Feasibility Study on the Use of Imagination-Based Orthographic Views for 3D CAD Modeling Assistance

Yeong-A Jang[1,2], Hogeon Seo[1,2]*
[1]Korea Atomic Energy Research Institute, [2]University of Science & Technology
*Corresponding Author: hogeony@hogeony.com

1. INTRODUCTION

- 3D CAD (Computer-Aided Design) modeling requires **expertise and manual work;** MLLMs (Multimodal Large Language Models) can generate CAD from images and drawings, but many existing methods need **inputs with sufficient shape information**, such as drawings.
- Often **only a single image is available**, and a single 3D view lacks information on **occluded geometry and depth**. Such methods are thus **hard to use without drawings**.
- We evaluate on **30 parts** whether **MLLM-generated "Imagination" orthographic views** can serve as **auxiliary input**.

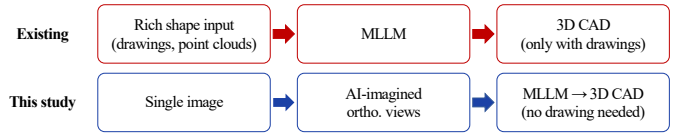


Fig. 1. Existing approach vs. this study

2. METHODOLOGY

2.1. CAD Generation Pipeline

- An MLLM generates **CAD code** from the input, which the OCCT (Open CASCADE Technology) geometric kernel executes to obtain 3D geometry.
- Evaluated **by pose-aligned volumetric Intersection over Union (IoU; 0–1)** against the ground truth.



Fig. 2. CAD generation and evaluation pipeline

2.2. Imagination Orthographic View Generation

- **Generator:** gemini-3-pro-image
- **Input: single 3D (isometric) view + part description only**; no ground-truth geometry, drawings, or dimensions exposed

Prompt to the image generator (abridged)
"You are a drafting assistant. Given the isometric render and the part description, infer the hidden geometry and draw a clean orthographic three-view drawing: FRONT, TOP, RIGHT SIDE in one row (third-angle projection); black lines on white, dashed lines for hidden edges."

- **Output:** 3 generated views → given to CAD-code MLLM with the single 3D view (condition B)
- **Settings: no prompt tuning, iterative generation, or verification**

2.3. Input Conditions (30 parts)

- **Dataset: 30** mechanical parts from a public dataset (Zenodo)
- **Ground truth:** original STEP model of each part → GT orthographic views and SVG drawings derived from it

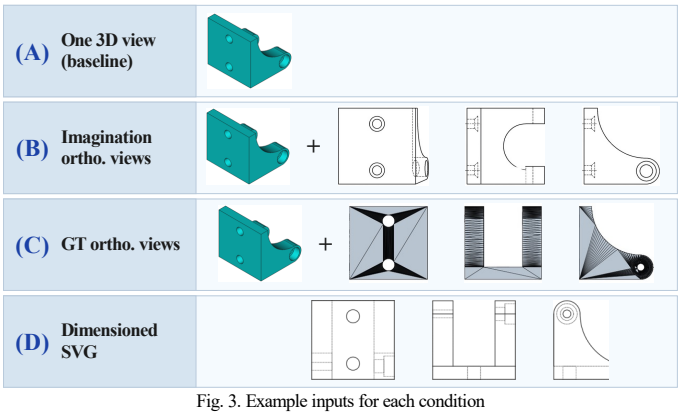


Fig. 3. Example inputs for each condition

3. RESULTS & DISCUSSION

3.1. IoU Distribution by Condition

Table I. IoU distribution statistics (n = 30)

Case	Input	Min	Q1	Med.	Q3	Max	Mean
(A)	One 3D view (baseline)	0.077	0.255	0.401	0.527	0.789	0.405
(B)	Imagination ortho. views	0.124	0.337	0.414	0.548	0.921	0.455
(C)	GT ortho. views	0.111	0.399	0.523	0.626	0.808	0.515
(D)	Dimensioned SVG	0.296	0.509	0.855	0.985	1.000	0.752

IoU = overlap with ground-truth shape (0–1; 1 = exact match).
Min–Max, Q1–Q3 = spread over 30 parts.

Table II. Paired comparison against the baseline (A)

Contrast	Mean diff.	p	r_{bc}	Win / Loss
(B)–(A)	+0.050	0.151	+0.31	21 / 8
(C)–(A)	+0.110	0.001	+0.69	21 / 8
(D)–(A)	+0.347	< 0.001	+0.99	29 / 1

Mean diff. = IoU change vs (A) (+ = better) · p = Wilcoxon signed-rank (< 0.05 = significant)
r_{bc} = effect size (larger = stronger) · Win / Loss = number of parts better / worse than (A)

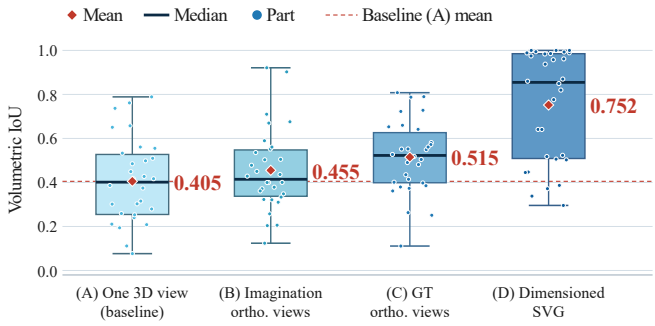


Fig. 4. 3D CAD generation accuracy by input condition (n = 30, paired samples)

- **GT orthographic views (C)**: mean IoU **+0.110** vs baseline (A) (p = 0.001) → effect of adding orthographic views confirmed
- **Dimensioned SVG (D)**: highest mean IoU, **0.752**

3.2. Performance of Imagination Views

- The mean IoU of (B) was **0.455**, **+0.050** over (A), and beat (A) on **21 of 30 parts**.
- The (A)–(B) difference was **not statistically significant** (p = 0.151); quality varied by part, with some outputs having holes or steps that differ from the original geometry.
- Default API capability was used: an **initial feasibility check**.

3.3. Upper-Bound Potential

- Best IoU of (B): **0.921**
- Parts with **occluded geometry**: IoU **0.240 → 0.902** (on par with or better than GT views)
- Stabilizing view quality → **room for further improvement**

4. CONCLUSIONS

- **Imagination orthographic views (B)**: mean IoU **0.405 → 0.455**; **21 of 30 parts** improved; best **0.921**
- Drawings previously required → **feasibility shown:** MLLM-generated orthographic views as auxiliary input, **even without drawings**
- **Future work:** automatic quality evaluation and selection of generated views → improved consistency

This work was supported by the Institute Strategic Development Project (ISD) funded by the Ministry of Science and ICT (No. 526140-26), the Substantiation Support Program through the Korea Innovation Foundation, and the NRF of Korea (No. RS-2026-25545949), funded by the Ministry of Science and ICT.