復旦大學
FUDAN UNIVERSITY

生成式AI实践

2025春 计算机图形学
Project 3

CAD概述

什么是CAD？

CAD(Computer Aided Design)是一种利用计算机软件进行设计、绘图、建模和分析的技术。它取代了传统的手工制图方式，通过数字化工具帮助工程师、设计师和建筑师更高效地完成从概念设计到成品优化的全流程工作。

CAD的主要用途有什么？

高效绘图和建模。快速生成精确的二维或三维模型，支持参数化设计，修改设计时自动更新相关部件。

CAD的核心应用产业？

机器制造、建筑与土木工程、汽车与航空航天、医疗领域、消费品设计...

CAD+AI

DeepSeek等大语言模型改变了生成式AI以及计算机图形学的研究范式。从启发式的设计变迁到对话式AIGC，生成的交互性和多样性得到了极大提升。CAD和AI的结合主要基于LLMs生成结构化的CAD数据，融合图片、文本等多种模态的进行融合和创新。

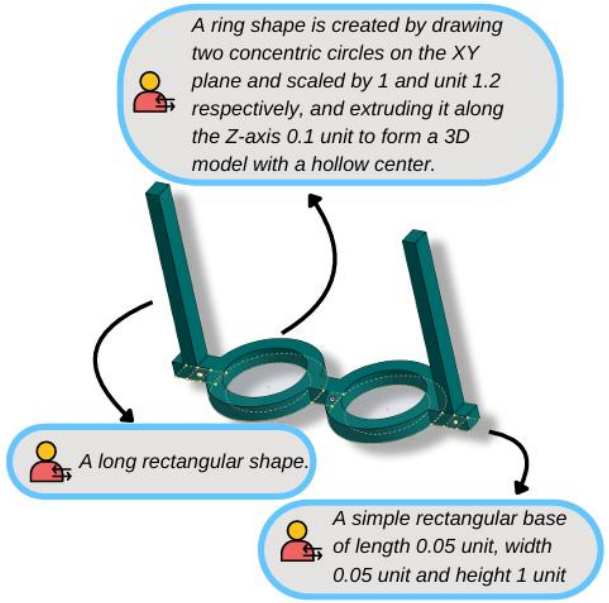


Figure 1: *Designers can efficiently generate parametric CAD models from text prompts. The prompts can vary from abstract shape descriptions to detailed parametric instructions.*

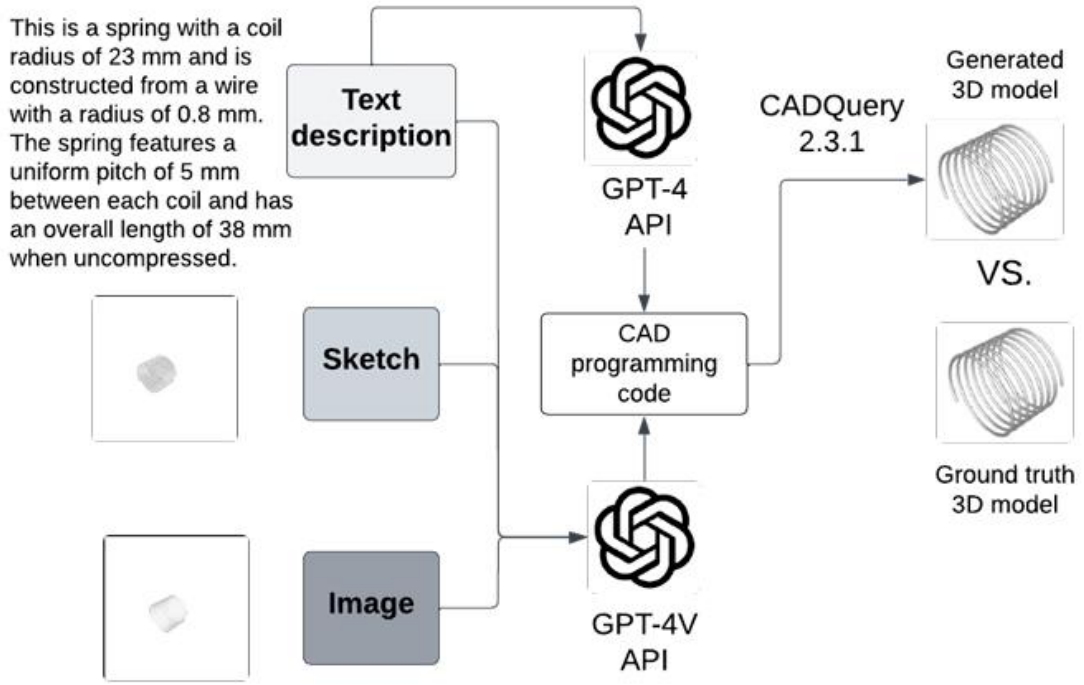


FIGURE 4. The pipeline for Code Generation and Evaluation

文本生成CAD - Text2CAD

传统人工的在软件中进行复杂CAD的设计非常消耗时间，论文提出Text2CAD来生成小的CAD模块，实现了依据文本描述生成CAD数据，用以辅助CAD的设计。论文还提出根据自然语言生成详细的CAD模块描述提示语的流程，基于DeepCAD数据集构建了对应的描述文本数据集。

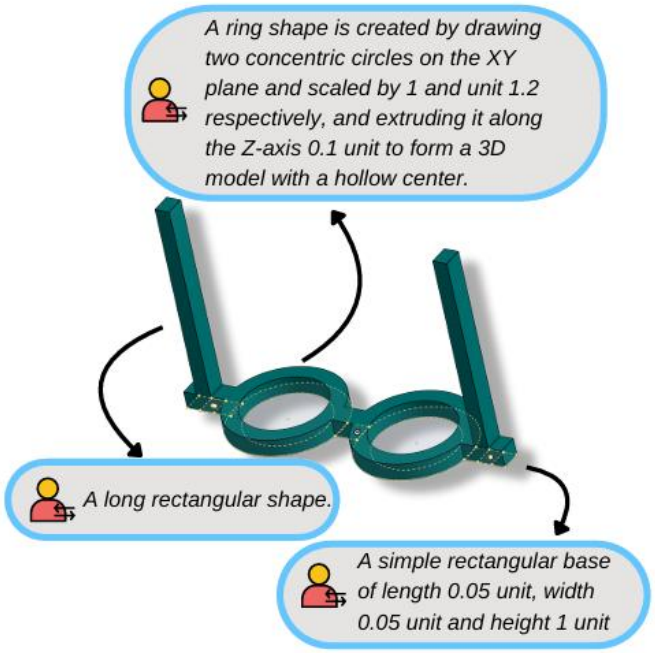


Figure 1: *Designers can efficiently generate parametric CAD models from text prompts. The prompts can vary from abstract shape descriptions to detailed parametric instructions.*

文本生成CAD - Text2CAD

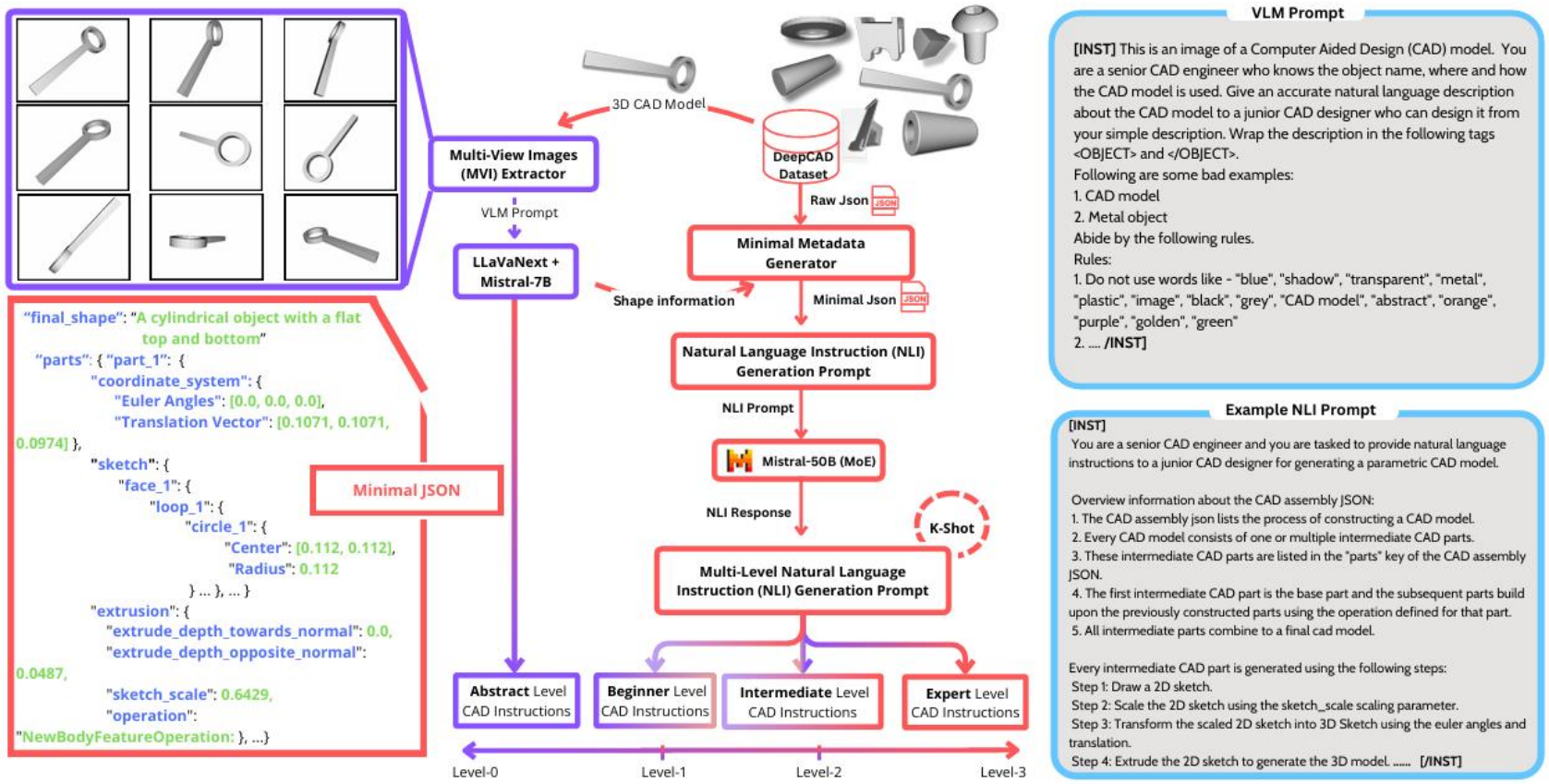


Figure 2: **Text2CAD Data Annotation Pipeline:** Our data annotation pipeline generates multi-level text prompts describing the construction workflow of a CAD model with varying complexities. We use a two-stage method - (Stage 1) Shape description generation using VLM (Stage 2) Multi-Level textual annotation generation using LLM.

文本生成CAD - Text2CAD

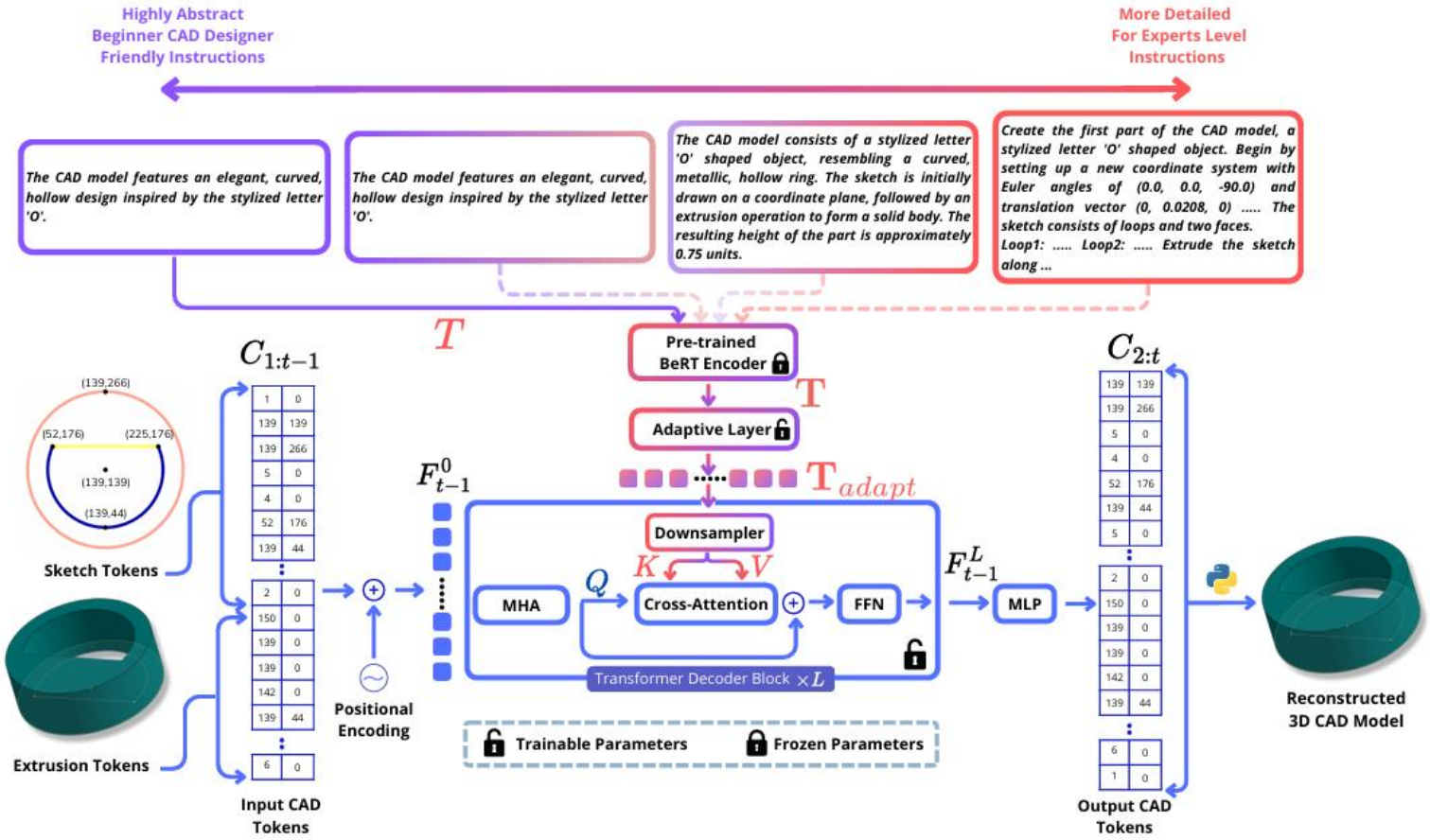


Figure 3: **Network architecture**: Text2CAD Transformer takes as input a text prompt T and a CAD subsequence $\mathbf{C}_{1:t-1}$ of length $t-1$. The text embedding $\mathbf{T}_{adapt}$ is extracted from T using a pretrained BeRT Encoder ([8]) followed by a trainable Adaptive layer. The resulting embedding $\mathbf{T}_{adapt}$ and the CAD sequence embedding $\mathbf{F}^0_{t-1}$ is passed through $\mathbf{L}$ decoder blocks to generate the full CAD sequence in auto-regressive way.

文本生成CAD - Text2CAD

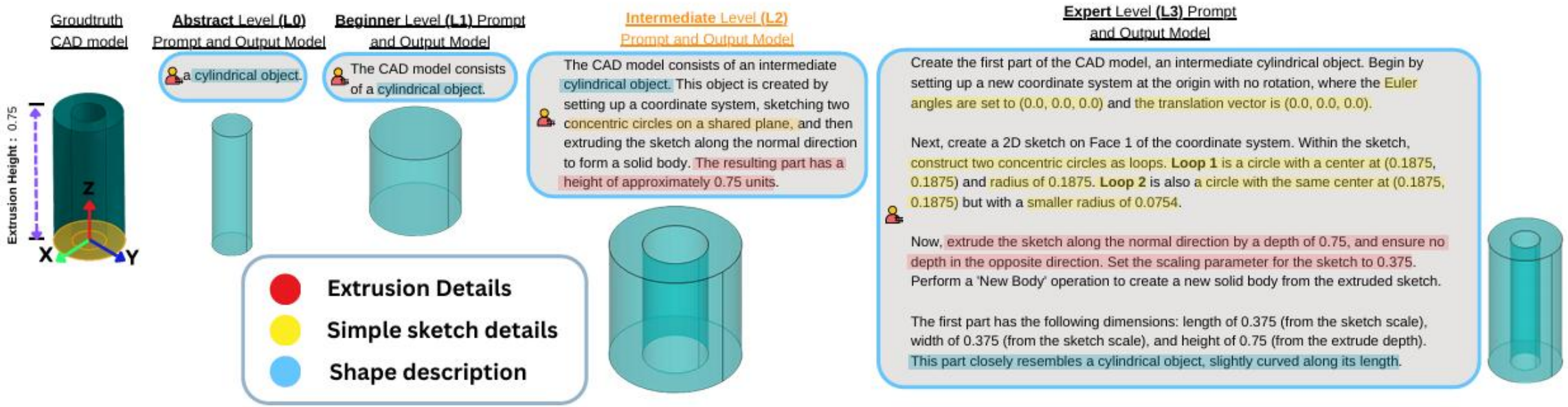


Figure 4: Parametric CAD model generation by Text2CAD transformer using different text prompts. Our text prompts follow a certain structure highlighting the different design aspects of CAD construction workflow (shown in different colors). Abstract (L0) and Beginner (L1) level prompts contain shape descriptions (teal color) whereas Intermediate (L2) and Expert (L3) level prompts are more parametric and contain design details for sketch and extrusion (yellow and red).

照片生成CAD - Img2CAD

Img2CAD实现了根据一张2D图片生成三维CAD建模的功能。
利用GPT-4V提取语义信息，预测全局离散基础结构。
提出了 TrAssembler，以具有语义的离散结构为条件，预测连续属性值。

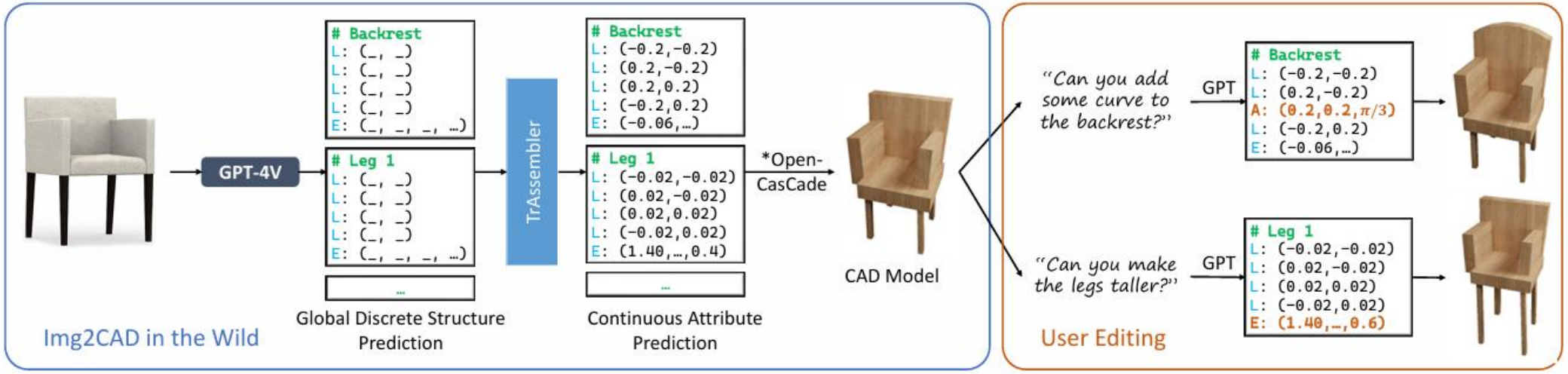


Figure 1: Img2CAD: A Framework for Reverse Engineering 3D CAD Models from Single-View Images. Our method leverages GPT-4V to predict the discrete CAD program structure and then uses a semantic-conditioned transformer to predict the continuous attributes. This approach allows users to easily reconstruct and edit a CAD model from a single-view input image. We use OpenCasCade [Capgemini [n. d.]] to convert the CAD program back to a 3D mesh.

照片生成CAD - Img2CAD

CAD能都是用线、弧、圆表示

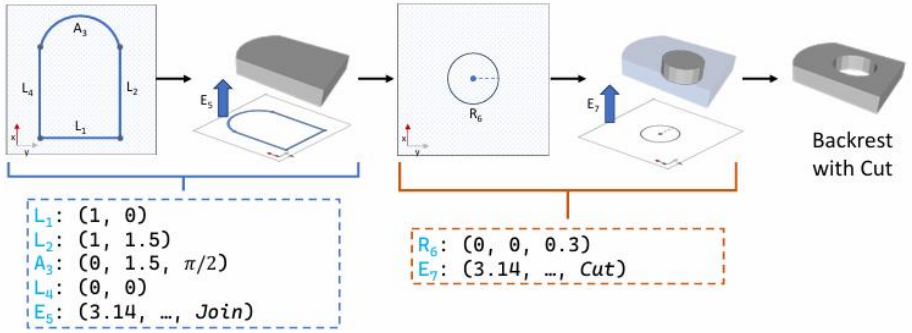


Figure 2: An illustrative example of a sketch and extrusion command sequence for a chair's backrest with a circular cutout. The backrest is created using two extrusion blocks: the first one employs the sketch commands *lines* and *arcs* with the extrusion type *NewBody*, while the second one uses the sketch command *circles* with the extrusion type *Cut*.

CAD结构数据的描述样例

- $L : (x, y)$, where (x, y) defines the endpoint of a line segment [1].
- $A : (x, y, \alpha)$ where (x, y) is the endpoint of an arc with a sweep angle α.
- $R : (x, y, r)$, which defines a circle with the center (x, y) and radius r.

先解析结构，再确定数字

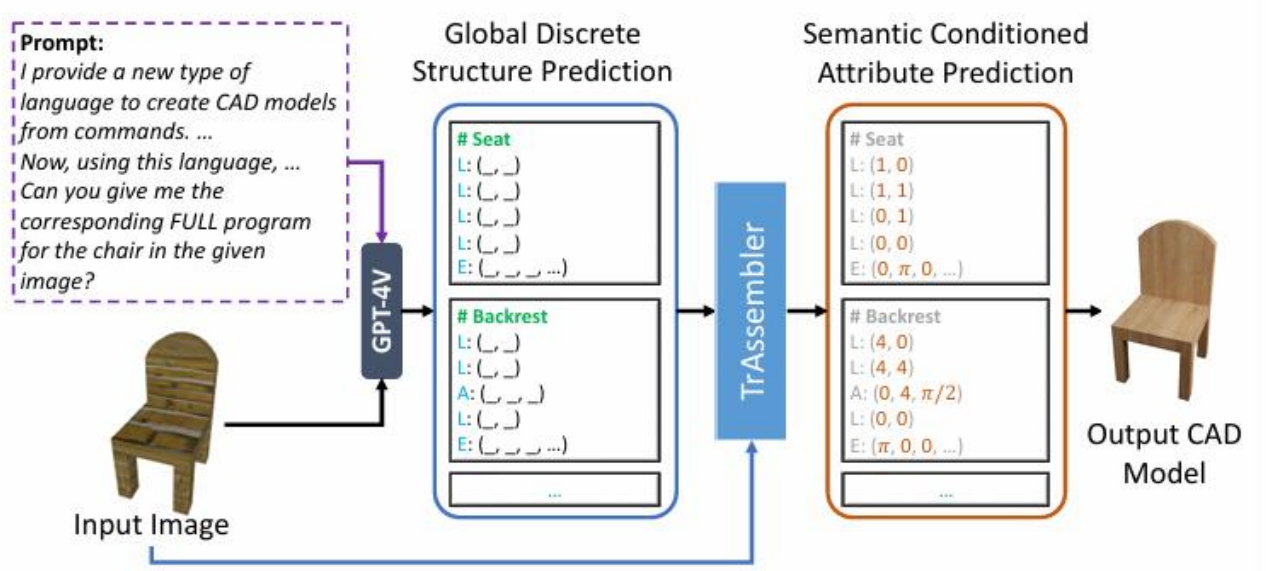


Figure 3: Overview of our two-stage pipeline. The first stage involves GPT-4V interpreting the input image to decompose it into semantic parts and generate a sequence of discrete CAD structures. The second stage uses a transformer-based model to predict continuous attributes for each part.

照片生成CAD - Img2CAD

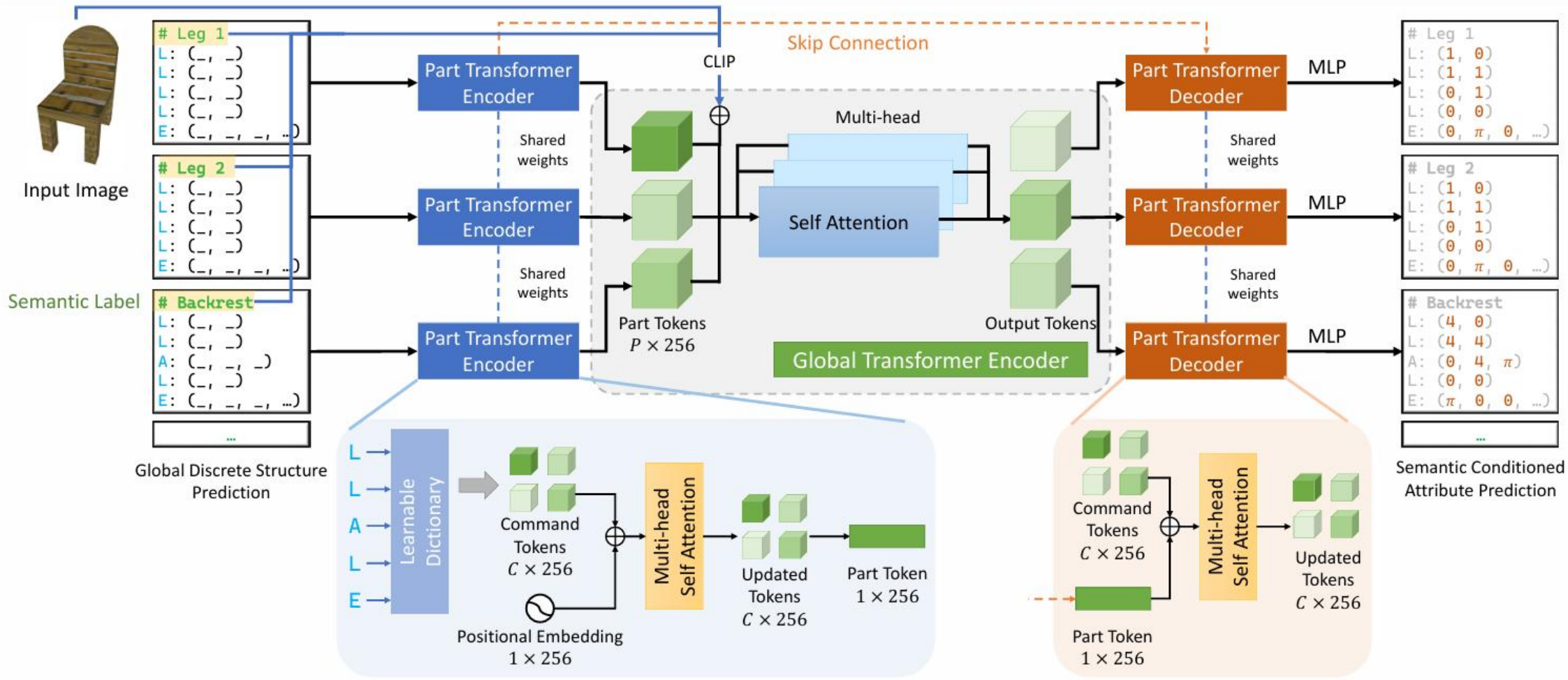


Figure 5: Overall architecture of TrAssembler. The model first leverages part transformer encoder to generate part embeddings with part structures given by GPT-4V. The part embeddings are then fed into a global transformer encoder and refines them with multi-head self-attention, followed by an MLP to decode the final attribute parameters for each command.

草图生成CAD - PHT-CAD

PHT-CAD是一个高效生成CAD表示的方法，其利用视觉语言模型(VLMs)实现2D参数化工程绘图到结构化CAD数据的高效且精确转化。

同时论文还提供了一个大规模Img2CAD数据集ParaCAD，是第一个明确集成几何层和注释层的大规模基准测试，涵盖100万个标注绘图数据用以训练，以及3000真实数据用于测试。

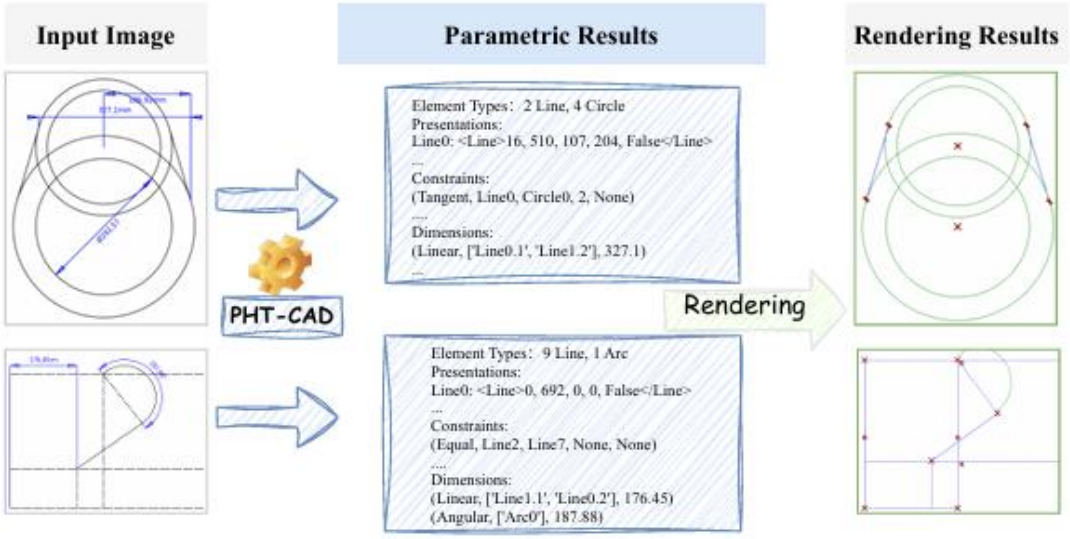
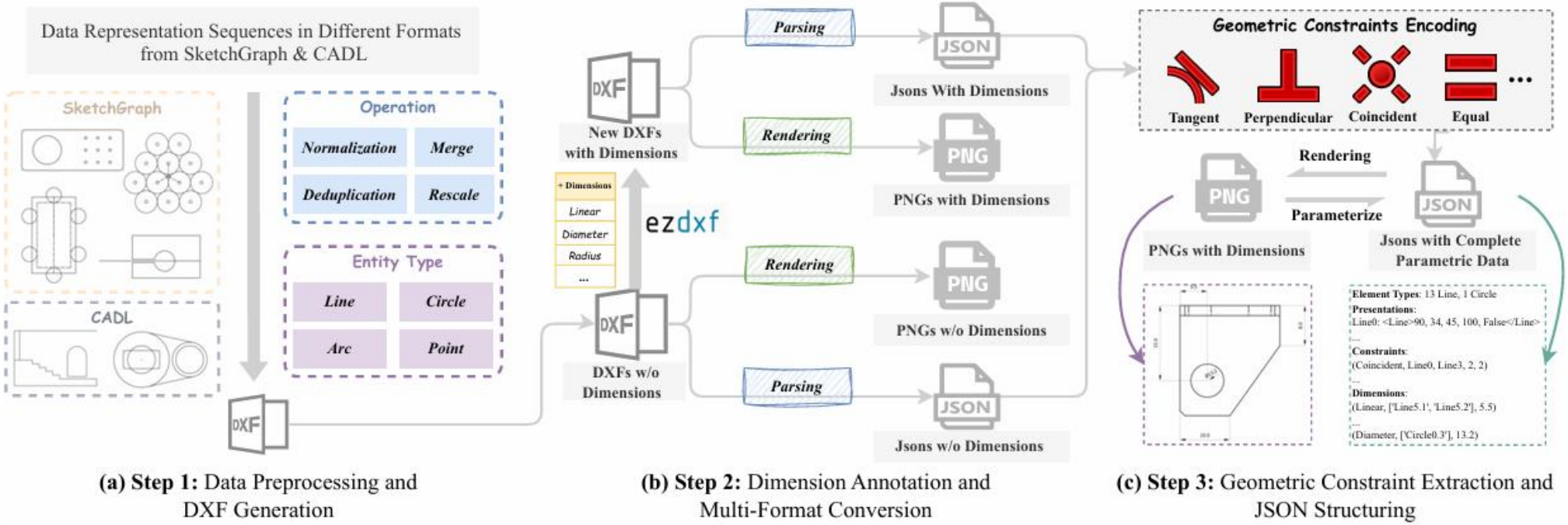


Figure 1. Comparison between sketches/hand-drawn images and structured engineering drawings.

草图生成CAD - PHT-CAD



(a) Step 1: Data Preprocessing and DXF Generation

(b) Step 2: Dimension Annotation and Multi-Format Conversion

(c) Step 3: Geometric Constraint Extraction and JSON Structuring

Figure 3. Figure: Data processing pipeline for engineering drawing parameterization. (a) Step 1: Preprocessing and DXF generation from SketchGraph and CADL datasets. (b) Step 2: Dimension annotation and conversion to multiple formats (DXF, PNG, JSON). (c) Step 3: Extraction of geometric constraints and structuring of parameterized data.

step1: 将素描转化为直线、圆弧、圆、点四种结构化数据，以DXF的形式记录
step2: 在DXF文件中添加尺寸标注信息，转化为DXF，PNG，JSON多种形式
step3: 分析基本元素之间的位置关系，例如平行、垂直等，把这些约束关系插入到JSON文件中
最终JSON文件包含三种信息，基本元素信息、标注信息和约束信息

草图生成CAD - PHT-CAD

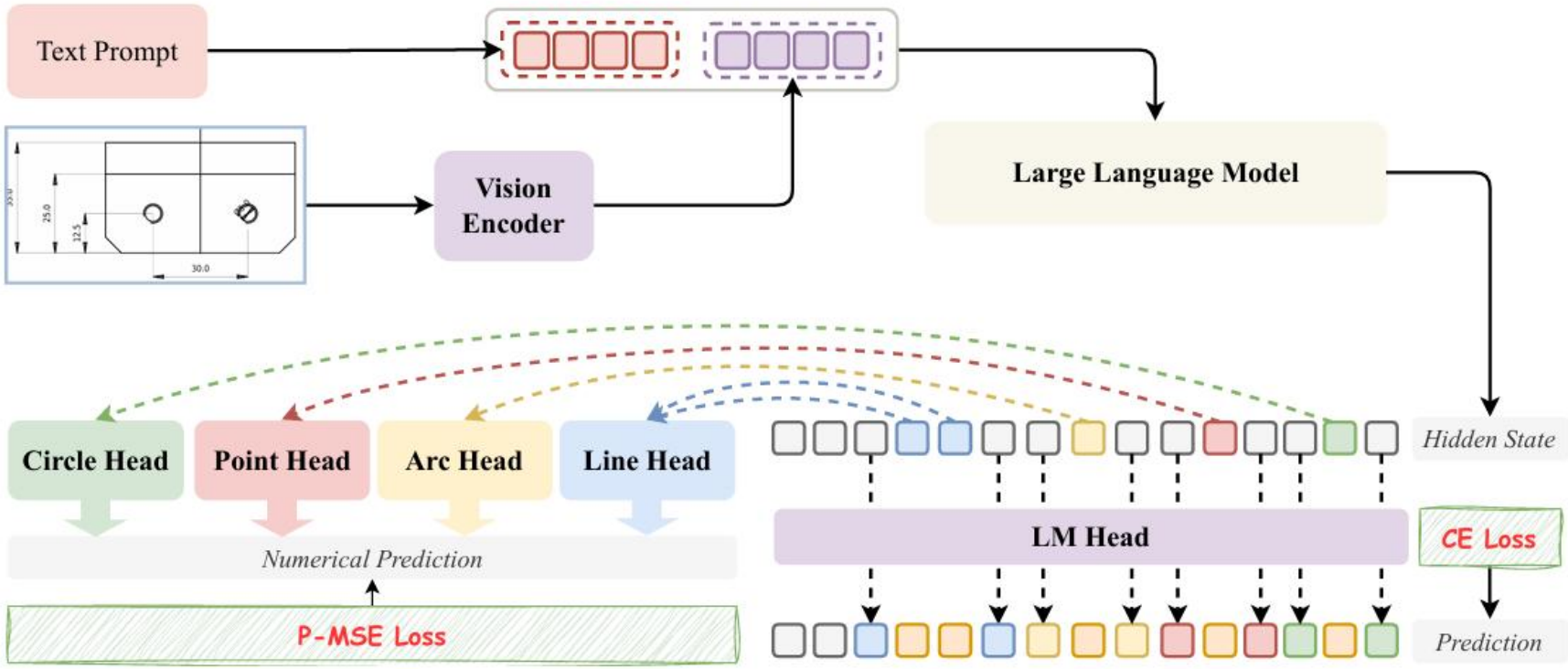
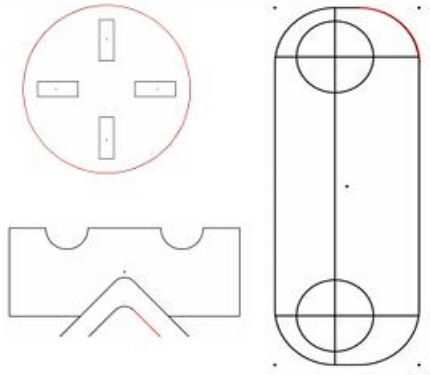
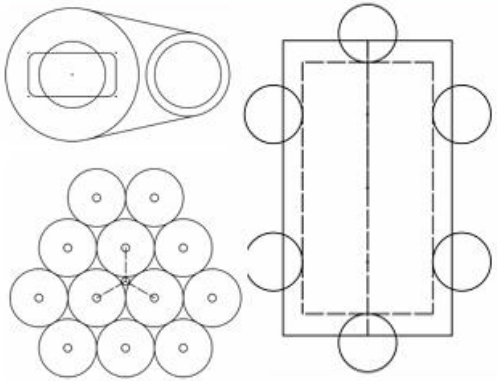
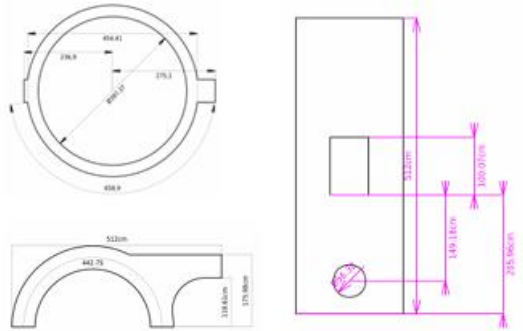


Figure 4. Overview of PHT-CAD framework.

Vison Encoder： ViT，用于抽取CAD绘图中的几何特征
LLM：Qwen2.5，用作TextDecoder，将视觉特征解析成CAD结构化的信息
CE Loss：监督原始语言头的输出
P-MSE Loss：监督4个专用回归头的预测输出

草图生成CAD - PHT-CAD

三种不同的训练阶段

Training Stage	Primitive Perception Tuning	Drawings Structural Perception Tuning	Annotation-Geometry Alignment
Training Data	 Sigle Primitive Recognition Data	 Sketch Structural Perceiving Data	 Dimentional Annotated Drawing Data
Prompt	Given the image, you should extract the primitive marked in red each type should be represented in the following format: <Line>.....</Line> <Circle>...</Circle> <Arc>.....</Arc>	Please think step by step. **First**, count the number of the primitives **Then**, each primitive should be represented in the following format: ..... **Last**, give the constraint relationship between the primitives. The format of the constraint relationship is as follows:	Please think step by step. **First,** *count* the number of the primitives **Second**, each primitive should be *represented* in the following format: ..... **Third**, give the *constraint relationship* between the primitives. **Finally**, you need to output the *dimension* informationin
Prediction	**1. Primitive Type** **2. Primitive Parameters**	**1. Primitive Type** **2. Primitive Presentations** **3. Constraint Relationship between Primitives**	**1. Primitive Type** **2. Primitive Presentations** **3. Constraint Relationship between Primitives** **4. Dimension Information**
	类型、参数	类型、参数、关系	类型、参数、关系、三维

CAD程序解析 - CADTalk

CADTalk以三维CAD程序作为输入，自动将程序按照语义进行分割并进行相关注释；论文对CADTalk在新的数据集上进行评估，包含人造和机器制造的数据。

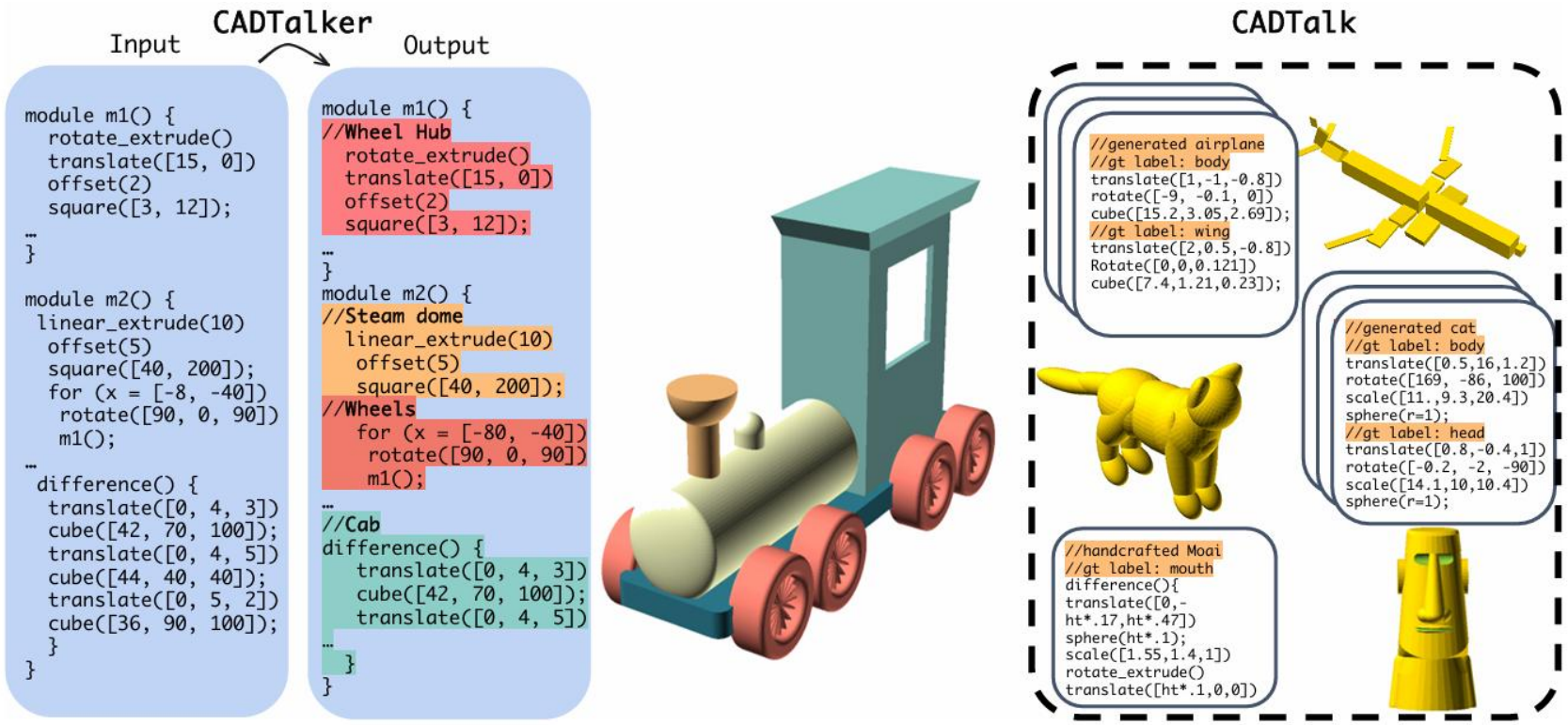


Figure 1. Given a CAD program as input, our algorithm – *CADTalker* – automatically generates comments before each code blocks to describe the shape part that is generated by the block (left). We evaluate our algorithm on a new dataset of commented CAD programs – *CADTalk* – that contains both human-made and machine-made CAD programs (right).

CAD程序解析 - CADTalk

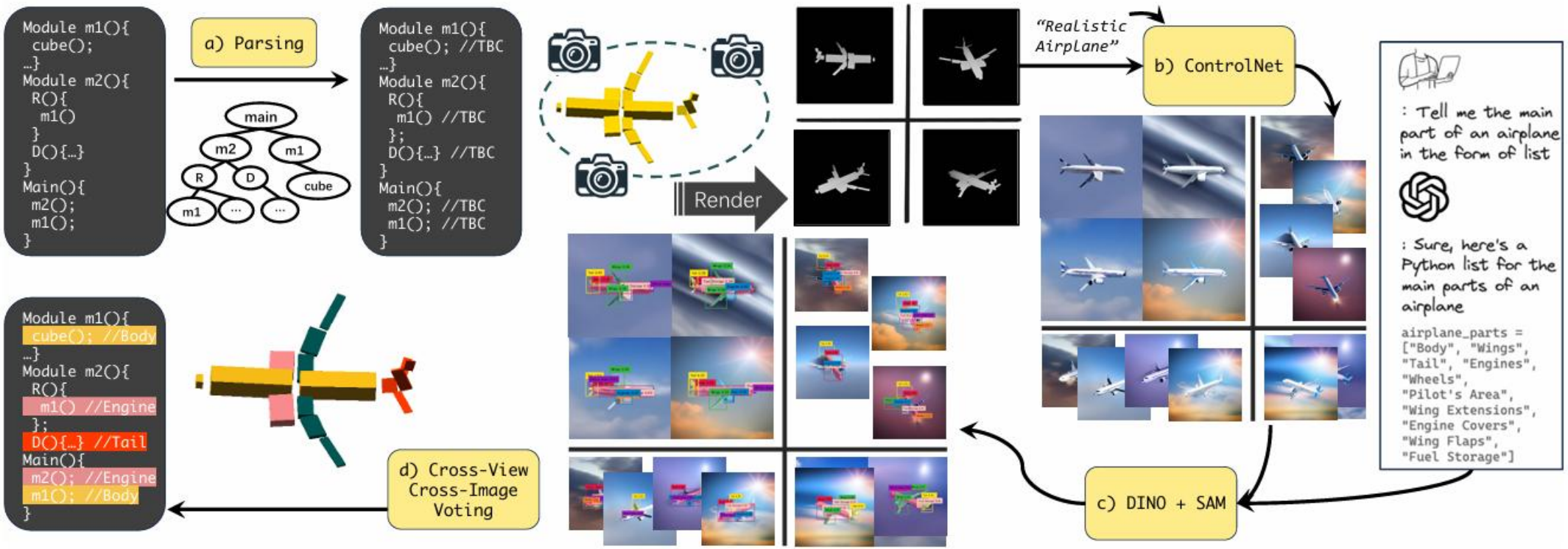
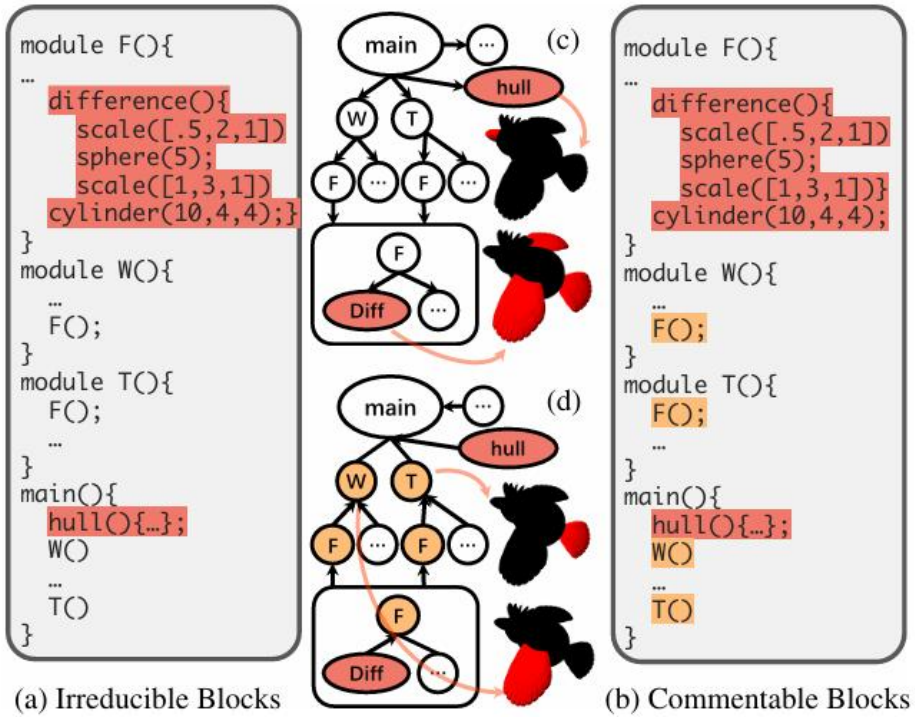


Figure 2. **Algorithm overview.** We first parse the input program to identify commentable code blocks, marked with TBC (a). We then execute the program and render the resulting shape under several viewpoints to obtain multiview depth maps, which we convert into realistic images using image-to-image translation (b). In addition, we obtain a list of part names of the shape from ChatGPT. We use these labels to segment semantic parts in the images using computer vision foundation models (c). Finally, we aggregate this semantic information across views by linking it to the code blocks that correspond to the segmented parts (d).

CAD程序解析 - CADTalk

程序不可约块解析，类似编译器



(a) Irreducible Blocks (b) Commentable Blocks

通过聚合DINO检测结果，生成语义标签

Aggregating semantic labels. We represent all possible label assignments via a matrix C, where each entry $C(b, l)$ quantifies the confidence of block b to be assigned label l. We fill in this matrix by accumulating labeling confidence over all 40 images of the shape, in three steps. In a first step we compute, for each image i generated from view v, the confidence of label l to be assigned to block b as:

$$C^i(b, l) = C_{DINO}(i, l) \times IoU(M_b^v, S_l^i), \qquad (1)$$

where $C_{DINO}(i, l)$ is the confidence of label l in image i provided by Grounding DINO, S_l^i is the segmentation mask provided by SAM, M_b^v is the binary mask rendered for block b in view v, and IoU is the Intersection-over-Union.

给分标准

- 扩展部分为开放式，要求使用大语言模型，由同学自选CAD+LLM相关课题完成，包括但不限于CAD生成、CAD解析等。
- 要求：
  - BE COOL, 发挥想象力, 展示创意。
  - 不得完全照搬，要体现自己的工作量。
  - 可以参考论文、博客与开源代码等资源，但需要给出明确的引用。
  - 提交代码链接、数据链接、模型链接、报告和PPT。
  - 汇报和demo演示。

- 给分标准：
  - 基础给分：10分，完成整个流程，准时提交报告，按时汇报
  - 质量给分：10分，按展示效果和创新性给分。
  - 如果包含原创性的、有效的创新，可获得整个PJ3满分（20分）

报告要求

- 使用英文，描述清晰，提交PDF文件。
- 推荐使用latex，可以参考NIPS会议模板。
（https://media.neurips.cc/Conferences/NeurIPS2024/Styles.zip）

- 格式：
  - ✓摘要
  - ✓介绍
  - ✓相关工作
  - ✓方法
  - ✓实验结果
  - ✓结论

如果使用公开代码和数据需要脚注标注。结尾给出具体组员的姓名、学号与分工情况。

汇报要求

- 制作PPT，讲解如下内容：
  - ✓成员与分工
  - ✓选题的内容与意义
  - ✓所用方法
  - ✓结果

- 5月29日（15周） 6月5日（16周）进行demo展示，每组时间控制在5分钟左右：
  - 展示内容可以是图片、GIF、视频、网站等
  - 推荐使用实时渲染、交互式的方式展示
  - BE COOL，发挥想象力，震撼的展示效果和创新的应用能够加分

- 压缩包标题：2025图形学Project3 姓名1 姓名2 姓名3 姓名4 （最多4人）
- 提交到elearning：报告+PPT打包成zip，DDL：5月29日中午9：55

参考文献

1. Text2CAD: Generating Sequential CAD Designs from Beginner-to-Expert Level Text Prompts 根据语言指令生成CAD设计结构 https://sadilkhan.github.io/text2cad-project/
2. PHT-CAD: Efficient CAD Parametric Primitive Analysis with Progressive Hierarchical Tuning 根据手绘图像生成CAD
3. Img2CAD: Reverse Engineering 3D CAD Models from Images through VLM-Assisted Conditional Factorization 根据图像生成三维CAD结构
4. Mamba-CAD: State Space Model For 3D Computer-Aided Design Generative Modeling
5. CAD-GPT: Synthesising CAD Construction Sequence with Spatial Reasoning-Enhanced Multimodal LLMs 从文本以及图像生成CAD 3D模型 https://openiwin.github.io/CAD-GPT/
6. LLM4CAD: MULTI-MODAL LARGE LANGUAGE MODELS FOR 3D COMPUTER-AIDED DESIGN GENERATION 借助GPT以及GPT-4V利用多模态数据（文本，图像以及草图）生成CAD建模程序的方式 https://sidilab.net/wp-content/uploads/2024/07/idetc2024_llm4cad_final.pdf
7. SolidGen:An Autoregressive Model for Direct B-rep Synthesis 不同于预测cad建模序列，直接预测表示B-reps顶点、边缘等的全新表示形式 https://arxiv.org/abs/2203.13944
8. Cadtalk: An algorithm and benchmark for semantic commenting of cad programs https://enigma-li.github.io/CADTalk/

参考数据集

名称	网址	描述
DeepCAD	https://github.com/ChrisWu1997/DeepCAD	The data we used are parsed from Onshape public documents with links from ABC dataset. We also release our parsing scripts here for anyone who are interested in parsing their own data.
ABC-Dataset	https://archive.nyu.edu/handle/2451/61215	explicitly parametrized curves and surfaces, providing ground truth for differential quantities, patch segmentation, geometric feature detection, and shape reconstruction
ParaCAD	https://www.modelscope.cn/datasets/yuwenbonnie/ParaCAD-Dataset/summary	the first large scale benchmark that explicitly integrates both the geometric and annotation layers. 10 million for train and 3000 real data for test
CADTalk	https://enigma-li.github.io/CADTalk/	5,288 machine-made programs and 45 human-made programs with ground truth semantic comments
Text2CAD	https://huggingface.co/datasets/SadilKhan/Text2CAD	The dataset contains ~170K models and ~660K text annotations, from abstract CAD descriptions (e.g., generate two concentric cylinders) to detailed specifications (e.g., draw two circles with center (x,y) and radius r1, r2, and extrude along the normal by d...).