# Text-to-CAD Generation Through Infusing Visual Feedback in Large Language Models

Ruiyu Wang[1] Yu Yuan[2] Shizhao Sun[3] Jiang Bian[3]

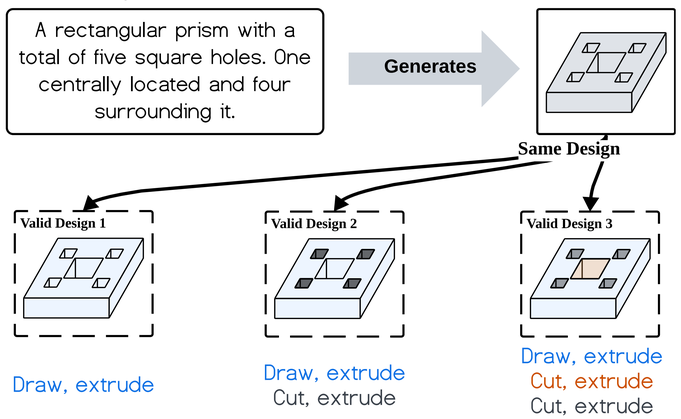
[1]University of Toronto, [2]University of Science and Technology of China, [3]Microsoft Research

Code

Paper

## Introduction

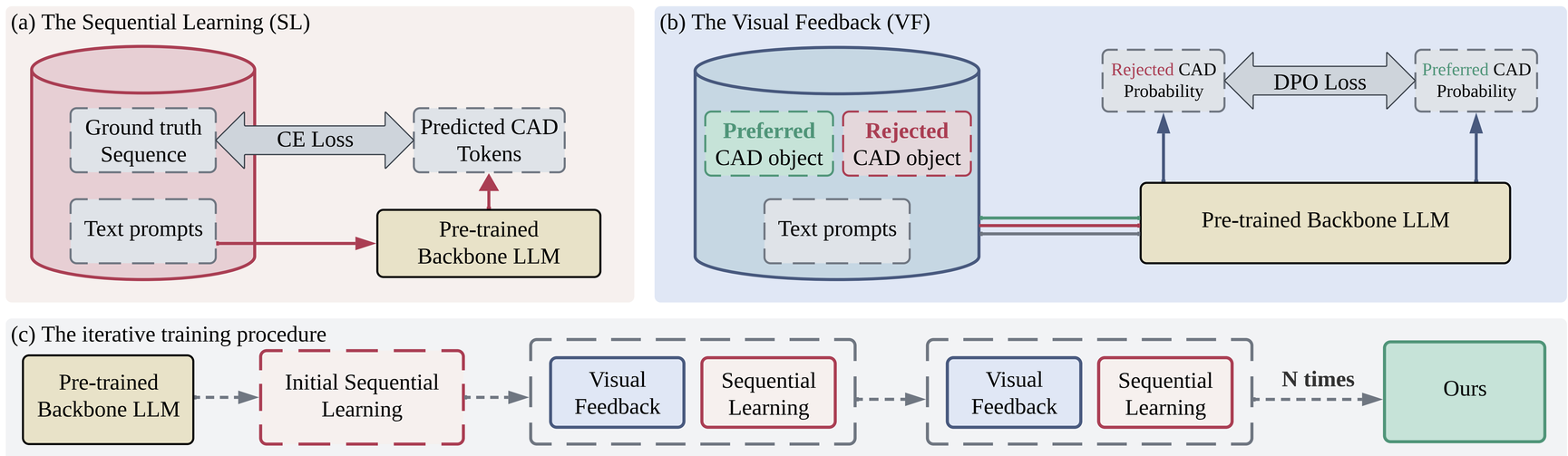
CAD is a many-to-one modality. SFT on only sequential data makes the CAD model overlook other shapes that are also valid.




**Contributions**: 1. integrated visual signals to the Text-to-CAD pipeline 2. improved CAD generation quality.

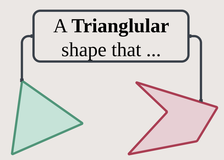
## Methodology

(a) Sequential Learning: trains LLMs using ground truth CAD parametric sequences.
(b) Visual Feedback: rewards visually preferred objects and penalizes the unpreferred.
(c) Iterative Training: balances the contribution of both signals.



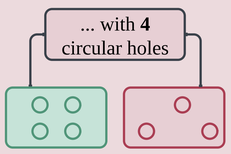

## Feedback Criteria

i. Shape Quality



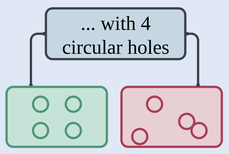

***Shape Quality***: *Does the model generate correct shapes?*

ii. Shape Quantity



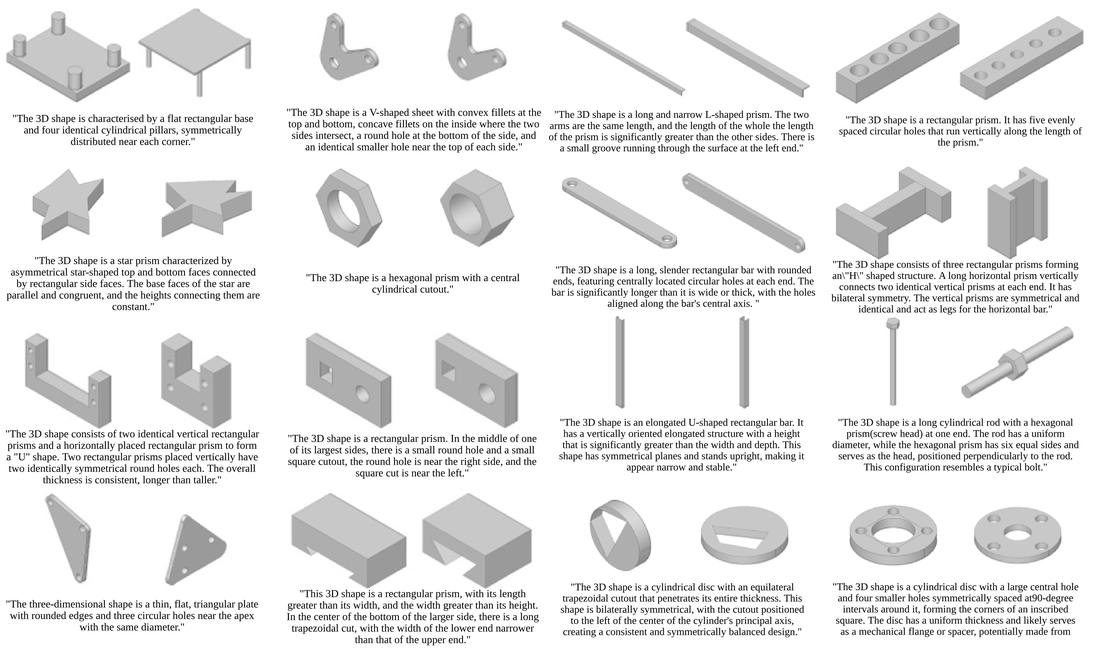

***Shape Quantity***: *Does the model generate correct # of shapes?*

iii. Item Distribution




***Distribution***: *Do the shapes distribute naturally?*

## Qualitative Evaluation




## Quantitative Evaluation

**Sequential-level** quality:

| Method | F1-Sketch | F1-Extrusion | Chamfer Dist. |
|---|---|---|---|
| **GPT-4o-**8shot | 82.96 | 85.72 | 68.50 |
| **Text2CAD** | 63.94 | 92.13 | 30.23 |
| **CADFusion** | **85.22** | **92.79** | **19.89** |

| Method | COV | MMD | JSD |
|---|---|---|---|
| **GPT-4o-**8shot | 72.40 | 6.60 | 37.93 |
| **Text2CAD** | -* | -* | -* |
| **CADFusion** | **90.40** | **3.49** | **17.11** |

**Visual-level** quality:

| Method | IR | LVM Score | Avg. Rank |
|---|---|---|---|
| **GPT-4o-**8shot | 74.26 | 5.13 | 3.22 |
| **Text2CAD** | **3.37** | 2.01 | 2.97 |
| **CADFusion** | 6.20 | **8.96** | **1.86** |

**We could not compute these due to different setups with Text2CAD.*

UNIVERSITY OF TORONTO

Microsoft

中国科学技术大学 University of Science and Technology of China

Code Paper

ICML International Conference On Machine Learning