MiCRoadDiff: AI-driven generative design of on-site haul road for modular constructions using fine-tuned diffusion models

Fangzheng Li[1], Rongyan Li[2], and Hung-Lin Chi[3]

[1]Department of Building and Real Estate, Faculty of Construction and Environment, The Hong Kong Polytechnic University, Hong Kong
[2]Department of Building and Real Estate, Faculty of Construction and Environment, The Hong Kong Polytechnic University, Hong Kong
[3]Department of Building and Real Estate, Faculty of Construction and Environment, The Hong Kong Polytechnic University, Hong Kong
fangzheng.li@connect.polyu.hk, rongyan.li@connect.polyu.hk, hung-lin.chi@polyu.edu.hk

Abstract

Modular construction projects rely heavily on large trucks to transport prefabricated modules, making the design of on-site haul roads critical for ensuring safe and efficient logistics operations. However, the manual design process suffers from inefficiencies. To address this issue, this study proposes MiCRoadDiff, a diffusion model-based intelligent approach for modular construction on-site haul road layout generation. MiCRoadDiff, adopts a pre-trained Stable Diffusion model as an intelligent design backbone. LoRA is employed to achieve domain adaptation by injecting construction-specific knowledge, design experience, and implicit road design rules into the generative model. In addition, multiple ControlNet modules are introduced to enforce geometric and road layout constraints. A comprehensive evaluation framework that incorporates image fidelity and layout validity is proposed. Experimental results demonstrate that MiCRoadDiff exhibits strong domain adaptation capability and can rapidly generate diverse, high-quality haul road layout alternatives that satisfy practical engineering requirements. By integrating multi-source conditional information, it can robustly adapt to varying site conditions, generating coherent and feasible haul road layouts. The proposed framework significantly improves the efficiency and intelligence of construction site layout design workflows.

Keywords –

Diffusion models; Generative design; On-site haul road design; Finetuning; Modular constructions.

1 Introduction

Modular Integrated Construction (MiC) [1] is an innovative construction approach where modules are prefabricated in factories and then assembled on-site. In recent years, the Government of the Hong Kong Special Administrative Region (HKSAR) has actively promoted MiC [2] to tackle key challenges in the local construction industry, including labor shortages, site limitations, and the growing demand for environmental sustainability. The modules are transported to the construction site by trucks, entering through the main gate, being placed at designated unloading points, and then exiting the site. On-site haul roads play a critical role in ensuring the safety and efficiency of truck transportation [3]. Proper design of haul roads can significantly enhance logistic efficiency, safety, and ease of operation for truck drivers, thereby minimizing potential damage to the modules. However, traditional manual design processes face several limitations, such as low efficiency and the necessity for repetitive adjustments, which hinder overall productivity and effectiveness.

Researchers have explored classical optimization methods, including greedy algorithms [4], genetic algorithms [5], and Monte Carlo techniques [6], in various aspects of road planning. While genetic algorithms demonstrated potential in exploring design spaces, they were limited to producing near-optimal solutions and fell short of capturing the complex nature of the layout design process. Mixed-integer linear programming [3] was also employed to achieve exact solutions, but it suffered from prolonged computational times.

In recent years, advances in deep learning have significantly transformed automated engineering and construction design, enabling more efficient and intelligent design workflows. Among these approaches, Convolutional Neural Networks (CNNs) have played a foundational role in processing spatial information and extracting meaningful features from visual data [7,8]. Their strong capability in analyzing complex spatial patterns has facilitated the automation of design tasks that traditionally relied on extensive manual effort. Subsequently, deep generative models based on convolutional architectures, particularly Generative Adversarial Networks (GANs) [9], have attracted considerable attention. Owing to their adversarial training mechanism, GANs are capable of producing designs that not only exhibit high visual realism but also satisfy certain engineering constraints. In the construction domain, GAN-based methods have been applied to layout generation [10], and defect detection [11]. Nevertheless, GAN models often suffer from training instability and limited output diversity [12], which restrict their scalability and robustness in practical engineering applications. Moreover, the rationality of a single generated design cannot be guaranteed, making model optimization and deployment challenging. In addition, existing studies on road design frequently depend on expert-based qualitative evaluation, which lacks standardized and objective metrics. Recent progress in state-of-the-art image generation techniques, particularly diffusion models [13,14] and their associated control mechanisms [15], which provide improved training stability, enhanced diversity of generated results, offers a promising pathway to overcome these limitations.

To address the aforementioned limitations, this paper presents MiCRoadDiff, a diffusion model-based intelligent approach for modular construction on-site haul road layout generation conditioned on construction plans and textual design conditions. Building upon the exceptional image synthesis capabilities of Stable Diffusion, this study aims to learn from designers' expertise to rapidly generate initial design proposals, serving as a valuable tool for assisting and streamlining the design process. This approach significantly contributes to the automation and efficiency of construction design workflows. This study integrates both textual and image design conditions to enable automated design. By utilizing the ControlNet architecture, the generated designs demonstrate strong alignment with the input conditions, facilitating the creation of customized and precise layout solutions. Domain adaptation technology, LoRA, is incorporated to improve model performance under limited data conditions. This effectively addresses the challenges of data scarcity frequently encountered in construction engineering applications. A comprehensive evaluation strategy is developed to assess the generated results, including the image quality aspect (FID, SSIM, and PSNR) and haul road layout aspect (transportation cost, accessibility, and number of corners). This approach aids in model optimization while ensuring both visual realism and compliance with engineering requirements.

2 Related Works

2.1 Deep Learning Methods for Site Layout Generation

Currently, there is limited research on using deep learning methods for site layout generation. Li et al. [10] utilized GAN-based methods to design tower crane locations for high-rise buildings, which overcomes complexity limitations and facilitates layout planning decision-making process. GAN models were adopted mostly for building layouts and floor layouts design [16]. However, GAN-based approaches encounter challenges such as training instability and limited output diversity. Additionally, existing research has yet to effectively address the critical issue of limited datasets, a challenge in the construction domain.

2.2 Diffusion Models in Image Generation

Diffusion models [13,14] are a powerful class of generative models that have recently gained significant attention for their ability to produce high-quality and diverse outputs. By employing an iterative denoising process, these models excel at capturing and modeling complex data distributions, making them particularly effective for tasks such as image synthesis and video generation.

To improve controllability and flexibility compared to GAN-based models, a multi-conditional model for residential floor plan design was developed using a diffusion model [17]. This approach demonstrated high-quality, diverse, and controllable results through experimental validation. An image-prompt diffusion model showed exceptional feature extraction capabilities and design effectiveness in shear wall design [18].

3 Methodology

3.1 Problem Formulation

MiCRoadDiff aims to automatically generate feasible and efficient haul road layouts for modular construction sites under complex spatial and operational constraints. Given a fixed construction site layout, and design conditions, the objective is to generate a temporary haul road configuration that ensures connectivity, safety, and constructability while minimizing conflicts with non-movable elements. Formally, the task is defined as a conditional image generation problem. Let A denote the site layout constraints (e.g., site boundary and building footprints), F represent fixed facilities (e.g., tower crane bases and site gates), and T indicate textual design conditions (e.g., road width and minimum turning radius). The goal is to learn a generative model G that produces a haul road layout R: $R = G(A, F, T)$, where R satisfies geometric constraints imposed by A and F, and design logic encoded in T.

3.2 Overall Framework of MiCRoadDiff

The overall architecture of MiCRoadDiff is illustrated in Figure 1. The framework consists of three main components: (1) multi-layer conditional encoding, (2) conditional diffusion-based generation with ControlNet and LoRA, and (3) post-generation evaluation. First, for a construction site layout to be designed, multi-conditional maps (including boundary map and fixed facility map) are constructed, along with textual design requirements. These inputs are encoded and fed into a fine-tuned Stable Diffusion model. Then, the model leverages LoRA to perform experience-based haul road design, while ControlNet ensures compliance with the specified rigid constraints. Finally, the ideal generated layouts are evaluated using computer vision-related metrics and road design-specific models.

The core idea of MiCRoadDiff is to achieve flexible layout designs while satisfying rigid design constraints. Multiple condition maps are enforced through different ControlNet modules, while generalizable haul road design principles are learned via a lightweight LoRA adaptation of a pre-trained Stable Diffusion model.

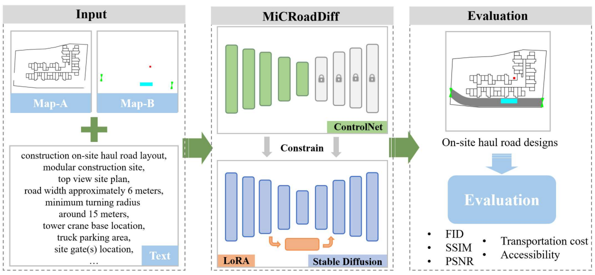

Figure 1. The framework of the research

3.3 Stable Diffusion

Stable Diffusion is a latent diffusion-based generative model that synthesizes images by iteratively denoising a latent representation guided by conditional information. Instead of performing diffusion directly in pixel space, Stable Diffusion operates in a compressed latent space obtained via a variational autoencoder (VAE), which significantly reduces computational cost while preserving high-level semantic structure. During training, Gaussian noise is progressively added to latent variables, and a denoising U-Net is trained to predict and remove this noise step by step. At inference time, the model starts from random noise and gradually reconstructs an image that aligns with the provided conditions.

In the context of MiCRoadDiff, Stable Diffusion serves as the core generative engine responsible for synthesizing plausible haul road layout designs. Rather than explicitly encoding construction rules or geometric algorithms, the diffusion backbone provides strong priors for producing continuous, coherent, and visually structured layouts.

Among available Stable Diffusion variants [19], MiCRoadDiff adopts Stable Diffusion XL (SDXL) [20] as the backbone model. SDXL introduces a larger U-Net architecture, a higher-capacity text encoder, and a dual-stage conditioning mechanism compared to earlier versions such as SD v1.5. These improvements enable SDXL to better capture complex spatial relationships and long-range dependencies within an image, which are critical for construction site layout generation where global topology and local geometry must be jointly satisfied.

3.4 LoRA and ControlNet

Low-Rank Adaptation (LoRA) [21] is a parameter-efficient fine-tuning technique designed to adapt large pre-trained models to downstream tasks without updating the full set of model parameters. Instead of directly modifying the original weight matrices in the neural network, LoRA introduces a pair of low-rank matrices that approximate the weight updates. During training, only these low-rank matrices are optimized, while the original backbone parameters remain frozen. This significantly reduces memory consumption and training cost, while preserving the generalization capability of the pre-trained model. In MiCRoadDiff, LoRA plays a critical role in learning domain-specific haul road design logic that is not explicitly encoded in the condition informations. LoRA can capture higher-level design principles, including road continuity, transportation efficiency, and typical geometric patterns of temporary haul roads in modular construction sites. By integrating LoRA with the SDXL backbone, MiCRoadDiff leverages the strong generative prior of a large diffusion model while maintaining efficient and controllable domain adaptation.

ControlNet [15] is a conditional control architecture designed to extend large diffusion models with fine-grained and spatially aligned guidance, without disrupting their original generative capabilities. It introduces an auxiliary neural network that mirrors the structure of the diffusion model' s denoising U-Net and injects condition-specific features into the generation process through zero-initialized convolution layers. This design ensures that, at the beginning of training or inference, the ControlNet branch has no influence on the output, thereby preserving the behavior of the pre-trained backbone, while gradually enabling strong conditional control as learning progresses.

In MiCRoadDiff, ControlNet is responsible for enforcing rigid spatial constraints that must be strictly preserved in construction site road layout generation. These constraints include site boundaries, building footprints, and fixed entities such as tower crane bases, ideal truck parking areas, and site entrances. Such elements are non-negotiable in practice and cannot be reliably handled by text prompts or learned implicitly through data-driven adaptation alone. MiCRoadDiff employs multiple ControlNet modules, each associated with a specific type of condition map.

3.5 Multi-conditional Image Generation

MiCRoadDiff performs haul road layout synthesis through a multi-conditional image generation process that jointly integrates multiple spatial condition maps and structured textual input within a unified diffusion framework, as shown in Figure 2. This design enables the model to handle heterogeneous constraints commonly encountered in construction site planning, where geometric feasibility, operational requirements, and design intent must be satisfied simultaneously.

Each condition map encodes a single type of constraint. Map-A encodes rigid and non-movable geometric boundaries, including site boundaries and building boundaries. Map-B encodes fixed entities including tower crane bases, ideal truck parking areas, and site entrances. This multi-conditional formulation allows MiCRoadDiff to explore diverse feasible solutions within a well-defined constraint space.

In addition to image-based conditions, MiCRoadDiff incorporates structured textual input to specify general design requirements that are difficult to encode spatially. The textual prompt follows a consistent and interpretable structure, following the format: "construction on-site haul road layout, modular construction site, top view site plan, road width approximately [Value] meters, minimum turning radius around [Value] meters, tower crane base location, truck parking area, site gate(s) location, road connects [Facilities], avoid [Obstacle]" This format ensures the generated images remain contextually and environmentally accurate by integrating both spatial and climatic factors. For example: *"construction on-site haul road layout, modular construction site, top view site plan, road width approximately 9 meters, minimum turning radius around 15 meters, tower crane base location, truck parking area, site gate(s) location, road connects site gate(s) and truck parking area, avoid buildings and crane base."* This structured prompting framework enables effective conditioning across haul road design intent. Through the joint use of multiple condition maps and structured textual input, MiCRoadDiff achieves controllable, robust, and flexible multi-conditional image generation tailored to the complex requirements of haul road layout design for MiC.

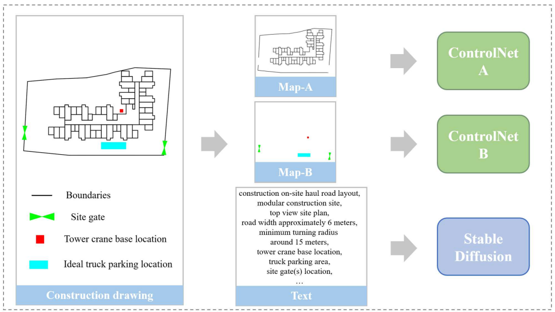

Figure 2. Multi-condition encoding for image generation

3.6 Post-generation Evaluation

To evaluate the quality and perceptual fidelity of the generated road layout images, three established metrics FID, SSIM, and PSNR are used. FID assesses realism and perceptual similarity, while SSIM and PSNR measure structural consistency and reconstruction quality. Besides, domain-specific metrics including transportation cost and accessibility, are used to evaluate layout performance in terms of efficiency and safety. Transportation cost is calculated as the product of the road centerline length and the unit construction cost. Accessibility evaluates how the designed road connects site components including gates and truck parking locations.

4 Experiment and Results

4.1 Experimental details

To train and evaluate LoRA, a dedicated dataset was constructed to reflect realistic modular construction site conditions. Ten real-world projects were sourced from renowned contractors in Hong Kong, resulting in the creation of a total of 10 data. Each data consists of image and corresponding textual description, enabling joint image–text conditioning during training. Textual descriptions are manually curated and aligned with each image sample. The constructed dataset highlights diversity in road types (one-way or two-way), road width (6m-10m) and turning radius (12m-18m), gate configurations (single or multiple), modular building shapes, and boundary shapes. This diversity allows the model to learn transferable haul road design logic rather than memorizing specific layouts. During training, the SDXL backbone parameters were kept frozen, and LoRA modules were inserted into the attention layers of the U-Net architecture. Only the low-rank adaptation parameters were updated. No augmentation techniques were used during training. LoRA training and inference were performed on a workstation equipped with 13th Gen Intel(R) Core (TM) i7-13700KF@3.40 GHz CPU and NVIDIA GeForce RTX 4090 GPU. AdamW8bit [22] was used to optimize the low-rank decomposition of a subset of weights to minimize this loss. The batch size was set to 1. The learning rate for U-Net was set to $1e^{-4}$, while the learning rate for the text encoder network was $1e^{-5}$.

4.2 Training results of LoRA

The LoRA adaptation was trained for a total of 20 epochs, each consisting of 50 training steps. The evolution of the training loss over the entire training process is illustrated in Figure 3. The loss value shows a gradual decreasing trend with increasing training iterations, indicating stable convergence of the LoRA training process. The results generated by the SDXL model under the same text prompt under different training epochs were compared, as shown in Figure 4.

From the 5th to the 10th epoch, the generated results exhibit a clear shift toward a simplified semantic style suitable for site layout representation. By the 15th epoch, most critical layout components (haul road, tower crane base, truck parking area, site gate, and building boundaries) are correctly generated except site boundaries. While the 20-epoch model produces complete and well-organized layouts with road networks that satisfy engineering design requirements.

Overall, the LoRA-enhanced model shows a clear improvement in semantic alignment with modular construction on-site haul road planning intent. It also shows the effectiveness of the LoRA-based adaptation in capturing domain-specific haul road design logic for modular construction sites rather than memorizing specific site configurations.

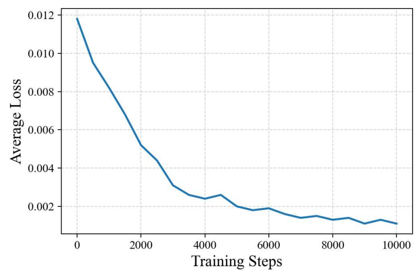

Figure 3. Training loss curves of LoRA

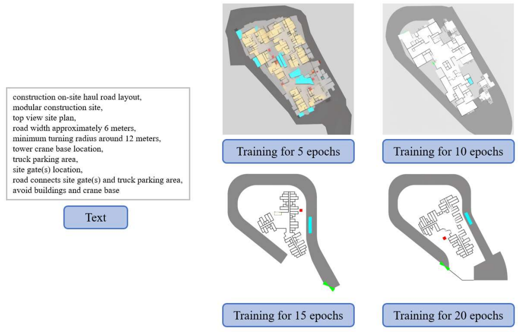

Figure 4. Comparison of model performance under different training epochs

4.3 Design results of MiCRoadDiff

The MiCRoadDiff model was built upon SDXL as the base diffusion model and incorporated a LoRA module trained for 20 epochs. In addition, two ControlNet modules were employed for image-based constraint control, with their parameters carefully tuned through multiple iterations to achieve optimal performance.

The MiCRoadDiff model was evaluated using new project cases that were not included in the training dataset. The generated haul road layouts were assessed through both qualitative and quantitative evaluations to comprehensively examine the model's generalization capability and design performance.

The project used in this study is a modular residential project located in Hong Kong. After simplification and semantic processing, the resulting design representations include condition image, design text, and corresponding haul road layouts, as shown in Figure 5. Black lines represent building boundaries and site boundaries, red squares denote tower crane locations, blue rectangles indicate truck parking areas with dimensions corresponding to real truck sizes, green symbols represent site gates, and grey regions denote haul roads, whose widths and turning radius reflect actual design dimensions. MiCRoadDiff was applied to design the haul road layout for this project, and the generated results are shown in Figure 6.

In this study, input construction drawings were pre-processed to standardize spatial resolution, enabling precise conversion of textual constraints into pixel units. Dimensional constraints are embedded into structured prompts, guiding the model toward geometrically plausible outputs during diffusion. Additionally, ControlNet ensures road layouts align with site geometry, preventing violations of spatial constraints. This ensures that the textual dimension can be accurately translated into spatial pixels.

From the design results, it can be observed that the model is capable of producing diverse haul road layout alternatives for designers' selection. The generated layouts form relatively complete site plans with high image quality and clarity. Moreover, the generated haul roads satisfy the prescribed design conditions and are feasible from an engineering perspective.

Layout (a) was selected as the final design solution for quantitative analysis, with the results shown in Figure 7. The generated layout achieved favorable FID, SSIM, and PSNR scores, indicating high image clarity and strong structural similarity to the ground truth. In addition, the total road length of the designed haul road is 128 m, with an estimated construction cost of 385,101 HKD. The layout exhibits good connectivity by linking two site gates and effectively serving the designated truck parking areas. However, the presence of three turning corners may slightly increase driving difficulty in practice.

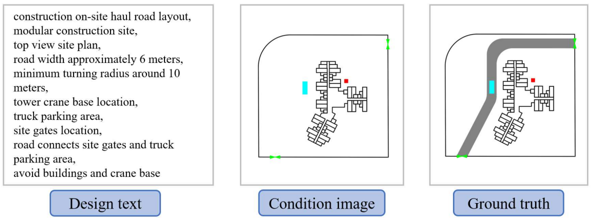

Figure 5. Design conditions and real on-site haul road layout of the case project

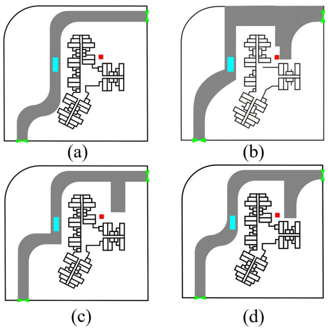

Figure 6. Design results of MiCRoadDiff

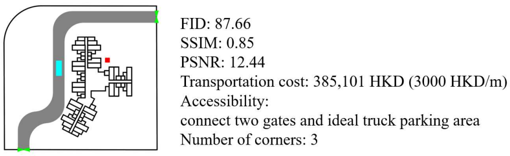

Figure 7. Evaluations on layout design result

5 Discussion

5.1 Ablation test of MiCRoadDiff

To investigate the contribution of different modules to the generation performance of MiCRoadDiff, an ablation study was conducted. The experimental results are presented in Figure 8 and Table 1.

As shown in Figure 8, when the ControlNet modules are removed, the model no longer produces engineering drawings with fixed spatial constraints. Instead, it generates layouts with varying building and site shapes, randomly arranged tower crane locations, truck parking areas, and site gates. Although these results exhibit a certain degree of creativity, the model still maintains a consistent visual style and the design experience learned during training. When the LoRA module is removed, the model shifts toward a more photorealistic visual style. While the overall building and site boundaries are still preserved, the model fails to correctly interpret the design conditions specified in the text prompts, such as road width, turning radius, and connectivity requirements. When only the SDXL baseline model is used for text-to-image generation, the generated results appear visually realistic but lack a consistent design style and fail to satisfy engineering design constraints.

Reducing each module leads to variations in site configurations or drawings that cannot be evaluated using road-related metrics, as shown in Figure 8. Consequently, only image similarity metrics were employed for quantitative comparison. The results of the compared evaluation metrics are listed in Table 1. The symbol ↓ indicates that lower values are better while ↑ the opposite is. The results show that the MiCRoadDiff model achieves the greatest performance in all three metrics. As shown in Figure 9, all three metrics exhibit clear improving trends as the model evolves from the SDXL baseline to MiCRoadDiff. Compared with SDXL, MiCRoadDiff achieves improvements of 83%, 214%, and 214% in FID, SSIM, and PSNR, respectively. The introduction of the ControlNet and LoRA modules effectively enhances overall model performance. Removing the LoRA module results in a significant decline in performance, with FID, SSIM, and PSNR dropping by 272%, 50%, and 18%, respectively. Moreover, MiCRoadDiff significantly outperforms models with a single enhancement strategy, thereby verifying the effectiveness of the proposed framework for modular construction haul road design.

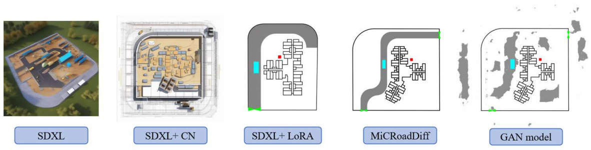

Figure 8. Comparison with baseline models

Table 1 Test results of ablation experiment

Model	FID↓	SSIM↑	PSNR↑
SDXL	530.41	0.27	3.95
SDXL+CN	326.85	0.42	10.19
SDXL+LoRA	243.40	0.83	11.51
MiCRoadDiff	87.66	0.85	12.44

(SDXL+LoRA+CN)

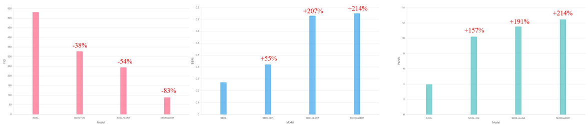

Figure 9. Comparison between different models regarding FID, SSIM, and PSNR

5.2 Layout design with different conditions

To investigate the adaptability of MiCRoadDiff under different design conditions, a series of experiments were conducted by modifying key site constraints, including truck parking area locations, site gate positions, and conditional maps. The qualitative results are illustrated in Figure 10.

When the truck parking area is relocated, MiCRoadDiff successfully generates distinct haul road layouts that build connectivity between the site gates and the designated parking area.

When the site gate position is changed, the generated haul roads could provide feasible access to the truck parking area, while respecting site boundaries and site gates. These results indicate that MiCRoadDiff is capable of handling changes in global access conditions.

When only Map-A (Building and site boundaries) is provided, the generated haul road generally follows the site boundary but lacks precise alignment with functional facilities, resulting in ambiguous or suboptimal connectivity. Conversely, when only Map-B (Site gates, tower crane, and truck) is used, the model correctly recognizes facility locations but may produce road geometries that are less compatible with the overall site shape. These observations highlight that a single conditional map is insufficient to fully constrain the haul road design task. The combination of multiple condition maps is necessary to simultaneously enforce geometric boundaries and functional requirements.

Overall, the experimental results demonstrate that MiCRoadDiff can robustly adapt to varying site conditions. The model effectively integrates multi-source conditional information to generate coherent and feasible haul road layouts.

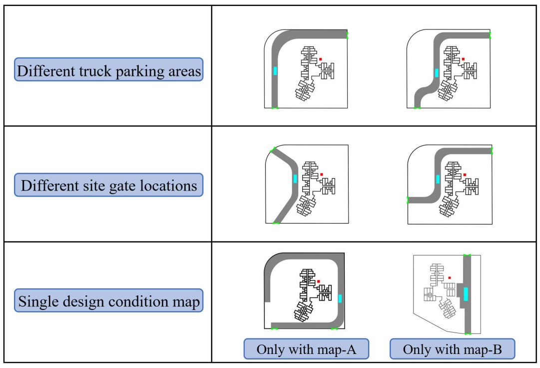

Figure 10. Comparison with different design conditions

6 Conclusion and Future Work

This paper proposed MiCRoadDiff, a novel generative design framework for automated on-site haul road layout generation in modular construction. Besides, the evaluation strategy takes into account both visual and layout-based metrics, ensuring a reliable assessment of design quality. By integrating a pre-trained Stable Diffusion backbone with LoRA-based domain adaptation and ControlNet-based geometric constraint control, the proposed method enables the efficient generation of feasible and high-quality haul road layouts under complex site conditions. The generated layouts of MiCRoadDiff form relatively complete site plans with high image quality and clarity, satisfying the prescribed design conditions and are engineering feasible. Experiments also highlight the significant contribution of LoRA to generation quality. Removing the LoRA module results in a significant decline in performance, with FID, SSIM, and PSNR dropping by 272%, 50%, and 18%, respectively. Meanwhile, ControlNet ensures the controllability of the design process, maintaining its reliability and precision. The experimental results demonstrate that MiCRoadDiff can robustly adapt to varying site conditions such as different site gate locations and truck parking areas. The proposed framework significantly reduces the manual effort required for haul road design while maintaining strong engineering feasibility. Future work will focus on extending the proposed framework to more complex construction scenarios, including larger-scale sites and irregular geometries, as well as incorporating road structural properties such as pavement strength classes. In addition, the influence of parametric variations in textual design conditions on generation outcomes will be systematically investigated. Integrating the generated haul road layouts with Building Information Modeling (BIM) platforms or professional traffic and road simulation tools provides a promising direction, which could further support engineers in preliminary design evaluation and feasibility analysis.

References

[1] W. Pan, C.K. Hon, Briefing: Modular integrated construction for high-rise buildings, Proceedings of the Institution of Civil Engineers - Municipal Engineer 173 (2020) 64–68. https://doi.org/10.1680/jmuen.18.00028.

[2] CIC, Why MiC, (2020). https://mic.cic.hk/en/WhyMiC (accessed January 26, 2026).

[3] C. Yi, M. Lu, A mixed-integer linear programming approach for temporary haul road design in rough-grading projects, Automation in Construction 71 (2016) 314–324. https://doi.org/10.1016/j.autcon.2016.08.022.

[4] D. Xiang, H. Lin, J. Ouyang, D. Huang, Combined improved A* and greedy algorithm for path planning of multi-objective mobile robot, Scientific Reports 12 (2022) 13273.

[5] B. Huang, L. Yao, K. Raguraman, Bi-level GA and GIS for Multi-objective TSP Route Planning, Transportation Planning and Technology 29 (2006) 105–124. https://doi.org/10.1080/03081060600753404.

[6] D. Lenz, T. Kessler, A. Knoll, Tactical cooperative planning for autonomous highway driving using Monte-Carlo Tree Search, in: 2016 IEEE Intelligent Vehicles Symposium (IV), IEEE, Gotenburg, Sweden, 2016: pp. 447–453. https://doi.org/10.1109/IVS.2016.7535424.

[7] W. Fang, L. Ding, B. Zhong, P.E. Love, H. Luo, Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach, Advanced Engineering Informatics 37 (2018) 139–149.

[8] H. Luo, C. Xiong, W. Fang, P.E. Love, B. Zhang, X. Ouyang, Convolutional neural networks: Computer vision-based workforce activity assessment in construction, Automation in Construction 94 (2018) 282–289.

[9] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, Commun. ACM 63 (2020) 139–144. https://doi.org/10.1145/3422622.

[10] R. Li, H.-L. Chi, Z. Peng, X. Li, A.P.C. Chan, Automatic tower crane layout planning system for high-rise building construction using generative adversarial network, Advanced Engineering Informatics 58 (2023) 102202. https://doi.org/10.1016/j.aei.2023.102202.

[11] G. Zhang, Y. Pan, L. Zhang, Semi-supervised learning with GAN for automatic defect detection from images, Automation in Construction 128 (2021) 103764.

[12] D. Saxena, J. Cao, Generative Adversarial Networks (GANs): Challenges, Solutions, and Future Directions, ACM Comput. Surv. 54 (2022) 1–42. https://doi.org/10.1145/3446374.

[13] J. Ho, A. Jain, P. Abbeel, Denoising Diffusion Probabilistic Models, in: Advances in Neural Information Processing Systems, Curran Associates, Inc., 2020: pp. 6840–6851.

[14] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer, High-Resolution Image Synthesis with Latent Diffusion Models, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, New Orleans, LA, USA, 2022: pp. 10674–10685. https://doi.org/10.1109/CVPR52688.2022.01042.

[15] L. Zhang, A. Rao, M. Agrawala, Adding Conditional Control to Text-to-Image Diffusion Models, in: 2023 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, Paris, France, 2023: pp. 3813–3824. https://doi.org/10.1109/ICCV51070.2023.00355.

[16] N. Nauata, K.-H. Chang, C.-Y. Cheng, G. Mori, Y. Furukawa, House-GAN: Relational Generative Adversarial Networks for Graph-constrained House Layout Generation, (2020). https://doi.org/10.48550/arXiv.2003.06988.

[17] P. Zeng, W. Gao, J. Yin, P. Xu, S. Lu, Residential floor plans: Multi-conditional automatic generation using diffusion models, Automation in Construction 162 (2024) 105374. https://doi.org/10.1016/j.autcon.2024.105374.

[18] Y. Gu, Y. Huang, W. Liao, X. Lu, Intelligent design of shear wall layout based on diffusion models, Computer-Aided Civil and Infrastructure Engineering (2024) mice.13236. https://doi.org/10.1111/mice.13236.

[19] Stability AI, generative-models, GitHub Repository (2023). https://github.com/Stability-AI/generative-models.

[20] D. Podell, Z. English, K. Lacey, A. Blattmann, T. Dockhorn, J. Müller, J. Penna, R. Rombach, SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis, (2023). https://doi.org/10.48550/arXiv.2307.01952.

[21] E.J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, W. Chen, LoRA: Low-Rank Adaptation of Large Language Models, (2021). https://doi.org/10.48550/arXiv.2106.09685.

[22] D.P. Kingma, J. Ba, Adam: A Method for Stochastic Optimization, (2017). https://doi.org/10.48550/arXiv.1412.6980.