2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

A ZERO-SHOT APPROACH TO DECOMPOSING GRAPHICAL RULES IN ACCESSIBILITY REGULATIONS

Yonghan Kim[1], Ghang Lee[1,2], and André Borrmann[2]
[1]Technical University of Munich, Munich, Germany
[2]Yonsei University, Seoul, South Korea

Abstract

Composite graphical rules hinder interpretation by multimodal large language models (MLLMs), and existing annotation-isolation methods require labeled training data. To address this, we propose a training-free approach using the Segment Anything Model 3 (SAM3) to automatically isolate annotations, remove cross-rule visual clutter, and reconstruct individual rules. Using manually decomposed rules as ground truth, we compared our method against prompt-based generative baselines. Results indicate that while baselines produce visually cleaner diagrams, MLLM-as-a-judge evaluations demonstrate our SAM3-based approach more reliably preserves the original regulatory semantics.

Introduction

Normative knowledge in regulatory documents appears in multiple forms, including natural language text, tables, illustrations, calculation methods, and commentaries (Amor and Dimyadi, 2021). Among these, illustrations play a critical role in conveying requirements that are difficult to express using text alone. In this work, we define illustrated provisions that explicitly encode declarative regulatory requirements as *graphical rules*.

However, graphical rules are often presented as composite diagrams that contain multiple regulatory requirements within a single illustration. Such composite representations can hinder rule interpretation, particularly when multimodal large language models (MLLMs) are used to analyze regulatory documents. Decomposing composite diagrams into atomic, rule-level units is therefore necessary for improving the interpretability and downstream usability of graphical regulatory knowledge.

Existing approaches, however, have important limitations. First, annotation-isolation methods typically rely on supervised learning with labeled datasets, which are costly and difficult to generalize to new regulatory domains (Zhang et al., 2023; Maupou et al., 2024). Second, prompt-based generative approaches can modify diagrams without a trained annotation-isolation model, relying instead on image regeneration, which may fail to preserve the original geometry and introduce semantic distortion (Wu et al., 2023).

To address these challenges, this paper proposes a zero-shot framework for decomposing composite graphical rules and pairing them with corresponding textual provisions. The proposed method leverages the zero-shot segmentation capability of SAM3 (Carion et al., 2025) to isolate annotation elements without labeled training data. These elements are grouped by spatial proximity, and a base image is reconstructed by removing annotations via inpainting. Each annotation group is then overlaid onto the base image to produce rule-level graphical units. Finally, the reconstructed graphical rules are paired with textual provisions using semantic similarity.

Related Works

LLM and MLLM for Rule Interpretation

Recent studies have explored the use of LLMs as reasoning engines for automated code compliance checking (ACC), enabling direct interpretation of regulatory text without explicit rule formalization (N. Chen et al., 2024; Ying and Sacks, 2024; Madireddy et al., 2025; Li, 2026; Iversen and Huang, 2026).

Extending this paradigm, recent work has investigated multimodal large language models (MLLMs) to incorporate graphical regulatory information alongside text (Kim et al., 2025). However, empirical findings show that MLLM performance degrades as visual complexity increases (Hou et al., 2025), particularly when multiple regulatory requirements are encoded within a single diagram.

This limitation is especially pronounced in tasks requiring precise spatial grounding. While MLLMs can often recognize numerical annotations, they frequently fail to associate them with the correct geometric entities or spatial context (Khan et al., 2025; Picard et al., 2025). These observations highlight the need for preprocessing strategies that reduce visual complexity and disentangle composite graphical rules prior to model interpretation.

Annotation isolation on a technical drawing

Annotation isolation aims to separate annotation elements (e.g., text, dimensions, symbols) from geometric content so that each component can be analyzed independently. Because regulatory semantics are primarily conveyed through annotations, this step is essential for interpreting graphical rules.

Early approaches relied on rule-based segmentation using drafting standards and geometric constraints (Dori and Velkovitch, 1998; Lin and Ting, 1997). While effective

under controlled conditions, these methods depend on strong assumptions about drawing regularity and are difficult to generalize across heterogeneous regulatory documents.

More recent approaches formulate annotation isolation as a supervised learning problem, using deep neural networks to detect and segment annotation elements (Zhang et al., 2023; Maupou et al., 2024). Although these methods improve robustness, they require large, labeled datasets and are therefore limited in scalability across domains.

Generative and Prompt-based Diagram Manipulation

Recent generative approaches enable selective modification of diagrams using natural language prompts (Jeong et al., 2025; Goswami et al., 2025; Yang et al., 2025). While these methods can simplify visual content, they rely on image regeneration rather than explicit decomposition. As a result, they may fail to preserve the original geometry and can introduce semantic distortions. This limitation is critical in regulatory contexts, where even minor visual changes may alter the intended meaning of a rule.

Positioning of This Work

In contrast to prior studies, this work focuses on the zero-shot decomposition of composite graphical rules in regulatory documents. Rather than introducing new algorithms, the contribution lies in framing graphical rule decomposition as a zero-shot problem, motivated by the high variability of regulatory drawings, which makes large-scale labeled training impractical.

Unlike annotation-isolation approaches that focus on segmenting individual elements, the proposed method reconstructs rule-level graphical units and pairs them with corresponding textual provisions. Furthermore, in contrast to generative diagram editing methods, it preserves the original diagram structure through decomposition rather than image synthesis.

By leveraging a general-purpose segmentation model without labeled training data, the method enables rule-level understanding while maintaining fidelity to the original illustration.

Proposed Method

The proposed method aims to decompose composite graphical rules into atomic rule-level graphical units and pair them with corresponding textual rules. To achieve this without labeled training data, the framework leverages the zero-shot segmentation capability of SAM3 (Carion et al., 2025), together with geometric and semantic post-processing steps.

As illustrated in Figure 2, the pipeline consists of four main stages. First, graphical and textual provisions are extracted from regulatory documents. Second, composite graphical rules are decomposed through zero-shot annotation segmentation. Third, textual provisions are decomposed into atomic rules. Finally, graphical and textual rules are paired based on semantic similarity.

Extraction of graphical rules and textual provisions

The input to the pipeline is a regulatory document in PDF format. Graphical rules and textual provisions are extracted using PyMuPDF (McKie, 2024), which provides structured access to page-level elements.

Graphical rules are identified by detecting raster images and vector graphics within PDF pages. Each detected graphical region is treated as a candidate composite graphical rule, and its bounding box and page index are preserved to maintain contextual consistency.

Textual provisions are extracted from the same pages and segmented at the paragraph level. To associate text with graphical rules, text blocks are filtered based on layout proximity, without relying on predefined templates or rule identifiers.

Graphical Rule Decomposition

Each composite graphical rule is processed independently to isolate annotation elements. Zero-shot annotation segmentation is performed using SAM3, guided by a fixed set of annotation-related prompts (e.g., arrows, dimension lines, and text labels).

To improve detection of small elements such as arrowheads and numeric annotations, Slicing Aided Hyper Inference (SAHI) (Akyon et al., 2022) is applied. SAM3 generates multiple candidate masks, which are filtered using geometric and visual heuristics (e.g., area, aspect ratio, and stroke density).

The extracted annotation elements are then grouped into rule-level units. Since individual regulatory requirements are typically represented by clusters of annotations, masks are grouped based on spatial proximity using distance-based clustering.A base image is constructed by removing annotation elements via image inpainting. A classical, non-learning-based inpainting method is used to preserve geometric

Textual rule decomposition

Textual provisions associated with a graphical rule often describe multiple regulatory requirements within a single paragraph. To align textual granularity with rule-level graphical units, each textual provision is decomposed into atomic textual rules using a two-stage, instruction-only prompt chaining strategy.

In the first stage, *requirement identification*, the language model is instructed to identify all independent regulatory requirements contained within the input provision. The system prompt explicitly guides the model to separate requirements based on distinct compliance conditions and to report the total number of identified requirements together with brief summaries. This stage focuses on structural decomposition, without modifying the original regulatory intent.

In the second stage, *rule normalization*, each identified requirement is reformulated into a standalone textual rule. The system prompt enforces normalization constraints, requiring that (1) each rule is understandable in isolation, (2) regulatory language using “shall” is consistently applied, (3) only one requirement is expressed per rule,

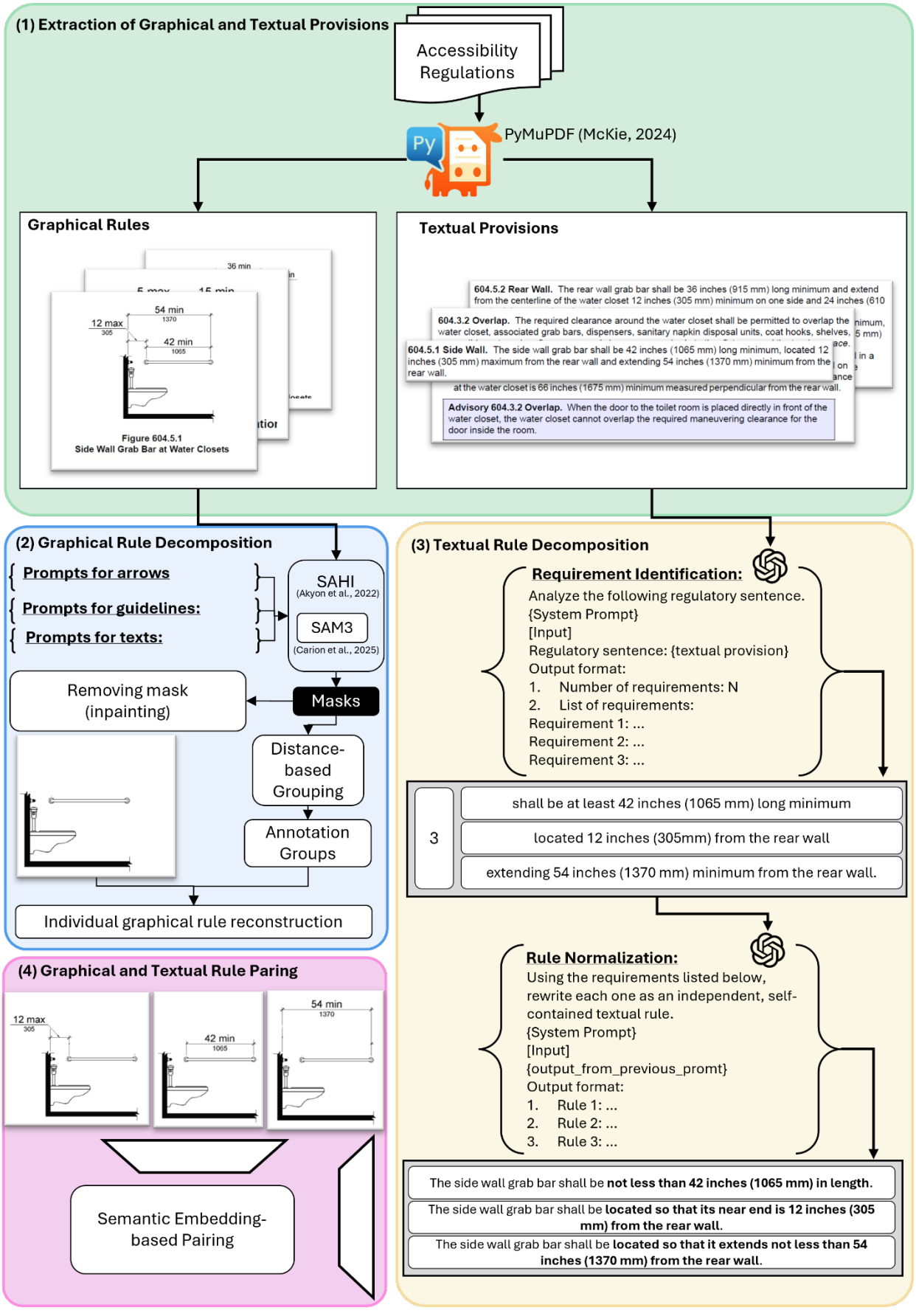


Figure 1: Overview of the proposed framework for graphical rule decomposition and pairing

and (4) all relevant measurement references and conditions are explicitly stated in a concise manner.

An instruction-only prompting strategy is adopted throughout both stages, without using any few-shot examples. The output of the requirement identification stage is provided as input to the rule normalization stage, enabling progressive refinement of the textual decomposition while preserving the original normative meaning.

Graphical and textual rule pairing

Finally, each reconstructed graphical rule is paired with its corresponding textual rule. To achieve this, semantic embeddings are computed using a pretrained CLIP model (ViT-B/32), which maps images and text into a shared embedding space.

Rule pairing is performed by maximizing cosine similarity between graphical and textual embeddings. The resulting graphical–textual rule pairs can be directly used for downstream interpretation by MLLMs.

Experimental design

The experimental evaluation focuses on assessing whether the proposed method correctly decomposes composite graphical rules into atomic, rule-level graphical rules that each correspond to a single regulatory requirement.

Based on the proposed method, this section describes the experimental design used to evaluate its effectiveness.

The evaluation is designed to assess two key aspects: (1) whether the decomposed diagrams represent a single regulatory requirement, and (2) whether they preserve the same regulatory meaning as the ground truth.

To this end, an MLLM-as-a-judge framework (D. Chen et al., 2024) is employed as the primary evaluation method. The framework is used to compare automatically decomposed graphical rules with manually decomposed ground-truth diagrams and to assess whether the predicted results preserve the intended regulatory meaning.

Ground truth generation

For each composite graphical rule, a set of ground-truth rule-level diagrams was manually generated using image editing software. Each ground-truth diagram isolates exactly one regulatory requirement by retaining only the annotation elements relevant to that requirement, together with the minimal graphical context necessary for interpretation.

In total, ground truth was constructed from 48 composite graphical rules, resulting in 134 individual graphical rules. The manual decomposition process was guided by the semantics of the corresponding textual provisions to ensure that each ground-truth diagram represents a single, independent regulatory rule without ambiguity or overlap.

These manually decomposed rule-level diagrams serve as the reference standard for evaluating the correctness of graphical rule decomposition.

Baseline definition

To contextualize the performance of the proposed method, a prompt-based annotation filtering baseline is employed. The baseline adopts a multi-step prompting strategy using a multimodal generative model to produce rule-specific diagrams by selectively suppressing irrelevant annotations in the original composite graphical rule.

First, the model is prompted to identify and enumerate the individual regulatory requirements encoded in the input composite diagram. These extracted requirements are organized into a structured list and serve as the basis for subsequent annotation filtering.

Next, for each identified regulatory requirement, the original composite diagram is provided again, and the model is instructed to remove all annotation elements that are unrelated to the target requirement, while preserving the original geometric drawing and the annotation elements corresponding to the selected rule. In this step, the model is explicitly guided to retain only the dimensional markers, symbols, and textual annotations necessary to convey the target regulatory requirement, and to suppress all others.

As a result, the baseline produces a set of rule-specific diagrams modified from the original image, each intended to express a single, self-contained regulatory requirement.

This baseline reflects recent prompt-driven approaches to diagram editing and controlled visual abstraction using generative multimodal models.

Evaluation metrics

The evaluation employs an MLLM-as-a-judge framework to assess rule-level semantic correctness.
For each decomposed graphical rule, the MLLM is first asked to determine whether the diagram represents exactly one regulatory requirement.
Next, the MLLM compares the automatically decomposed graphical rule with the manually decomposed ground-truth diagram and determines whether both convey the same regulatory requirement.

In addition to this semantic evaluation, we incorporate objective segmentation-level metrics to assess the quality of annotation isolation. Specifically, the mask-level intersection of union (IoU) is computed against manually annotated ground truth to quantitatively evaluate the accuracy of the extracted annotation elements.

Results and discussion

Experiment results

Table 1 summarizes the evaluation results obtained using the MLLM-as-a-judge framework.

Among the images generated by the proposed method, 84% were judged to contain exactly one regulatory requirement. In contrast, all images produced by the prompt-based baseline were classified as single-requirement diagrams. However, when comparing the regulatory semantics conveyed by each generated image with the manually decomposed ground truth, the proposed

Table 1: Overall experiment results from MLLM-as-a-Judge

MLLM-as-a-Judge	Proposed Method	Baseline
Has a single requirement	84 %	100 %
Same as ground truth	82 %	78 %

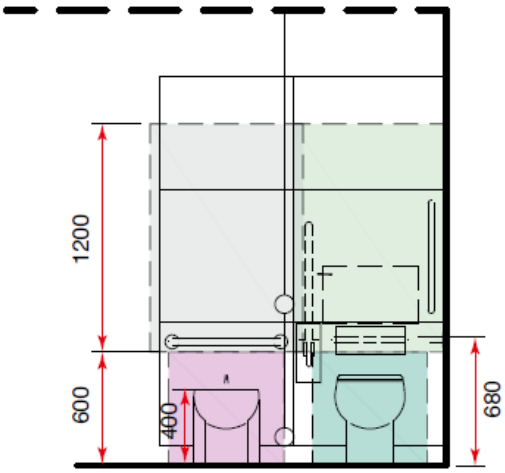
method achieved a higher agreement rate, showing a modest improvement (+4%) compared to the baseline.

In addition to semantic evaluation, we further assessed the quality of annotation isolation using segmentation-level metrics. The proposed method achieved a mask-level IoU of 88%, highlighting that segmentation quality remains a key bottleneck, and further improvements are required to achieve more reliable rule-level decomposition.

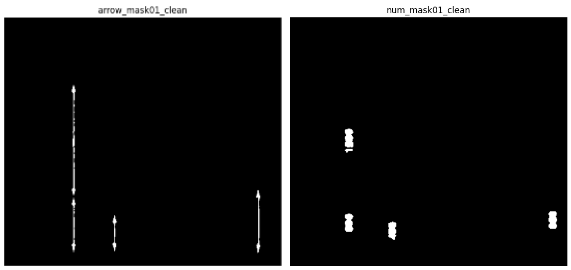
Discussions

To provide a clearer understanding of how the proposed method operates, we first present an intermediate example illustrating the decomposition process.

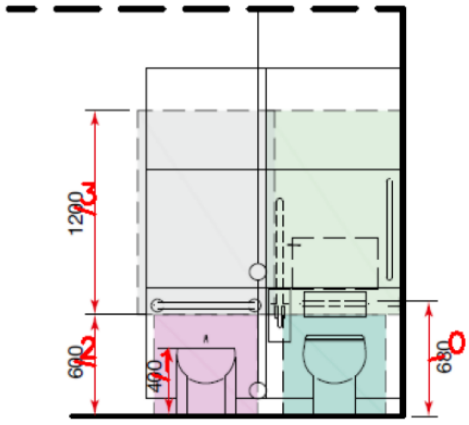
Figure 2 shows the intermediate results of the proposed pipeline, including annotation segmentation, clustering, and reconstruction.



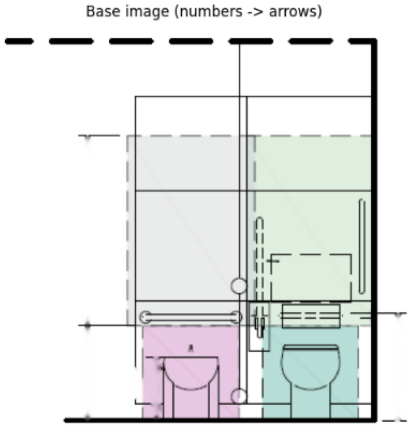
(a) Original composite graphical rule



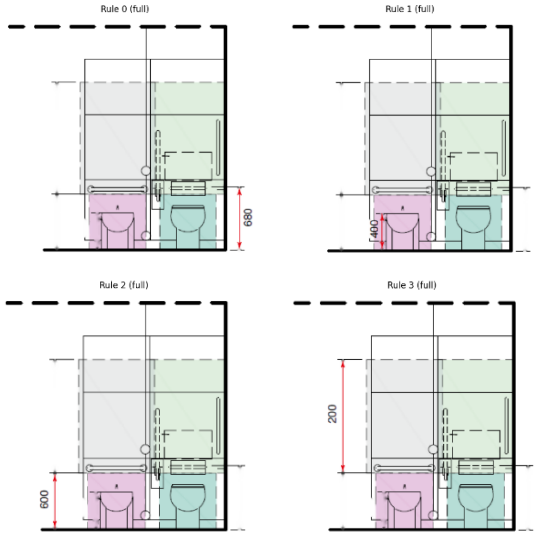
(b) Annotation segmentation results



(c) Grouping



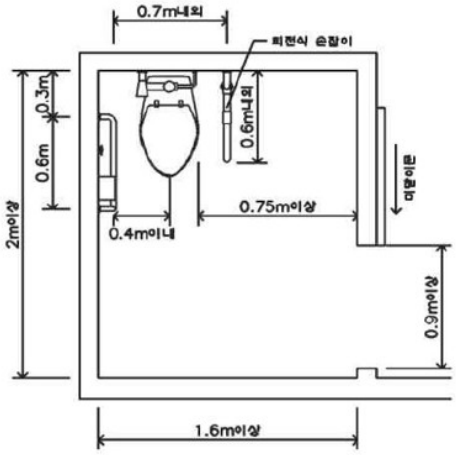
(d) Base image



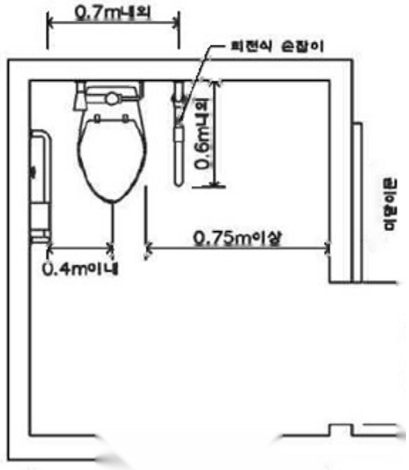
(e) Rule decomposition results

Figure 2: Intermediate results of the proposed pipeline for graphical rule decomposition. (a) Original composite graphical rule. (b) Annotation segmentation results generated by SAM3. (c) Grouping of annotation elements into rule-level clusters based on spatial proximity. (d) Base image obtained by removing annotation elements through inpainting. (e) Final rule-level graphical unit reconstructed by overlaying clustered annotations onto the base image.

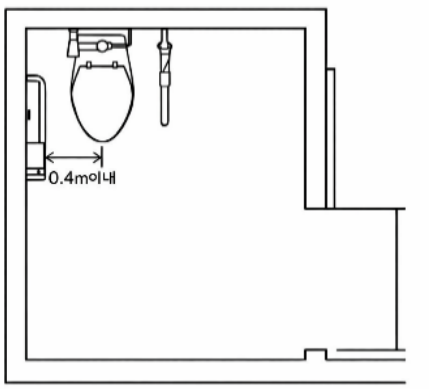
The experimental results indicate that the proposed method is less robust than the baseline in strictly decomposing composite diagrams into visually clean single-rule images. This limitation is particularly evident in challenging input cases, such as the example shown in Figure 3-(a), where failures occur during the annotation segmentation and grouping stages.



(a) A composite graphical rule for the decomposition



(b) A decomposition output from the proposed method



(c) A decomposition output from the baseline method

Figure 3: Example of a composite graphical rule (a) where the proposed method (b) fails to isolate single requirements, while the baseline succeeds in isolating single requirements and preserving the regulatory meaning.

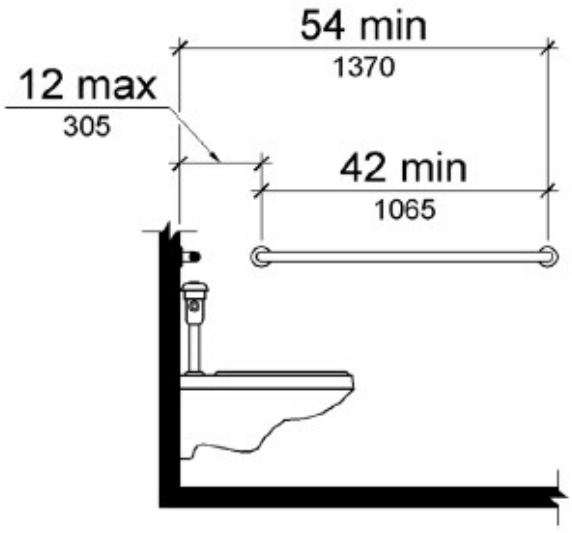
As shown in Figure 3, the baseline successfully isolates the regulation specifying that the spacing between the handrail and the toilet should be approximately 0.4 m, yielding a clear, self-contained graphical rule. In contrast, the diagram produced by the proposed method retains residual non-target graphical elements in the base drawing, indicating that unrelated annotations were not fully removed. Consequently, the resulting image does not form a fully isolated single graphical rule.

This failure mode is consistently observed in source images with low resolution or very small annotation elements, where reliable segmentation and grouping become particularly difficult.

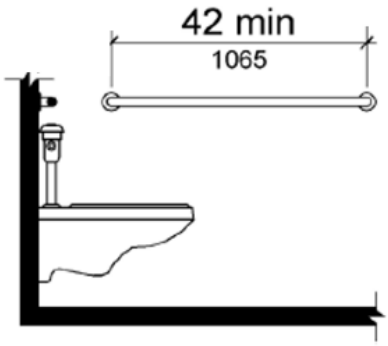
Despite this limitation, the proposed method demonstrates a complementary strength in preserving the regulatory semantics of the original graphical rules, as shown in Figure 4, where prompt-based baselines result in semantic distortion.

The diagram generated by the proposed method preserves the essential spatial relationships and numerical constraints required to convey the target regulation, despite minor visual imperfections introduced during the annotation isolation process. In contrast, the baseline produces a visually clean but semantically incorrect diagram that conveys a substantially different regulatory interpretation from the ground truth.

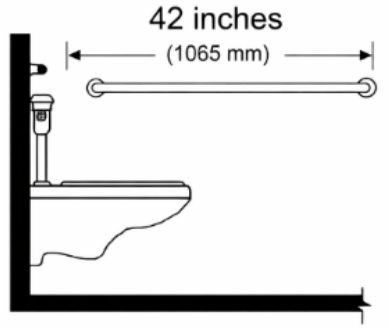
This tendency is also supported by the “same as ground truth” metric in Table 1, where the proposed method achieves 82%, slightly outperforming the baseline (78%). Although the numerical improvement is modest, it indicates that the proposed method more reliably preserves the intended regulatory meaning.



(a) A composite graphical rule for the decomposition



(b) A decomposition output from the proposed method



(c) A decomposition output from the baseline method

Figure 4: Example of a composite graphical rule (a) where the baseline (c) fails to preserve the regulatory meaning, while the proposed method (b) maintains semantic consistency with the ground truth.

Taken together, these results reveal a clear trade-off between the two approaches. The baseline is more effective at producing visually clean, strictly isolated single-rule diagrams, whereas the proposed method is more reliable at preserving regulatory semantics by maintaining the original geometric structure.

This behavior can be attributed to the reliance of the baseline on generative image synthesis, which reconstructs the diagram from scratch. While this enables strong visual simplification, it also increases susceptibility to hallucinations and unintended misinterpretation of regulatory constraints, leading to semantic drift.

In contrast, the proposed method operates by decomposing and reconstructing the original diagram, which can compromise visual cleanliness in some cases but helps prevent semantic distortion.

Conclusions

This paper presented a zero-shot framework for decomposing composite graphical rules into atomic graphical rules and pairing them with corresponding textual rules. By leveraging text-prompt-guided zero-shot segmentation with SAM3, context-aware inpainting for faithful graphical reconstruction, and instruction-only prompt chaining for textual rule decomposition, the proposed method enables rule-level understanding of graphical regulations without requiring labeled training data.

Experimental results suggest that the proposed method exhibits a trade-off compared to prompt-based image generation approaches. While it is less robust in enforcing visually clean single-rule isolation, it tends to better preserve regulatory intent and semantic consistency with the ground truth.

The evaluation combines semantic assessment using an MLLM-as-a-judge framework with segmentation-level metrics (e.g., IoU), but remains limited in several aspects, including reliance on a single MLLM-based evaluation, the absence of human validation and inter-rater reliability analysis, and a relatively small dataset.

Future work will incorporate human expert evaluation, in which domain experts define ground-truth interpretations of graphical rules and assess whether the decomposed outputs correctly match these references. Such an evaluation will enable more reliable validation of the proposed method.

Moreover, the current study focuses on dimensional technical drawings in which regulatory requirements are primarily conveyed through distance-based annotations, specifically within accessibility regulations for disabled restroom facilities. While this scope reflects a common form of graphical regulation, the experimental validation remains limited to a single facility type. In particular, the present study does not yet account for variations in drawing configurations. Therefore, extending the framework to a wider range of facility conditions and drawing variations, as well as to other types of graphical rules, remains for future work.

Acknowledgments

This work was supported by the Hans Fischer Senior Fellowship program at the Technical University of Munich – Institute of Advanced Studies (TUM-IAS) and the TUM Georg Nemetscheck Institute Artificial Intelligence for the Built World (GNI).

References

Amor, R. & Dimyadi, J. (2021) The promise of automated compliance checking. Developments in the Built Environment, 5, 100039.

Carion, N., Gustafson, L., Hu, Y.-T., Debnath, S., Hu, R., Suris, D., Ryali, C., Alwala, K.V., Khedr, H., Huang, A., Lei, J., Ma, T., Guo, B., Kalla, A., Marks, M., Greer, J., Wang, M., Sun, P., Rädle, R., Afouras, T., Mavroudi, E., Xu, K., Wu, T.-H., Zhou, Y., Momeni, L., Hazra, R., Ding, S., Vaze, S., Porcher, F., Li, F., Li, S., Kamath, A., Cheng, H.K., Dollár, P., Ravi, N., Saenko, K., Zhang, P. & Feichtenhofer, C. (2025) SAM 3: Segment Anything with Concepts. arXiv preprint, arXiv:2511.16719.

Chen, D., Chen, R., Zhang, S., Liu, Y., Wang, Y., Zhou, H., Zhang, Q., Wan, Y., Zhou, P. & Sun, L. (2024) MLLM-as-a-Judge: Assessing Multimodal LLM-as-a-

Judge with Vision-Language Benchmark. arXiv preprint, arXiv:2402.04788.

Chen, N., Lin, X., Jiang, H. & An, Y. (2024) Automated Building Information Modeling Compliance Check through a Large Language Model Combined with Deep Learning and Ontology. Buildings, 14(7), 1983.

Dori, D. & Velkovitch, Y. (1998) Segmentation and Recognition of Dimensioning Text from Engineering Drawings. Computer Vision and Image Understanding, 69, p.pp.196–201

Goswami, K., Mathur, P., Rossi, R., Dernoncourt, F. (2025) PlotEdit: Natural Language-Driven Accessible Chart Editing in PDFs via Multimodal LLM Agents. arXiv preprint, arXiv:2501.11233.

Hettiarachchi, H., Dridi, A., Gaber, M.M., Parsafard, P., Bocaneala, N., Breitenfelder, K., Costa, G., Hedblom, M., Juganaru-Mathieu, M., Mecharnia, T., Park, S., Tan, H., Tawil, A.-R.H. & Vakaj, E. (2025) CODE-ACCORD: A Corpus of building regulatory data for rule generation towards automatic compliance checking. Scientific Data 12, 170

Hou, Y., Giledereli, B., Tu, Y. & Sachan, M. (2025) Do Vision-Language Models Really Understand Visual Language? arXiv preprint, arXiv:2410.00193.

Iversen, O. & Huang, L. (2026) Leveraging large language models for BIM-based automated compliance checking. Automation in Construction, 182, 106707.

Jeong, D., Kang, D., Park, J., Lee, H. & Paik, J. (2025) Structure-Preserving Zero-Shot Image Editing via Stage-Wise Latent Injection in Diffusion Models. arXiv preprint, arXiv:2504.15723.

Khan, M.T., Yong, Z., Chen, L., Tan, J.M., Feng, W. & Moon, S.K. (2025) Automated Parsing of Engineering Drawings for Structured Information Extraction Using a Fine-tuned Document Understanding Transformer. arXiv preprint, arXiv:2505.01530.

Kim, Y., Borrmann, A. & Lee, G. (2025) A Preliminary Study on Design Rule Derivation from Graphical Representations Using Multimodal Large Language Models. In: Proceedings of the 2025 European Conference on Computing in Construction, Porto, Portugal. European Council on Computing in Construction.

Li, Y., 2026. An Adaptive LLM-Based Agent Framework with Self-Extending Tool Library for Automated Building Code Compliance Checking. Master's thesis, Technical University of Munich, Munich, Germany.

Ma, N., Labib, R., Amor, R., Chong, A., Fan, C., Forth, K., Fu, X., Fuchs, S., Hong, T., Klimenkova, N., Koo, J., Li, S., McCullough, S.T., Park, J.Y., Shraga, R., Yoon, S., Zhang, L. & Zhang, Y. (2026) Ten questions concerning large language models (LLMs) for building applications. Building and Environment, 291, p.114260.

Madireddy, S., Gao, L., Din, Z.U., Kim, K., Senouci, A., Han, Z. & Zhang, Y. (2025) Large Language Model-Driven Code Compliance Checking in Building Information Modeling. Electronics 14, 2146.

Maupou, C., Yang, Y., Fodop, G., Qie, Y., Migliorini, C., Mehdi-Souzani, C. & Anwer, N. (2024) Automatic raster engineering drawing digitisation for legacy parts towards advanced manufacturing. Procedia CIRP 129, 234–239.

Ministry of Health and Welfare (MOHW) (2023) The enforcement rule of the act on promotion for persons with disabilities, the elderly, and pregnant women. Appendix 1: detailed standards for structure and materials of convenience facilities. National Law Information Center of Korea.

Picard, C., Edwards, K.M., Doris, A.C., Man, B., Giannone, G., Alam, M.F. & Ahmed, F. (2025) From concept to manufacturing: evaluating vision-language models for engineering design. Artificial Intelligence Review, 58, 288.

Lin S. & Ting C. (1997) A new approach for detection of dimensions set in mechanical drawings. Pattern Recognitition Letters, 18, p.pp.367–373.

U.S. Department of Justice (DOJ). 2010. ADA Standards for Accessible Design. U.S. Department of Justice. U.S. Department of Justice, Washington, DC

Wu, C., Yin, S., Qi, W., Wang, X., Tang, Z. & Duan, N. (2023) Visual ChatGPT: Talking, Drawing and Editing with Visual Foundation Models. arXiv preprint, arXiv:2303.04671.

Yang, D., Zhang, L., Yue, Z., Chen, L., Xu, Y., Wang, W. & Jin, Q. (2025) ChartM3: Benchmarking Chart Editing with Multimodal Instructions. https://doi.org/10.48550/arXiv.2507.21167

Ying, H. & Sacks, R. (2024) From Automatic to Autonomous: A Large Language Model- driven Approach for Generic Building Compliance Checking. arXiv preprint. In Proceedings of the CIB W78 Conference.

Zhang, W., Joseph, J., Yin, Y., Xie, L., Furuhata, T., Yamakawa, S., Shimada, K. & Kara, L.B (2023) Component segmentation of engineering drawings using Graph Convolutional Networks. Computers in Industry, 147, p.103885