Optimizing Generative Diffusion Models with Reinforcement Learning from Human Feedback (RLHF) in architectural: A Case Study of Campus Layouts

Ying Lin[1] , Fei Ye*[2]

Xi'an University of Architecture and Technology

Abstract: Generative diffusion models are extensively used in architectural design, such as campus layout planning, but often fail to align with precise aesthetic and functional expectations. This study enhances these models by implementing Reinforcement Learning from Human Feedback (RLHF), integrating expert architectural feedback directly into the training process. A systematic annotation pipeline was developed to gather qualitative evaluations from experienced architects, which served as direct feedback for model training. Using RLHF, the models are iteratively fine-tuned to more accurately incorporate key architectural elements and spatial configurations that meet professional preferences. The models refined through this method demonstrated substantial improvements, producing architectural images that closely match desired styles and are superior to traditional models in terms of both technical metrics and expert assessments of practical and aesthetic attributes.

Keywords: Deep learning；Generative design；optimization methods；RLHF；Layout plan

1 Introduction

With the rapid development of technology and the rise of machine-assisted design, artificial intelligence, automation and machine learning have brought inevitable opportunities and challenges to various disciplines, especially the field of architecture. In traditional architectural design methods based on manual drawing and physical models, designers express their design concepts through sketches and technical drawings, paying attention to details and proportions. The process includes preliminary research, schematic design, design development and construction drawing design, with an emphasis on teamwork and communication. Generative design, on the other hand, uses computer algorithms and artificial intelligence to automatically generate multiple design schemes based on set parameters and rules. Designers play a significant role in setting parameters and selecting schemes, allowing them to quickly explore complex design forms and optimisation schemes, improving design efficiency and creativity.

1.1 Background

Generative diffusion models have become powerful tools in a variety of design disciplines, particularly in architectural design tasks such as campus layout planning. These models use deep learning techniques to generate high quality architectural visualisations, assisting architects in the early stages of design by providing a wide range of design options. However, while generative models offer significant advantages in terms of speed and creativity, they often fail to meet the exacting aesthetic and functional requirements expected in professional architectural practice. Despite the ability to generate realistic images, the lack of a nuanced un-

[1] Y. Lin, Xi'an University of Architecture and Technology. 71000, China, e-mail: lynn813910@gmail

[2] F. Ye (✉), Xi'an University of Architecture and Technology. 71000, China, e-mail: feiye@xauat.edu.cn

derstanding of architectural principles means that the output may not meet the architect's specification criteria.

1.2 Problem statement

Despite advances in generative modelling, current diffusion models struggle to meet the precise expectations of architects, particularly when it comes to incorporating specific aesthetic and functional elements required in professional practice. Generated designs often lack coherence in spatial configurations and do not adhere to established architectural standards, reducing their utility in real-world applications. Furthermore, the lack of human-in-the-loop feedback during the model training process exacerbates this problem, as the models are unable to learn from expert judgements and adjust their outputs accordingly. This disconnect between the generated outputs and the practical needs of architects is a significant barrier to the adoption of these models in professional settings. Therefore, there is an urgent need to develop a method that allows for the integration of expert architectural feedback into the training process, thereby enhancing the ability of the model to produce designs that are both technically sound and aesthetically pleasing.

1.3 Objectives

The primary objective of our study is to enhance the capabilities of diffusion models in architectural design by integrating Reinforcement Learning from Human Feedback (RLHF) into the training process.

Firstly, to align the generated output with professional architectural standards: By incorporating expert feedback, the research aims to improve the model's ability to generate designs that meet aesthetic and functional expectations, ensuring that the output is more closely aligned with the needs of practising architects.

Secondly, this study will establish a structured process for collecting qualitative assessments from experienced architects. These assessments will be used systematically to guide model training, ensuring that feedback is effectively integrated into the generation process.

At the end of the article, the performance of the RLHF enhanced model is compared with that of the conventional generative model using both quantitative metrics and qualitative expert judgement. The aim is to demonstrate that the enhanced model has improved in terms of generating architectural designs that meet professional standards.

2 Methodology

This study proposes a generative framework for architectural design that integrates Reinforcement Learning from Human Feedback (RLHF) to enhance the practical application of diffusion models. The process consists of four phases: In the training phase, architectural design data is pre-processed and the pre-trained model is fine-tuned; in the generation phase, low-rank adaptation is used to optimise model outputs; in the evaluation phase, model performance is refined through genetic algorithms and expert feedback; finally, in the optimisation phase, expert review guides iterative adjustments to produce high-quality design solutions (Fig.1). This approach significantly improves the accuracy of the model and its practical utility in architectural design.

2.1 Data processing

Fig. 2 shows the complete process of architectural data processing and functional area division. Figure a) details the steps from data collection to processing: first, relevant data is collected from Mapbox, OpenCase, and the design project through a Python script, and then the data is annotated using the BLIP tool to generate training images and a corresponding label file. Next, these images and labels were resized to fit the training requirements of various models, including 512x512 pixel images for generating adversarial networks (GANs), 1024x1024 pixel datasets for diffusion models (DMs), and 128x128 pixel images for variational autoencoders (VAEs). Figure b) clearly indicates the division of different functional areas in the building de-

sign, such as playgrounds, teaching buildings, gymnasiums, etc., through colour coding, so that the results of data processing can be directly related to the practical application of building design.

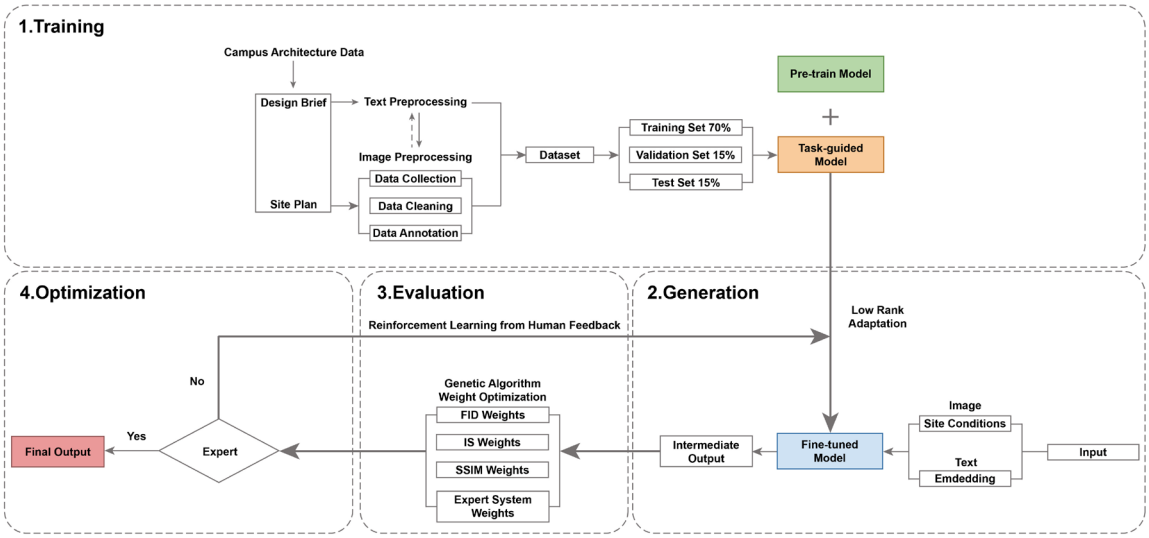


Fig. 1 The framework of the optimization of generative model

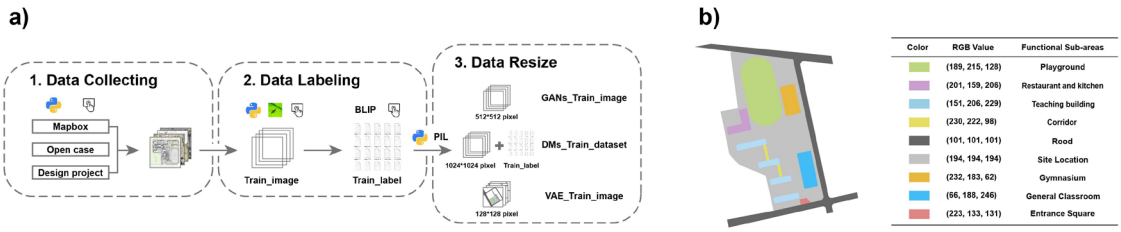


Fig. 2 Dataset construction process and labelling rule

2.2 Model training

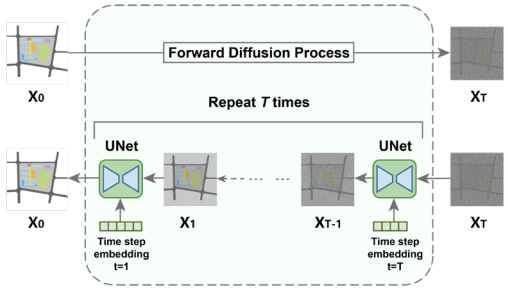


Fig. 3 Principle of Diffusion Model for Layouts

Figure 3 shows a method for generating building layouts based on a diffusion model. In this method, the initial architectural design image X_0 is gradually transformed into a completely noise-free image X_T through T diffusion iterations. At each iteration step, the U-Net model is used in combination with time-step embedding to process the image, gradually removing the noise and inverting it to restore a clearer image. Ultimately, through this diffusion and denoising process, the model can generate architectural layout drawings of higher quality and in line with design requirements. The effectiveness of the diffusion model in image generation is confirmed, especially its potential for application in architectural design tasks.

2.3 Reinforcement Learning from Human Feedback

OpenAI's ChatGPT dialogue model has ignited a new wave of AI innovation, demonstrating an ability to respond fluently to a wide range of questions and seemingly breaking down the barriers between machine and human interaction. At the heart of this breakthrough is a new training paradigm in the field of large language model (LLM) development: Reinforcement Learning from Human Feedback (RLHF). RLHF enhances reinforcement learning models by incorporating direct feedback from human experts. This method aims to guide the training process by not only relying on a predefined reward function but also dynamically adjusting the model's behavior to better meet the needs of specific tasks and align with human expectations. The RLHF training process is typically divided into three key steps: pre-training a language model, aggregating question-answer data and training a reward model, and fine-tuning the language model using reinforcement learning.

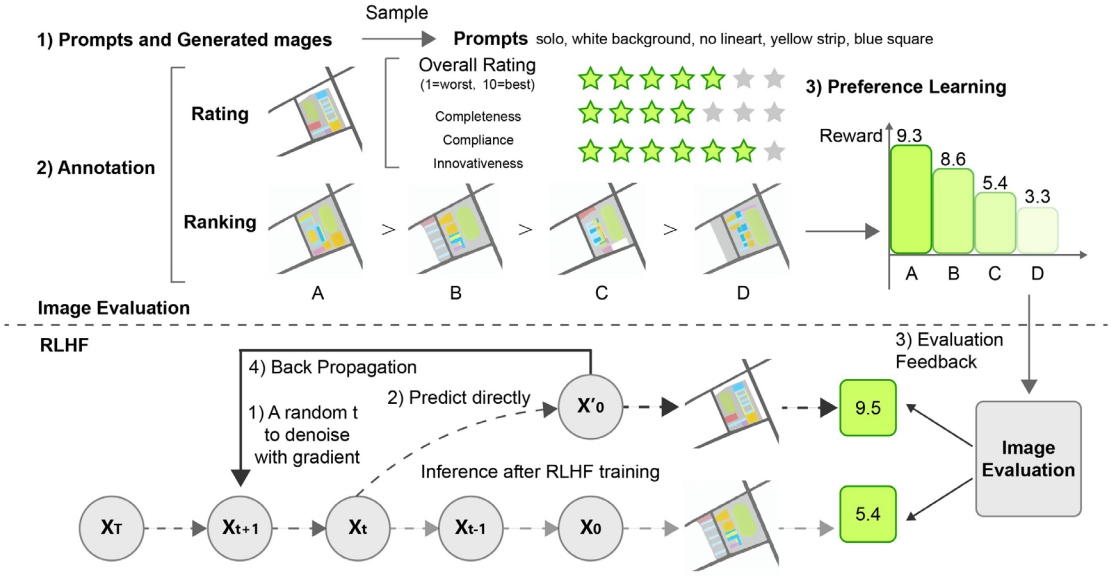


Fig. 4 An overview of the RLHF for Image Generation and Optimization

Figure 4 illustrates two key processes in image generation and optimization. The top part shows the image annotation and training process, which includes data collection, image annotation and preference learning. By collecting a large dataset of images and having them annotated by experts, put the rating and ranking results in the reward， then the model learns preferences through this feedback, improving its generative capabilities. The lower part shows the RLHF process, where RLHF uses the evaluation feedback from experts to directly optimize the diffusion models in a random denoising step. This approach ensures that the model not only produces high quality images, but also better matches human preferences and design requirements.

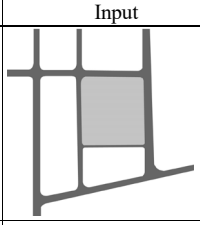
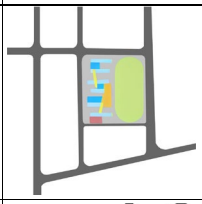
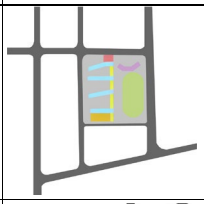
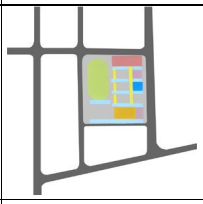
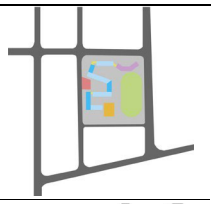
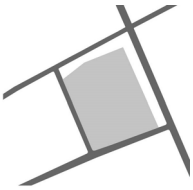
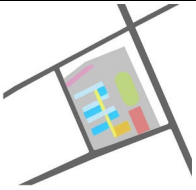
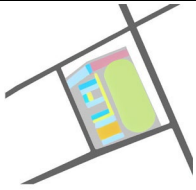
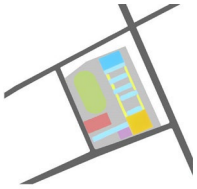
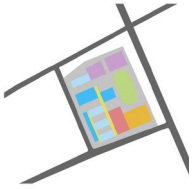
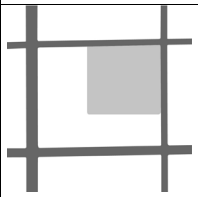
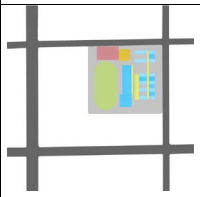
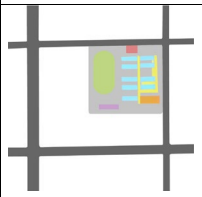
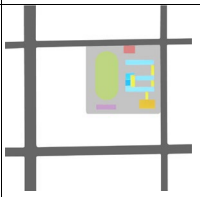
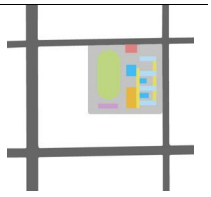
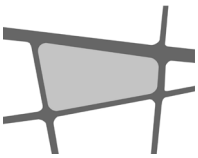
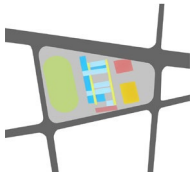
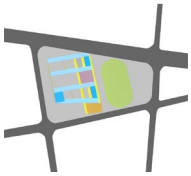
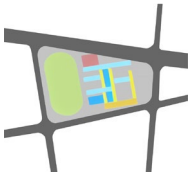
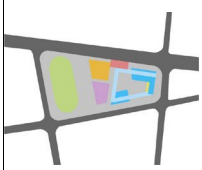
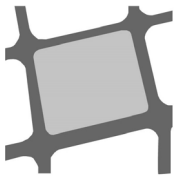
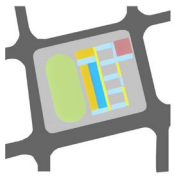
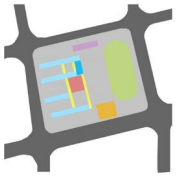
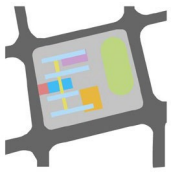
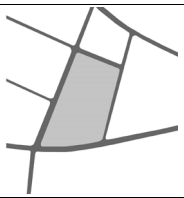
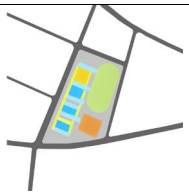
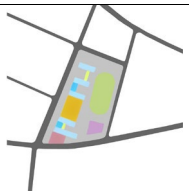
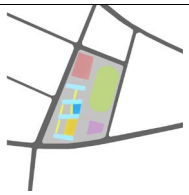
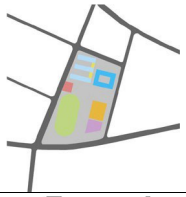
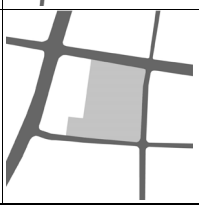
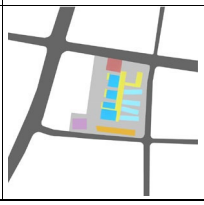
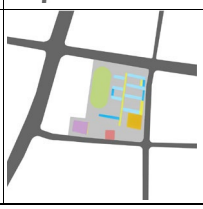
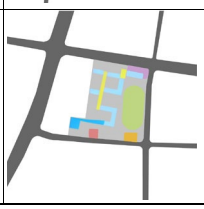
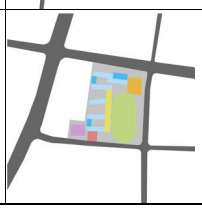
2.4 Generating Result

Table 1 compares the layouts generated by different models based on different prompts. The table shows the process of generating output from initial input designs to the standard diffusion model (SD), the extended diffusion model (SDXL), and the model trained through human feedback reinforcement learning (RLHF).

SD represents the layout generated by the standard diffusion model, reflecting how it incorporates the specified elements into the design. SDXL shows the output from the extended diffusion model, which generally offers more detailed and refined layouts. RLHF illustrates the layout produced by the model trained with Reinforcement Learning from Human Feedback, designed to more closely match the desired outcomes specified by the prompts. Ground Truth is column presents the actual design result from the real-world case, which the models aim to replicate or approximate.

The results are then evaluated qualitatively and quantitatively.

Table 1 Comparison of Model-Generated Layouts

Prompt	Input	SD	SDXL	RLHF	Ground Truth
No.054: yellow strip, blue square, green figure, 2d					
No.071: no humans, white background, simple background, yellow strip, blue square, green figure, 2d					
No.077: solo, yellow strip, blue square, green figure, 2d					
No.091: abstract, yellow strip, blue square, green figure, 2d					
No.104: abstract, no humans, simple background, white background, yellow strip, blue square, green figure, 2d					
No.132: no humans, white background, abstract, yellow strip, blue square, green figure, 2d					
No.155: solo, yellow strip, blue square, green figure, 2d					

3 Result evaluation

3.1 Qualitative Evaluation

From a qualitative perspective, the RLHF model demonstrates superior visual quality and design alignment

in the generated architectural layout images. By incorporating human feedback, the model better understands and meets practical design requirements, producing images that closely match expert expectations. This approach ensures that the model generates images that are not only technically proficient, but also meet high standards of aesthetics and functionality. In comparison, while the SD and extended diffusion models (SDXL) can generate images based on prompts, they fall short of the RLHF model in terms of accuracy and design consistency.

3.2 Quantitative Evaluation

3.2.1 SSIM

SSIM (Structural Similarity Index) measures the structural similarity between the generated image and the ground truth image. The RLHF model shows higher SSIM values for most images, indicating that its outputs are structurally closer to the real designs. PSNR (Peak Signal-to-Noise Ratio) calculation is based on the Mean Squared Error (MSE), which represents the average of the squared differences between the pixel values of the original and reconstructed images. PSNR is expressed in decibels (dB), with higher values indicating less difference between the two images, meaning the reconstructed image is of higher quality. LPIPS (Learned Perceptual Image Patch Similarity) measures the perceptual distance between images, with lower values indicating closer perceptual similarity.

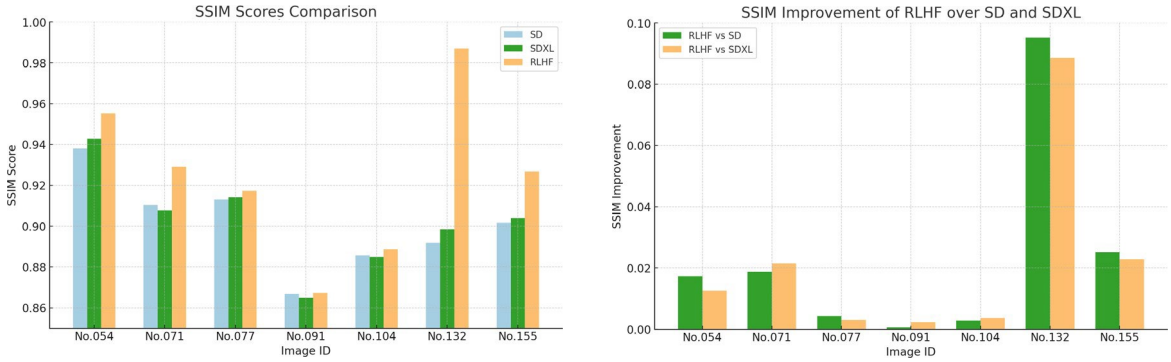


Fig. 5 Comparison of SSIM Performance Across RLHF, SD, and SDXL Models

Figure 5 (left) directly compares the SSIM scores of images generated by the SD, SDXL, and RLHF models. For each image ID, there are three bars representing the SSIM scores of the SD, SDXL, and RLHF models, respectively. The results shows the RLHF model consistently shows higher SSIM scores across most image IDs, especially in Image No.132 and No.155, where the SSIM scores approach 1.0. This suggests that the images generated by the RLHF model are structurally very similar to the ground truth.

Figure 5 (right) the extent of SSIM improvement that the RLHF model achieves over the SD and SDXL models. The green bars represent the improvement of RLHF over SD, while the orange bars show the improvement of RLHF over SDXL. In most cases, the RLHF model outperforms both the SD and SDXL models in terms of SSIM, particularly in Image No.132, where the improvement is most significant. This indicates that the RLHF model generates images with structural similarities closer to the ground truth.

3.2.2 PSNR

Figure 6 (left) compares the PSNR scores across different models (SD, SDXL, and RLHF) for each image ID. The lines indicate how each model performs in terms of PSNR for the given images. The RLHF model consistently shows higher PSNR values compared to SD and SDXL models, particularly for certain images like No.132 and No.155. This suggests that the RLHF model generally produces higher quality images in terms of signal-to-noise ratio.

Figure 6 (right) shows the improvement in PSNR (Peak Signal-to-Noise Ratio) achieved after applying the RLHF (Reinforcement Learning from Human Feedback) model compared to other models. The bars repre-

sent the PSNR difference in dB for each image ID, showing how much better the RLHF model performs in terms of PSNR improvement. The RLHF model generally improves PSNR, particularly for images like No.091 and No.104. However, there are instances, such as No.132, where the improvement is minimal or even slightly negative, indicating that the RLHF model might not always enhance image quality in every scenario.

These charts together indicate that the RLHF model tends to improve image quality as measured by PSNR, especially in certain images, while it may not always guarantee improvement in every case. The RLHF model generally outperforms SD and SDXL models in terms of PSNR, which suggests its effectiveness in enhancing image quality through reinforcement learning with human feedback.

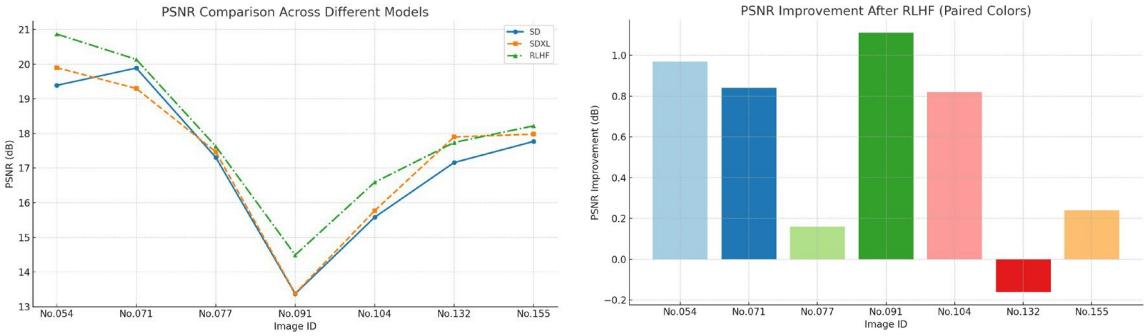


Fig. 6 Comparison of PSNR Performance Across RLHF, SD, and SDXL Models

3.2.3 LPIPS

Figure 7 compares the LPIPS (Learned Perceptual Image Patch Similarity) distances across three models: SD, SDXL, and RLHF, for different image IDs. Lower LPIPS values indicate better perceptual similarity to the ground truth image. The RLHF model generally shows lower LPIPS distances compared to SD and SDXL, suggesting that it produces images with perceptual qualities closer to the ground truth. For images like No.091, all models exhibit higher LPIPS distances, indicating a greater perceptual difference from the ground truth. However, the RLHF model still outperforms the other two models. In contrast, for images like No.132, the RLHF model achieves a significantly lower LPIPS distance, further confirming its superior performance in generating perceptually accurate images.

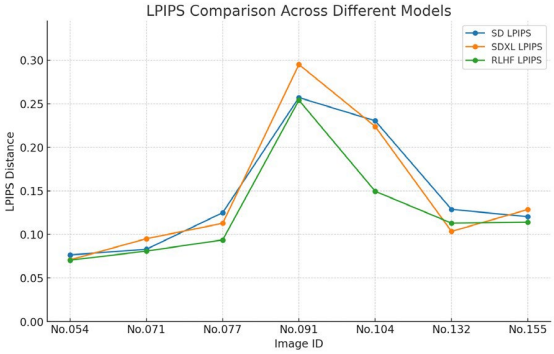


Fig. 7 LPIPS Comparison Across Different Models

The RLHF model shows a consistent trend of lower LPIPS values across most image IDs, highlighting its advantage in producing images that are visually closer to the reference.

3.2.3 Relationship Between Image Quality Metrics and Performance

Figure 8 (left) is a scatter plot further illustrates the positive correlation between SSIM and PSNR across the three methods (SD, SDXL, RLHF). The RLHF model generally shows higher SSIM values for given PSNR values compared to SD and SDXL, indicating that RLHF provides better structural similarity (SSIM) for im-

ages even when PSNR is not significantly higher. The trend lines for each method clearly demonstrate that RLHF tends to outperform both SD and SDXL in terms of SSIM for a given PSNR, suggesting that the fine-tuning process in RLHF effectively improves the balance between structural similarity and noise reduction in images.

Figure 8 (right) shows strong correlations between PSNR, SSIM, and LPIPS across different models (SD, SDXL, RLHF). Specifically, PSNR and SSIM have a strong positive correlation, indicating that as PSNR improves (higher values), SSIM also tends to improve, suggesting better image quality. Conversely, LPIPS shows a negative correlation with both PSNR and SSIM, which is expected since lower LPIPS values indicate better perceptual similarity.

The "Improvement (%)" metric, which measures the relative improvement of RLHF over SD and SDXL, shows a negative or low correlation with most other metrics, indicating that the improvement metric does not consistently correlate with raw scores of PSNR, SSIM, or LPIPS. This suggests that while RLHF may improve image quality in certain aspects, the relationship isn't straightforward across all metrics.

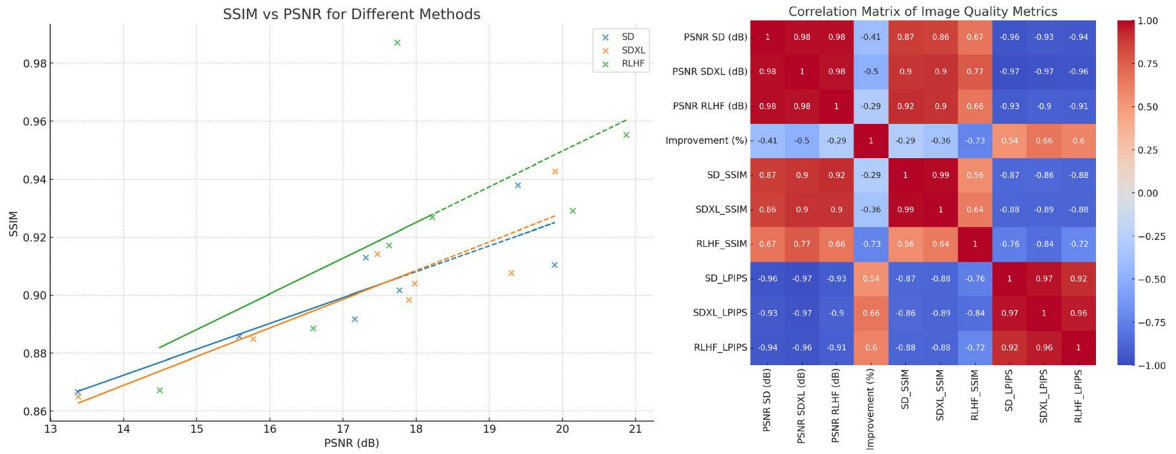


Fig. 8 LPIPS Comparison Across Different Models

4 Conclusion and discussion

This study demonstrated the potential of applying Reinforcement Learning from Human Feedback (RLHF) to improve generative diffusion models in the context of architectural design, specifically in campus layout planning. By integrating expert feedback directly into the training process, the models produced outputs that better aligned with professional standards and aesthetic preferences. The case study of campus layouts showed that RLHF-refined models achieved superior performance across key technical metrics such as PSNR, SSIM, and LPIPS, indicating a substantial improvement in generating designs that meet both technical and aesthetic criteria.

The results of this study suggest that RLHF can significantly enhance the practical application of AI in architecture, particularly for tasks that require a blend of creativity and functional accuracy. The improved consistency and quality of the generated designs underscore the value of incorporating expert feedback into AI-driven design processes. Moving forward, this approach could be extended to other areas of architectural design, offering a pathway for more sophisticated and contextually relevant AI-assisted design tools. Future research could further explore the implications of RLHF in automating complex design tasks while maintaining the creative and functional standards essential in architecture.

Acknowledgments

This research was funded by the Ministry of Science and Technology under the 14th Five-Year National Key R&D Program (Project No. 2023YFC3805505-04). We sincerely thank the editors and reviewers for valuable comments and suggestions.

References

1. Xu J, Liu X, Wu Y, et al. Imagereward: Learning and evaluating human preferences for text-to-image generation[J]. Advances in Neural Information Processing Systems, 2024, 36.
2. Ye J, Liu F, Li Q, et al. Dreamreward: Text-to-3d generation with human preference[J]. arXiv preprint arXiv:2403.14613, 2024.
3. Yuan Y, Hao J, Ma Y, et al. Uni-RLHF: Universal Platform and Benchmark Suite for Reinforcement Learning with Diverse Human Feedback[J]. arXiv preprint arXiv:2402.02423, 2024.