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, Republic of Korea*
[2]*University of Science and Technology, Republic of Korea*
E-mail: hogeony@hogeony.com

***Keywords : Imagination, 3D CAD Generation, MLLM, Agentic CAD**

1. Introduction

3D CAD models are used across a wide range of engineering fields, including the design, manufacturing, and analysis of mechanical parts. Generating a CAD geometry requires specialized knowledge, and the user must model the shape manually while referring to images and drawings. This process demands time and effort. Recently, studies have been conducted to automatically generate 3D CAD models from various types of input information by exploiting the image understanding and code generation capabilities of multimodal large language models (MLLMs) [1-3].

In 3D CAD generation, the shape information available to the model differs according to the form of the input. Text can describe the characteristics of an object but has difficulty expressing complex spatial relationships directly, whereas an image provides the shape of an object but often does not contain accurate numerical information such as lengths and positions. In contrast, a vector drawing that includes dimensions can provide shape and numerical information.

In a preliminary experiment, 3D CAD generation performance was compared using a single 3D view and an SVG vector containing dimensional information. The SVG input yielded the highest generation performance, which was attributed largely to the explicit dimensional information contained in the vector input. In practice, however, there are cases in which no CAD drawing or dimensional information exists and only a single image can be obtained.

With a single 3D view, it is difficult to accurately determine occluded geometry and structures along the depth direction, depending on the viewing direction. Orthographic views such as the front, top, and side views provide the shape of the object from different directions and can therefore complement the spatial information that is missing from a single image. Previous work proposed methods for inferring shape and dimensional parameters from orthographic drawings [4]; however, such approaches are limited in cases where the orthographic views do not exist.

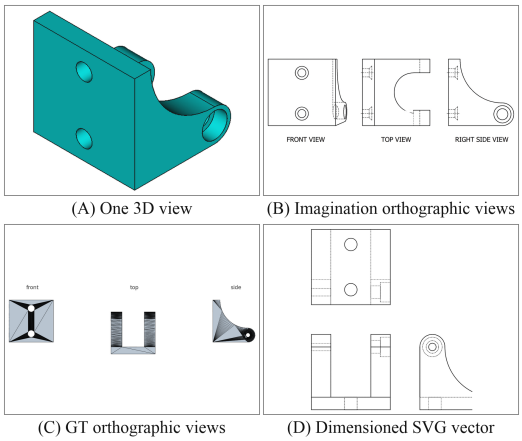
In this paper, we experimentally examine whether adding orthographic views as auxiliary inputs improves generation performance in single-3D-view-based 3D CAD generation. The effect of multi-view information on CAD generation is quantitatively confirmed through a condition in which ground-truth orthographic views are added, and, assuming a situation in which orthographic views do not exist, an Imagination condition in which the MLLM generates orthographic views from a single 3D view is evaluated together. The generation accuracy of four input conditions is compared for 30 mechanical parts, and the feasibility of using MLLM-generated orthographic views as auxiliary inputs for 3D CAD generation is assessed.

2. Methodology

2.1. CAD generation pipeline

In the proposed pipeline, an MLLM generates parametric CAD code from an input (an image or a drawing) and a part description, and the code is executed in a geometric kernel (OCCT) to obtain 3D geometry in STEP and STL formats. When code execution fails, the error message is returned to the model and a recovery loop allows up to one repair attempt. The model gemini-3.1-pro-preview was used for CAD code generation.

Output quality was evaluated using the pose-aligned volumetric intersection-over-union (IoU) with respect to the ground-truth geometry. Both shapes were normalized by the bounding-box center and the longest edge, voxelized on a 48^3 grid, and the maximum IoU over 24 axis-aligned rotations was adopted. The IoU ranges from 0 to 1, where a value closer to 1 indicates higher agreement with the original geometry.



(A) One 3D view (B) Imagination orthographic views
(C) GT orthographic views (D) Dimensioned SVG vector

Fig. 1. Example inputs for each condition

Four input conditions were used in the experiments, and examples of each condition are presented in Fig. 1. Condition (A) uses only a single isometric projection image as the input. Condition (B) uses the single 3D view together with the front, side, and top views generated by the MLLM from the image of Condition A. Condition (C) uses the single 3D view together with the ground-truth orthographic views. Condition (D) uses an SVG file containing dimensional and coordinate information as the input.

2.2. Imagination orthographic view generation

The imagination orthographic views were generated using gemini-3-pro-image. Only the single 3D view image and the part description text were provided as inputs, and the model was instructed to generate the front, side, and top views of the part in a third-angle projection layout. Information that is not given in the CAD generation stage, such as the ground-truth geometry or the original drawing, was never exposed to the generator.

2.3. Experimental design

The experiment was configured by applying each of the four conditions (A, B, C, and D) to 30 mechanical parts and measuring the IoU. To characterize the distributions, the mean, median, quartiles, minimum, and maximum were computed. Significance between conditions was assessed using the Wilcoxon signed-rank test, and the effect size was evaluated using the rank-biserial correlation coefficient (r_bc).

3. Results and Discussion

3.1. IoU distribution by condition

The distribution statistics of the IoU for the four conditions over the 30 parts are presented in Table I, and the corresponding box plots are shown in Fig. 2. The results of the paired comparisons between conditions are summarized in Table II.

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 vector	0.296	0.509	0.855	0.985	1.000	0.752

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

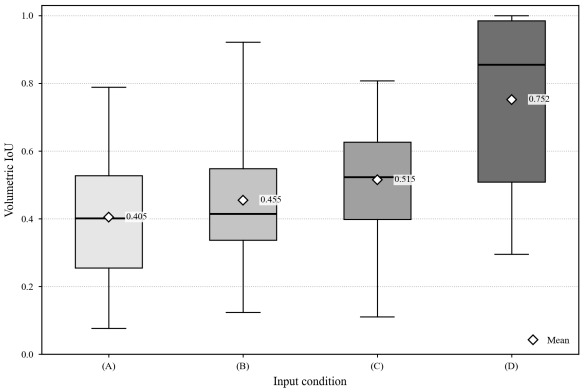


Fig. 2. 3D CAD generation accuracy by input condition (n = 30, paired samples)

The ground-truth orthographic views (C) improved the mean IoU by +0.110 relative to the baseline condition (A) (Wilcoxon signed-rank test, p = 0.001, r_bc = +0.69, 21 wins and 8 losses). This shows that generation accuracy improves significantly when orthographic views are added as auxiliary inputs. The dimensioned SVG vector (D) achieved the highest performance among all conditions, with a mean IoU of 0.752, and exceeded the baseline condition for 29 of the 30 parts ($p < 0.001$).

3.2. Performance of the imagination orthographic views

The mean IoU of Condition (B) was 0.455, which is 0.050 higher than that of the baseline Condition (A) (0.405), and the median of Condition (B) (0.414) also slightly exceeded that of Condition (A) (0.401). Condition (B) outperformed Condition (A) for 21 of the 30 parts, suggesting that MLLM-generated orthographic views may provide useful auxiliary information for 3D CAD generation even when no actual drawing is available. The mean difference from Condition (A) was +0.050 for Condition (B) and +0.110 for Condition (C). Although the two conditions show the same win-loss counts against the baseline, the magnitude of the improvement differs, which is reflected in the effect sizes.

The difference between (A) and (B) was not statistically significant (p = 0.151, r_bc = +0.31). In addition, the quality of the generated orthographic views varied across parts, and in some cases holes or steps that differ from the original geometry were generated, which affected the final CAD generation performance. In this experiment, no dedicated prompt optimization, iterative generation and verification, or model fine-tuning was applied to the orthographic view generation; the default generation capability of the API was used. The present results should therefore be regarded as a confirmation of initial feasibility rather than the optimal performance of the proposed method. Applying prompt designs that account for geometric consistency among the orthographic views, together with verification and correction of the generated results,

is expected to leave room for further performance improvement.

The first quartile of condition B increased from 0.255 in condition A to 0.337. This suggests that the use of imagination orthographic views may act to compensate for some of the low-performance cases. However, additional verification with a larger number of parts is required in order to generalize this trend.

3.3. Upper-bound potential of the imagination views

The maximum value of condition B, 0.921, is the highest among all image-based conditions (A, B, and C) and exceeds the maximum of the ground-truth orthographic views (C), 0.808, by 0.113. This shows that, depending on the part geometry, views generated by the MLLM can lead to results comparable to or even better than those obtained with actual orthographic views.

Large improvements were observed for parts whose geometry is difficult to infer from a single 3D view alone. For a part in which portions of the geometry are occluded in the isometric projection, the IoU improved from 0.240 to 0.902, and for another part from 0.194 to 0.710; in both cases the results were comparable to or better than those of the condition using ground-truth orthographic views. This suggests that the generated orthographic views complemented the geometric information occluded in the single view.

In the box plots of Fig. 2, the upper whisker of condition B is also the highest among the image-based conditions. This indicates that, for some parts, imagination orthographic views can lead to favorable results, and it suggests the practical potential of this approach.

The current performance variation appears to result from the lack of consistency in generation quality. If the generation quality is stabilized by increasing the number of generations, adjusting the prompts, and verifying the geometric consistency of the generated views, there is room for further improvement of the mean performance.

4. Conclusion

This study exploratively verified, for 30 parts, whether orthographic view generation based on the imagination capability of an MLLM can be applied to 3D CAD modeling assistance. The experimental results showed that the mean IoU of the imagination orthographic views (B, 0.455) improved over that of the baseline condition (A, 0.405), and condition B exceeded the baseline for 21 of the 30 parts. The best-performing case reached an IoU of 0.921, the highest maximum among the image-based conditions. These results demonstrate the feasibility of using MLLM-generated orthographic views as auxiliary inputs even in situations where actual orthographic views are not available. In future work, we plan to introduce a mechanism that automatically evaluates and selects the quality of the generated views and to analyze the generation characteristics for different geometry types, thereby extending the research toward improving the performance consistency of imagination-based orthographic view generation.

ACKNOWLEDGEMENT

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, funded by the Ministry of Science and ICT, and the Program for Strengthening Technological Capabilities in Nondestructive Testing through the National Research Foundation of Korea (No. RS-2026-25545949), funded by the Ministry of Science and ICT.

REFERENCES

[1] J. Xu, C. Wang, J. Zhao, W. Liu, Y. Ma, and S. Gao, CAD-MLLM: Unifying Multimodality-Conditioned CAD Generation with MLLM, arXiv preprint, arXiv:2411.04954, 2024.

[2] D. Rukhovich, E. Dupont, D. Mallis, K. Cherenkova, A. Kacem, and D. Aouada, CAD-Recode: Reverse Engineering CAD Code from Point Clouds, arXiv preprint, arXiv:2412.14042, 2024.

[3] D. Mallis, A. S. Karadeniz, S. Kabadayi, D. Rukhovich, N. Potamias, K. Cherenkova, A. Kacem, and D. Aouada, CAD Assistant: Tool-Augmented VLLMs as Generic CAD Task Solvers?, arXiv preprint, arXiv:2412.13810, 2024.

[4] K. Niu, Z. Chen, H. Yu, Y. Chen, T. Fu, M. Zhao, B. Li, and X. Xue, CReFT-CAD: Improving Orthographic Projection Reasoning for CAD via Reinforcement Fine-Tuning, arXiv preprint, arXiv:2506.00568, 2025.