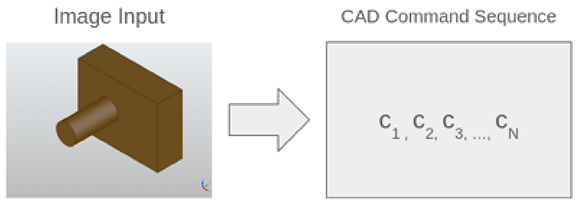
Image-Conditioned CAD Generation via Diffusion Models

Won Sup Song
Center for Global & Online Education, Stanford University

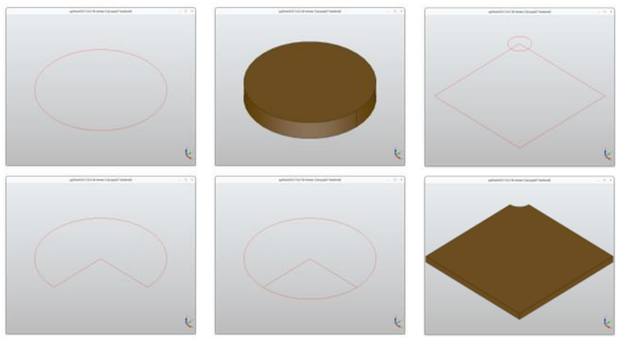
Motivation

The motivation is to develop sketch-based generating CAD command sequences directly from visual representations, aiming to enhance accessibility and streamline the design process for a broader user base.



Dataset / Methods

Image Dataset

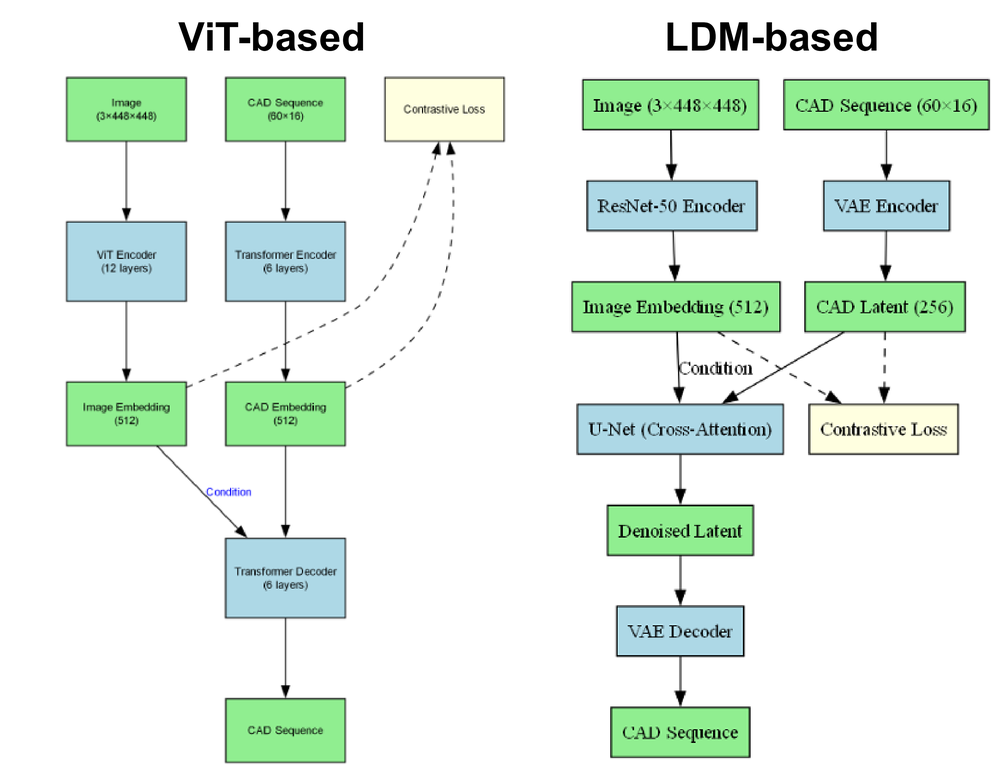


Hardware: NVIDIA RTX 4080 GPU
Processing:

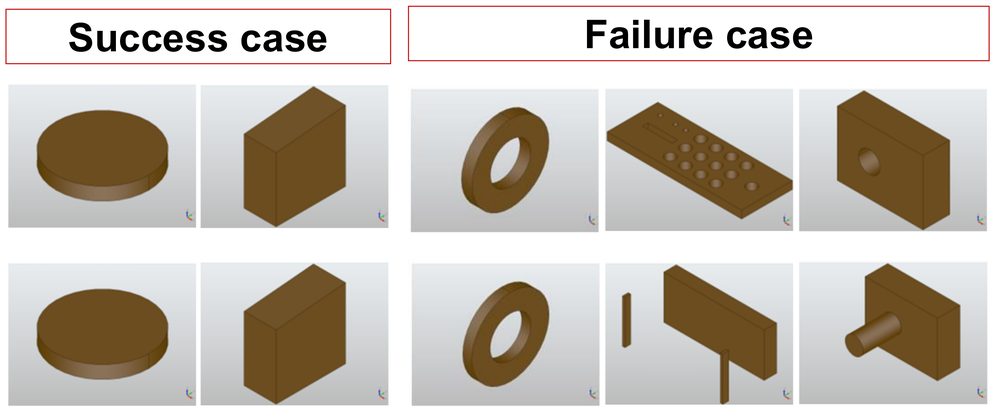
- Images: resized to 448x448, normalized to [0,1], and augmented with random rotations (±10◦) and horizontal flips.
- CAD sequences: normalized to a 2 ×2 ×2 cube, parameters quantized to 256 levels.

Features: CAD commands are embedded as 16-dimensional vectors, capturing command type (e.g., line, arc) and parameters (e.g., coordinates, radius).

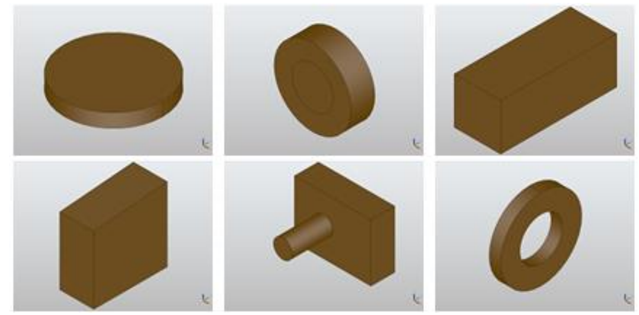
Model



Generation Results (Image-Conditioned)



Generation Results



Analysis

Token-based Result

Model	Tok. Acc.	Param. Acc.	CD ↓	ISR ↓
DeepCAD [2]	**99.50%**	**98.00%**	**0.750**	**3.20%**
ViT	99.30%	97.50%	0.790	3.50%
LDM	99.40%	97.60%	0.780	3.40%

Image-based Result

Model	FID ↓	COV ↑	MMD ↓	JSD ↓
DeepCAD [2]	0.80	80.00	1.40	3.70
ViT	0.25	79.50	1.45	3.65
LDM	**0.15**	**81.00**	**1.35**	**3.55**

Future Work

Limitation:

- Only supports the extrude command, leading to only simplistic CAD models
- Dataset size of 10,000 pairs is insufficient compared to larger datasets like those of CLIP,

Future Work:

- Incorporate more complex CAD commands
- Scaling datasets to improve embedding quality, optimizing for real-time use, and exploring hybrid ViT-LDM models with pre-training on larger CAD datasets