 3D shapes.

Format: JSON + 3D Meshes

[SketchGraphs Dataset](https://github.com/PrincetonLIPS/SketchGraphs) (Autodesk)

Contains 15M parametric sketches with constraints and operations.

Format: Graph representation of CAD sketches

# Models

Multimodal Large Language Model (MLLM)

* LLaVA (Large Language and Vision Assistant) – A fine-tuned LLaMA-2 model with vision capabilities.
* GPT-4V (Vision-enabled GPT-4) – Can process text and images for CAD generation.
* FLAN-T5 – Strong in text-to-action sequence generation for CAD commands.

CAD Generation Models

* ControlNet (Stable Diffusion + Sketches) – For text-to-2D-sketch generation.
* DreamFusion / DeepMarch – 3D model generation from text prompts.
* GeoGPT (3D-GPT) – Specialized LLM for procedural 3D modeling.

Geometric Deep Learning Models

* PointNet++ – Processes point clouds for refining CAD structures.
* MeshCNN – Works on mesh-based representations for complex CAD designs.

Constraint-based Parametric Models

* Program Synthesis with LLMs – Generates CAD scripts (e.g., OpenSCAD, Fusion 360 API).

# Pipeline for Development

Step 1: Data Preparation

* Collect CAD files, text descriptions, and sketches from the datasets.
* Convert CAD files into graph-based representations or latent embeddings.
* Preprocess sketches as edge maps for text-to-sketch models.

Step 2: Text to 2D Sketch Generation

* Fine-tune Stable Diffusion (ControlNet) on sketch datasets.
* Convert sketches into parametric constraints using ML models.

Step 3: 2D Sketch to 3D CAD Model

* Train a model (e.g., SketchGraphs Transformer) to infer constraints.
* Use Graph Neural Networks (GNNs) to reconstruct 3D parametric models.
* Apply ShapeGPT / 3D-GPT to refine and correct errors.

Step 4: End-to-End Text-to-3D CAD Generation

* Train LLM-powered CAD scripting models to generate Fusion 360 / OpenSCAD commands.
* Use procedural generation (CAD API calls) to execute scripts.
* Evaluate physical feasibility using constraints (e.g., Fusion 360 API).

Step 5: Post-processing & Optimization

* Use MeshCNN or PointNet++ to refine rough CAD models.
* Apply Neural Style Transfer for CAD aesthetics.
* Convert models to various formats (STEP, STL, OBJ).

# Related papers

[Md. Rakib Hasan Bhuiyan]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Paper | Dataset | Model | Results | Novelty |
| 1. [CAD-MLLM](https://web3.arxiv.org/pdf/2411.04954) | Omni-CAD dataset: Self-made with 453,220 augmented CAD command sequence data | Combination of a vision encoder, point encoder, and pre-trained Vicuna-7B LLM (fine-tuned using LoRA). | For image reconstruction:-  Chamfer Distance↓: 3.22  F-Score↑: 80.82  Normal Consistency↑: 62.07  Segment Error↓: 1.56  Dangling Edge Length↓: 0.51  Self-Intersection Ratio↓: 1.36  Flux Enclosure Error↓: 0.050 | Outperforms baseline methods (DeepCAD, Text2CAD, Img2CAD, etc.) in these metrics, demonstrating robustness against noise and missing data |
| 2. [Query2CAD](https://arxiv.org/pdf/2406.00144) | Self-made dataset with 57 queries ranging from easy (21) to medium (20) to hard (16). | GPT-4 Turbo and GPT-3.5 Turbo were used as the main LLM.  Uses BLIP2 as a caption model for feedback-based refinement.  Incorporates a self-refinement loop, using (VQAScore) to iteratively improve generated CAD designs. | Success rate:-  GPT-4 Turbo:  Easy queries: 95.23%  Medium queries: 70%  Hard queries: 41%  GPT-3.5 Turbo:  Easy queries: 85.71%  Medium queries: 35%  Hard queries: 37.5%  Success rate after fourth refinement step y3:-  GPT-3.5 Turbo: 53.4%  GPT-4 Turbo: 76.7% | The refinement loop improves the initially generated models to be more accurate. |
| 3. [ShapeGPT](https://arxiv.org/pdf/2311.17618) | ShapeNet: 16 object categories encompassing 50,000 models,  Text2Shape dataset for detailed text annotations. | Proposed new model ShapeGPT with T5 as language model and  clip-vit-large-patch14 model to extract features from images. | For Text to Shape:-  Intersection over Union↑: 0.587  Chamfer Distance↓: 1.256  F-Score↑: 0.402  ULIP↑: 0.189 | State-of-the-art results in generating complex 3D models from text and images. |

[Ashikur Rahman Ashik]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Paper | Dataset | Model | Results | Novelty |
| 3. [Text2CAD](https://arxiv.org/pdf/2411.06206) | New dataset that uses FreeCAD and includes 100,000 CAD models in the ".step" format, taken from the ABC dataset | Stable Diffusion v1-5: Fine-tuned to generate isometric CAD images from text prompts.  Zero-1-to-3 Model: Used for novel view generation, transforming isometric images into orthographic views (top, front, side).  Photo2CAD: A tool used to extract vector paths from orthographic drawings for 3D CAD reconstruction. | Evaluated on a scale of 0 to 10 based on human and GPT-4 reviews. Achieves overall average of 8.375 across 100 samples. | Generates a detailed isometric drawing and transforming it into consistent orthographic views, e.g., top, front, and side. |
| 2. [Text-to-CAD](https://arxiv.org/pdf/2501.19054) | Uses the DeepCAD dataset, containing 20,000 text-CAD parametric sequence pairs.  Additionally, constructs 1,500 preference pairs per iteration | LLaMA-3-8b-Instruct: Used as the backbone LLM for CAD sequence generation.  Large Vision-Language Models (LVMs): Used for evaluating visual accuracy of rendered CAD models.  Direct Preference Optimization (DPO): Optimizes LLMs based on user-preferred visual feedback. | F1 Score↑: 85.22 (Sketch), 92.79 (Extrusion)  Chamfer Distance↓: 45.67  Coverage↑: 90.40  Minimum Matching Distance↓: 3.49  Jensen-Shannon Divergence↓: 17.11  Inception Recall↑: 6.20  Latent Vector Matching Score↑: 8.96 | Leverages both the sequential signal and visual signal to train a Text-to-CAD model. To balance both signals, it alternate between the sequential learning and the visual feedback stage. |