Evaluation of charge storage ability of chrome doped Mn2O3 nanostructures derived by cathodic electrodeposition

Original
Evaluation of charge storage ability of chrome doped Mn2O3 nanostructures derived by cathodic electrodeposition / Darjazi, H., Hosseiny Davarani, S.S., Moazami, H.R., Yousefi, T., Tabatabaei, F.. - In: PROGRESS IN NATURAL SCIENCE. - ISSN 1002-0071. - 26:6(2016), pp. 523-527. [10.1016/j.pnsc.2016.09.006]

Availability:
This version is available at: 11583/2998629 since: 2025-03-27T08:30:00Z

Publisher:
Elsevier

Published
DOI:10.1016/j.pnsc.2016.09.006

Optimizing Inference Conditioning Techniques in Image Generation for Participatory Urban Transformation

Enrico Pupi[a], Piergiuseppe Rechichi[b]

[a] Department of Architecture and Design,Politecnico di Torino, Italy,
enrico.pupi@polito.it;
[b] Department of Energy, Systems, Territory and Buildings Engineering,
Università di Pisa, Italy,
piergiuseppe.rechichi@phd.unipi.it

ABSTRACT: Generative Artificial Intelligence (GenAI) is emerging as a transformative medium for democratizing participatory urban design, potentially bridging the gap between citizens' conceptualizations and professional representations. While current GenAI tools are divided between professional-grade platforms and accessible solutions for non-experts, technical challenges persist in generating precise and contextually relevant visualizations. This research investigates the optimization of inference conditioning techniques through an open-source approach, implementing an integrated framework within ComfyUI that leverages local computing resources. The methodology combines three key enhancements: prompt engineering through Large Language Models (LLMs), fine-tuning through Low-Rank Adaptation (LoRA), and structural control through ControlNet implementations. Testing this framework on two case studies in Pisa's historical center and suburban area demonstrated how the synergistic combination of these techniques significantly improves the quality and contextual relevance of generated visualizations. Results suggest that advanced conditioning strategies can effectively balance accessibility and precision in participatory urban design tools, supporting the development of more inclusive and sustainable urban transformation processes aligned with the UN's 2030 Agenda goals.

1. INTRODUCTION TO GENERATIVE AI FOR PARTICIPATORY PROCESSES

In recent years, more and more cities have shifted toward the smart city paradigm [1], adopting digital technologies to analyze complex social, economic, and environmental phenomena at an urban scale and pursue the optimization of public investment, service management, and urban development.

Generative Artificial Intelligence (GenAI) plays a key role in this shift. Recent reports on this topic [2, 3, 4] highlight the flexibility with which GenAI can be applied in urban management and planning. Research explores its application in energy management, mobility management, healthcare, hydraulic management, climatic concerns, and biodiversity studies. In Architecture Engineering Construction (AEC), GenAI is increasingly applied at both architectural and urban scales and may offer the potential to develop cross-scale tools and databases.

It is important to understand how digitizing assets and processes in urban management and planning should be oriented toward achieving the Sustainable Development Goals of the United Nations [5]. A UN-HABITAT report highlights the risks and challenges in applying GenAI in urban studies, and applications must consider the importance of a human-centered approach when confronting contemporary and future city challenges [2]. This approach is based on Sustainable Development Goal (SDG) 11, promoting safe, inclusive, resilient, and sustainable cities.

GenAI can facilitate data acquisition, data management, data analysis, data visualization, and predictive modeling in urban environments, tackling different tasks with different purposes. One relevant aspect of fulfilling SDG 11 is the role of GenAI in participatory processes.

Publiziert in: Lengyel, Dominik (Hrsg.): EVA Berlin 2025. Electronic Media and Visual Arts: 28th Issue of the EVA Berlin Conference, (EVA Berlin, Band 28), Heidelberg: arthistoricum.net, 2025, S. 225-235, https://doi.org/10.11588/arthistoricum.1568.c24100.

Existing literature shows GenAI supporting participatory processes by enabling the acquisition and processing of extensive qualitative data, converting it into quantitative data for visualization and interpretation [6], and leading to the development of predictive models [7, 8]. One interesting research focus is the analysis of relationships between urban environments and individuals' emotional well-being. Emotional response is fundamental for efficient involvement in participatory processes.

Recent research [9] emphasizes using Large Language Models (LLMs) to interact with citizens, collect and analyze textual data, and support urban planning and design. In the developed framework, LLMs refine prompts for image editing using Stable Diffusion. Prompts act on image portions with features that negatively affect emotional well-being. The images are acquired from street view databases and are edited to simulate renewal interventions. Recent literature on environmental psychology uses street view imagery to evaluate well-being measurements, achieving quantitative comparisons between environment metrics and psychological data, strengthening the scientific reliability of results [10].

However, relying solely on street view imagery risks diverting focus from user engagement to predictive models, limiting community involvement.

Visualization tools are key communicative mediums between professionals and communities in a top-down framework in participatory processes. Instead, it is possible to highlight how image generation can be part of co-design or bottom-up frameworks [11], engaging communities in generating diverse visions of urban development.

2. RECENT ADVANCEMENTS IN AI-BASED TOOLS FOR URBAN REGENERATION

Recent advancements in GenAI image processing enable the development of high-fidelity visualization tools for non-experts, allowing effective representation of human imagination in urban regeneration. GenAI has the potential to bridge the gap between citizens' conceptualizations and professional representations [12]. These visualizations can enhance individual satisfaction and engagement in participatory processes.

The unprecedented progress in image processing, particularly with Diffusion Models (DMs), has shown remarkable capabilities in generating, modifying, and manipulating visual content with increasing photorealism and contextual consistency [13]. Stable Diffusion marked a milestone, offering an open-source alternative to democratizing access to powerful image generation capabilities.

In the current landscape, multiple tools leverage GenAI for urban transformation processes, which can be categorized into two macro-groups based on their focus and target users.

- Professional-grade tools: Designed for sector experts and offer advanced urban design and analysis functionalities. Examples include Autodesk Forma, ArcGIS CityEngine, CITYPLAIN, and Digital Blue Foam. These platforms provide detailed analytical capabilities and support complex design processes.
- Tools for non-expert users: These aim to enhance bottom-up processes in urban design by offering intuitive interfaces and accessible functionalities. Examples include the CityScope project and the UrbanistAI platform.

These applications demonstrate how GenAI transforms professional work and citizen engagement in urban design processes. Professional tools focus on detailed analysis and advanced functionalities, whereas tools for non-expert users emphasize accessible representation and public participation.

While these tools focus on facilitating public participation, several challenges remain. The lack of prompt engineering expertise may limit the precision of generated visualizations. Using unoptimized DMs for specific urban contexts reduces image fidelity and relevance. Moreover, the iterative process of selective image modification, primarily limited to inpainting, restricts flexibility in adapting visualizations.

The convergence of advanced GenAI capabilities creates new opportunities for developing accessible yet sophisticated urban visualization tools [14]. These technologies can transform participatory urban design, enabling non-expert stakeholders to generate high-fidelity representations of their ideas. While professional tools advance precision and analytical depth, platforms for non-expert users democratize the process, fostering inclusivity and engagement. However, challenges remain, particularly in ensuring high-quality visualizations and meaningful community involvement.

3. GENERAL AIM AND RESEARCH-SPECIFIC OBJECTIVES

This research aims to investigate and optimize the application of advanced inference conditioning techniques within existing GenAI frameworks for participatory urban design (Fig. 1). This field of inquiry exists at the intersection of technological innovation in architectural representation and the development of participatory urban transformation processes [15]. Adopting an open-source approach represents a strategic choice aimed not only at ensuring the accessibility of developed tools but also at promoting the collaborative and transparent evolution of the implemented technologies.

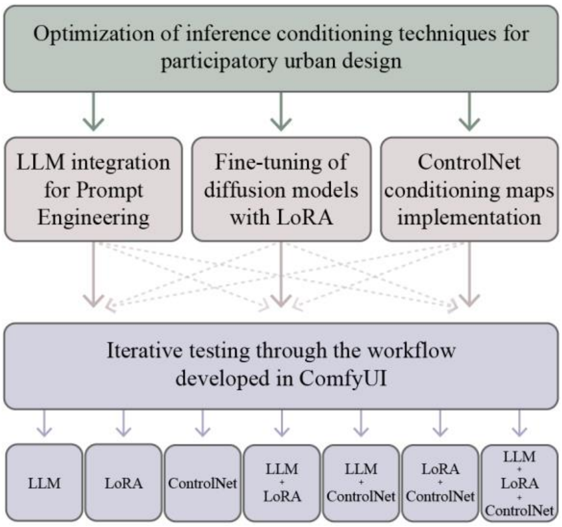


Figure 1: *Hierarchical methodological diagram illustrating the research structure: the first tier presents the general scope, the middle tier delineates specific research objectives focused on three inference conditioning techniques, and the bottom tier demonstrates how these objectives are pursued through iterative testing. This structured approach comprehensively evaluates each technique's contribution and synergistic effects.*

Integrating advanced GenAI capabilities with participatory urban design aligns with the emerging smart cities paradigm and the sustainable urban transformation objectives outlined in Agenda 2030 [4]. The accessibility to advanced visualization tools and the maintenance of high accuracy and contextual relevance standards catalyze more inclusive and effective urban planning processes.

The specific research objectives are articulated along four main directions. Analyzing and implementing the integration of LLMs in the prompt engineering process to overcome limitations due to non-expert users' lack of specific expertise. This approach aims to significantly improve the accuracy of generated visualizations through a natural and intuitive linguistic interface.

Evaluating the effectiveness of Low-Rank Adaptation (LoRA) in fine-tuning open-source DMs. The goal is to develop an efficient method for adapting generative models to specific urban contexts, producing visualizations that faithfully reflect local architectural and cultural characteristics.

Experimenting with ControlNet-based maps, particularly exploring the use of Depth Maps, Canny Edge Maps, and HED Maps (Soft Edge maps). This objective focuses on ensuring precise and flexible control over visual elements in proposed urban transformations.

Assessing the effectiveness of interpolating implemented techniques in generating visualizations that accurately reflect user intentions and coherently integrate with the existing urban fabric, maintaining a balance between design innovation and contextual accuracy.

4. APPLICATION METHODOLOGY OF PROCESS TESTING

In the context of the proposed methodology for optimizing inference conditioning techniques in image generation for participatory urban transformation, ComfyUI has been selected as the tool for experimenting with the synergistic integration of various methodological components (Fig. 2).

ComfyUI is an open-source graphical interface that utilizes a node-based system for executing custom workflows. Its modular architecture proves particularly suitable for implementing the proposed methodological framework for several reasons. The node-based interface enables the construction of complex pipelines that integrate LLMs for prompt engineering, implement workflows for fine-tuning through LoRA, and incorporate ControlNet conditioning maps.

Local computing, supported in this specific case by an NVIDIA® GeForce® RTX 4090 with 24GB GDDR6, allows for documenting and replicating methodological processes accurately and precisely. Computational autonomy ensures the ability to manage model fine-tuning with specific datasets and complete control over the generation process.

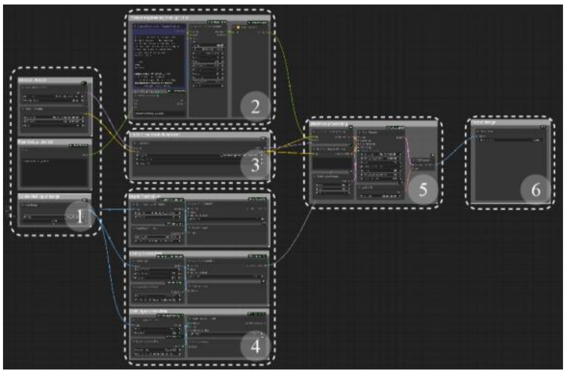
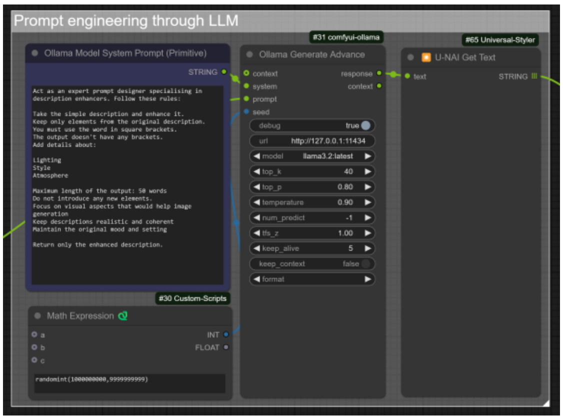


Figure 2: *Comprehensive visualization of the ComfyUI node-based workflow architecture. The diagram showcases how LLM prompt engineering, LoRA fine-tuning, and ControlNet implementations are orchestrated within a single computational pipeline running on local hardware (1 – Input as DM, raw prompt, and source image; 2 – LLM cluster; 3 – LoRA custom node; 4 – ControlNet cluster; 5 – Inference process; 6 – Output).*

ComfyUI's open-source approach facilitates research replicability and the framework's adaptability to different urban contexts.

4.1 PROMPT ENGINEERING THROUGH LARGE LANGUAGE MODELS

Prompt engineering is a relevant strategy in image generation through DMs, given the direct relationship between the quality of textual instructions and the accuracy of visual output [16].



```
Act as an expert prompt designer specialising in
description enhancers. Follow these rules:

Take the simple description and enhance it.
Keep only elements from the original description.
You must use the word in square brackets.
The output doesn't have any brackets.
Add details about:

Lighting
Style
Atmosphere

Maximum length of the output: 50 words
Do not introduce any new elements.
Focus on visual aspects that would help image
generation
Keep descriptions realistic and coherent
Maintain the original mood and setting

Return only the enhanced description.
```

```
debug true
url http://127.0.0.1:11434
model llama3.2:latest
top_k 40
top_p 0.80
temperature 0.90
num_predict -1
tfs_z 1.00
keep_alive 5
keep_context false
format
```

```
randomint(1000000000,9999999999)
```

Figure 3: *LLM-based prompt engineering system, showing the integration of Meta's Llama 3.2 3B model through the Ollama custom node.*

In this context, integrating LLMs in the prompt engineering process emerges as a methodological solution to bridge the gap between users' natural expressions and the technical specifications required by DMs, enabling automatic refinement and optimization of textual prompts.

Local computing capacity constraints have significantly influenced the implementation phase, particularly VRAM availability. This consideration guided the selection toward Meta's Llama 3.2 3B, which balances efficiency and performance. This choice consciously differs from heavier alternatives such as Llama 3.3 70B and Llama 3.1 405B, which are incompatible with the adopted local processing approach.

The implementation architecture is articulated through the local installation of the Ollama framework. Integration into ComfyUI occurs through a specific custom node for Ollama, where the prompt engineering system is structured according to a precise architecture (Fig. 3):

- Role definition as an expert prompt designer for descriptive enhancement.
- Operational constraints, including the preservation of original elements, respect for terms in square brackets (essential for activating the LoRA functionality in use), and the 50-word limit.
- Focus on visual aspects such as lighting, style, and atmosphere.
- Qualitative criteria such as realism, coherence, and maintenance of the original mood.

4.2 LORA FINE-TUNING OF THE OPEN-SOURCE DIFFUSION MODEL

DM fine-tuning represents an important methodological element for achieving contextually relevant results [17]. The current ecosystem of fine-tuning techniques offers various paths, including DreamBooth for detailed learning of specific subjects, Textual Inversion for acquiring new concepts through textual embeddings, and Hypernetworks for selective modifications to the model's behavior. However, LoRA stands out for its optimal balance between efficiency and performance.

Both technical and practical considerations guided the choice of utilizing LoRA. From a computational efficiency perspective, this technique requires significantly fewer hardware resources than alternative methods, making it possible to fine-tune models on consumer-grade GPUs while maintaining high-performance inference. The architectural implementation of LoRA presents additional advantages: it integrates with ComfyUI as an independent conditional node (Fig. 4), preserving the integrity of the base model and enabling the composition of multiple adaptations while facilitating selective application of fine-tuning during inference.

lora_name 1_ruined-building-san-lorenzo.safetensors
strength_model 1.00
strength_clip 1.00

Figure 4: *LoRA integration within the ComfyUI environment as an independent conditional node, enabling selective application of contextual fine-tuning during the inference process.*

The preparation of necessary datasets for fine-tuning is characterized by accentuated operational efficiency. Acquiring images focuses on selecting representative elements of the target urban context without requiring extensive pre-processing, focusing on contextual relevance. Dataset annotation was automated using Microsoft's Florence-2 vision foundation model, which generates accurate descriptions automatically through its multi-modal capabilities.

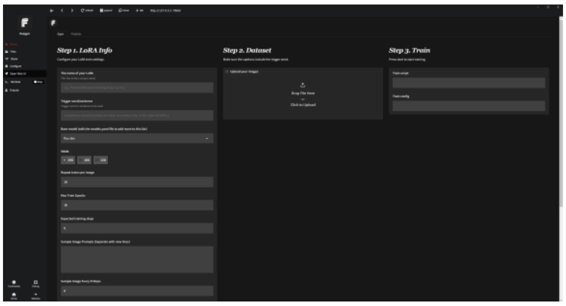


Figure 5: *Fluxgym web UI showing the elements for managing LoRA fine-tuning processes. The interface displays parameter controls, dataset management tools, and training progress monitoring features.*

The practical implementation of fine-tuning leverages the web UI Fluxgym, with the system architecture combining a frontend derived from AI-Toolkit (Fig. 5) with a backend powered by Kohya Scripts. This approach offers several implementational advantages: an intuitive interface for controlling the process, efficient resource optimization, and seamless integration with ComfyUI's comprehensive workflow.

All training utilized the Flux.1-dev as the base model, utilizing the following parameters:

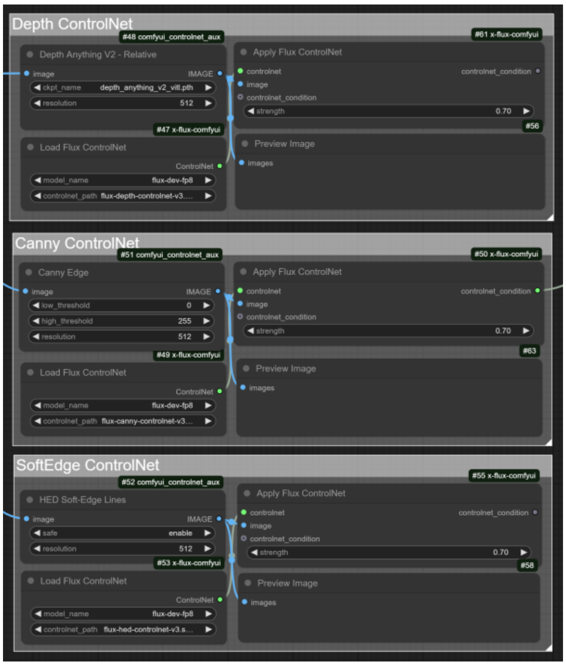
- 20 GB of VRAM
- 10 repeat trains per image
- 16 max train epochs
- Resized dataset images at 512px

Combining LoRA, optimized datasets, and efficient tools represents an optimal balance between accessibility, computational efficiency, and result quality.

4.3 IMPLEMENTATION OF CONTROLNET CONDITIONING MAPS

The ControlNet architecture introduced a significant innovation in image generation, providing a reference framework for inference conditioning through various control channels [13]. In the context of this research, this technique plays an essential role in preserving specific architectural and spatial relationships.

In this case, implementation was based on the Flux-controlnet-collections nodes developed by XLabs-AI, focusing on three primary types of control maps, each with distinct characteristics and functionalities.



ckpt_name depth_anything_v2_vitl.pth
resolution 512
model_name flux-dev-fp8
controlnet_path flux-depth-controlnet-v3....
strength 0.70

low_threshold 0
high_threshold 255
resolution 512
model_name flux-dev-fp8
controlnet_path flux-canny-controlnet-v3...
strength 0.70

safe enable
resolution 512
model_name flux-dev-fp8
controlnet_path flux-hed-controlnet-v3.s...
strength 0.70

Figure 6: *ControlNet conditioning implementation, showing the integration of multiple control maps (Depth, Canny Edge, and HED). Different control maps are processed and could be combined to maintain structural coherence in generated images.*

Depth Map ensures the preservation of spatial and volumetric relationships, guaranteeing coherence and accurate proportion control.

Canny Edge Detection emphasizes line contrast, preserving architectural details and clearly defining spatial boundaries.

Holistically-nested Edge Detection (HED), or Soft Edge Detection, operates through hierarchical border detection, capturing structural details at various scales and improving overall visual coherence.

The implementation process follows a structured methodology that unfolds into three main phases. First, the system needs to acquire the base image through a photograph of the urban context's current state, then generate specific control maps and calibrate parameter settings for detection. Then, map processing involves generating the three map types, with manual optimization of specific parameters leading to the validation of the quality of overall structural control. Lastly, the system can be integrated into the comprehensive pipeline, where maps are combined in ComfyUI workflow, balancing their influence and proceeding with iterative refinement (Fig. 6).

This component ensures that generated visualizations maintain a consistent structural coherence with existing urban contexts. Properly utilizing diverse control maps enables achieving an optimal balance between generative creativity and architectural constraints, producing spatial contextually relevant results.

5. SELECTED CASE STUDIES

The area identified for this research phase is the city of Pisa (Tuscany, Italy). Considering the current research and findings in the field, within our previous work, the results are strongly influenced by the coherence between the image generation process and the characteristics of the built and natural environments in the specific context.

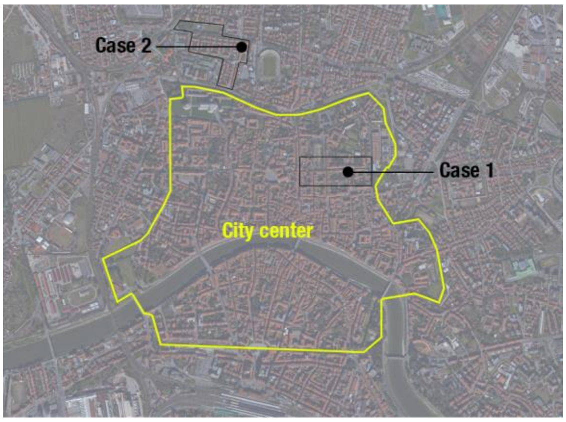


Figure 7: *The general plan of Pisa highlights the two selected case study areas: the historical city center featuring Palazzo Nissim (Case 1) and the suburban residential district of Via Piave/Via Rindi (Case 2). The map emphasizes the contrasting urban morphologies of selected areas.*

Deep knowledge of urban patterns and architectural features is fundamental for identifying areas of focus. Preliminary data collection is essential for the inference conditioning testing phase, especially when working with LoRA training.

Our analysis revealed two main areas of interest: the historical city center and a suburban residential area (Fig. 7). The distinct urban characteristics between these two areas will significantly impact the experiment's outcome.

The city center is characterized by narrow streets with tall buildings and reduced open green spaces. Only a few isolated criticalities can be localized in this area and identified as redevelopment opportunities. These emergencies are typically abandoned buildings, expressing the representation potential at the architectural scale.

5.1 FIRST CASE STUDY: HISTORICAL CITY CENTRE

The first case study is an abandoned building in Via San Lorenzo, also known as Palazzo Nissim (Fig. 8). This building has stood vacant for over two decades due to a long-standing dispute between public and private interests. However, despite the uncertainty, a resolution is within reach, potentially paving the way for revitalization.

Figure 8: *The current state documentation of Palazzo Nissim in Via San Lorenzo shows the abandoned historic building's facade. The image is the base reference for testing the inference conditioning framework in a heritage conservation context.*

In this case, the research tool under study could serve as a medium enabling users to imagine alternative scenarios and envision how the building might look if development were possible. Due to heritage conservation constraints, only limited renovations or restorations are permissible in this case. The experimental tool could foster inclusive and sustainable decision-making by incorporating heritage considerations.

The input data used in the generation process processed in ComfyUI are described below:

- Raw prompt: New life to the [ruined building San Lorenzo] with a conservative restoration project design through new finishes and overall urban redevelopment. The building is again in use with social and accommodation functions
- DM: flux1-dev-fp8.safetensors
- LoRA dataset: 39 images
- Inference step: 20
- Ratio: 1.1 it/s
- Inference time: 15-20 s per image

5.2 SECOND CASE STUDY: SUBURBAN RESIDENTIAL AREA

The suburban neighborhoods of Via Piave and Via Rindi are characterized by wide streets with consistent vehicular traffic. The residential area was initially designed as a public housing project, but its original function has diminished.

The municipality has launched a comprehensive redevelopment plan, which includes recent improvements to green areas and streets. In this case, the area's public spaces, parks, and streets require renovation to improve their livability and introduce urban furniture and services to enhance the quality of life for residents through high-quality urban design (Fig. 9).

Figure 9: *Via Piave/Via Rindi suburban area, showcasing the existing public spaces and residential buildings. The image highlights the current state of urban furniture and green areas targeted for redevelopment.*

The examined research tool is designed as a platform for citizens to evaluate redevelopment proposals, exploring reuse and reconfiguration scenarios that reflect their personal preferences and lived experiences.

The input data used in the generation process processed in ComfyUI are described below:

- Raw prompt: Urban design introducing new forms of ephemeral architecture in the [residential area park]. Street furniture and a public green environment offer innovative spaces for the neighborhood and the citizens
- DM: flux1-dev-fp8.safetensors
- LoRA dataset: 18 images
- Inference step: 20
- Ratio: 1.1 it/s
- Inference time: 15-20 s per image

6. ANALYSIS, INTERPRETATION AND RESULTS COMPARISON

The systematic analysis of the results obtained by implementing inference conditioning techniques revealed significant patterns in the effectiveness of the different applied methodologies. The most relevant aspect emerged from merging the three analyzed conditioning techniques: LLM for prompt engineering, LoRA for contextual adaptation, and ControlNet for structural control. This integration produced superior qualitative results regarding architectural coherence and contextual fidelity, significantly surpassing the performance of individual methods used separately. Using Canny Edge and HED maps through ControlNet proved particularly effective in preserving the morphological characteristics of historic buildings and urban spaces in structural control. Depth maps revealed significant limitations in handling complex architectural images, suggesting the need for further optimizations for this specific mode of control.

The comparative analysis of individual implementations revealed distinct specificity and limitations: while the exclusive use of LLM enhanced descriptive accuracy without guaranteeing sufficient architectural coherence, LoRA implementation alone increased stylistic fidelity but occasionally exhibited compositional inconsistencies. ControlNet, used in isolation, preserved spatial relationships but produced sometimes unconvincing results. We observed a significant amplification of generative capabilities by combining these three techniques. The evaluation of contextual fidelity in the two case studies revealed distinct yet equally significant results.

In the historic center (Fig. 10), the conditioned inference process excelled in preserving the architectural historical characteristics of Palazzo Nissim. In the suburban context of Via Piave (Fig. 11), the greater project flexibility allowed for more innovative explorations while maintaining a convincing contextual coherence.

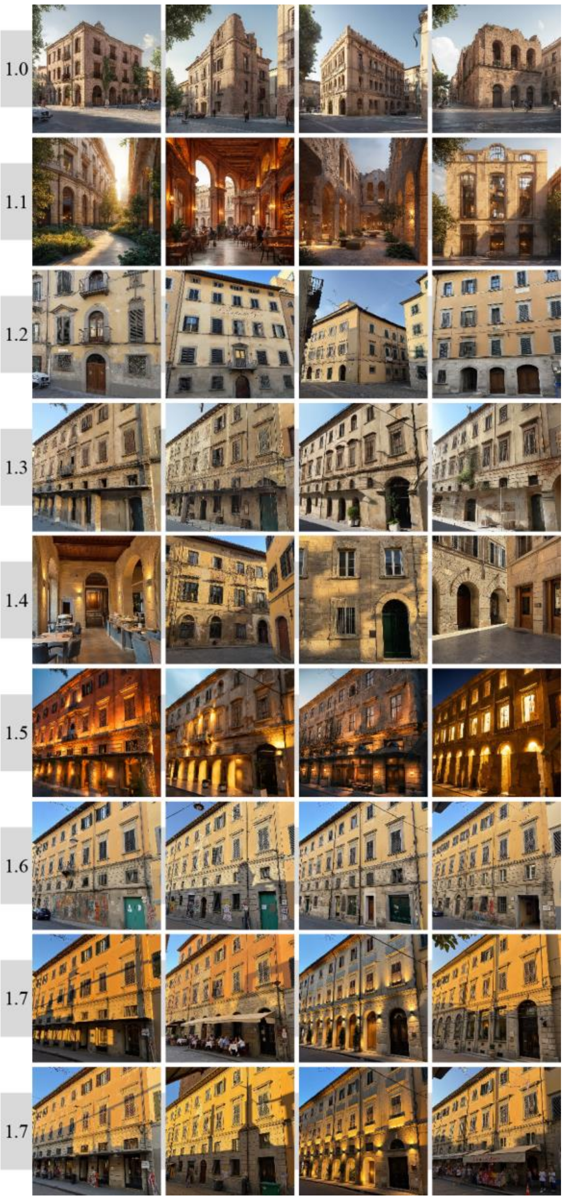


Figure 10: *Comparative analysis of generated visualizations for Palazzo Nissim, showing the progression from the base image through various conditioning stages to the final output (1.0 – Raw prompt; 1.1 – LLM; 1.2 – LoRA; 1.3 – ControlNet; 1.4 – LLM + LoRA; 1.5 – LLM + ControlNet; 1.6 – LoRA + ControlNet; 1.7 – LLM + LoRA + ControlNet).*

A particular aspect emerges from analyzing semantic pertinence in generated visualizations. The prompt engineering through LLM demonstrated notable efficacy in translating abstract concepts into concrete architectural and urban elements, with further enhancement provided by the integration with ControlNet's structural control. This behavior allowed for an optimal balance between conservation and contemporary design in the first case study while facilitating the interpretation of social requalification needs through diverse spatial solutions in the second.

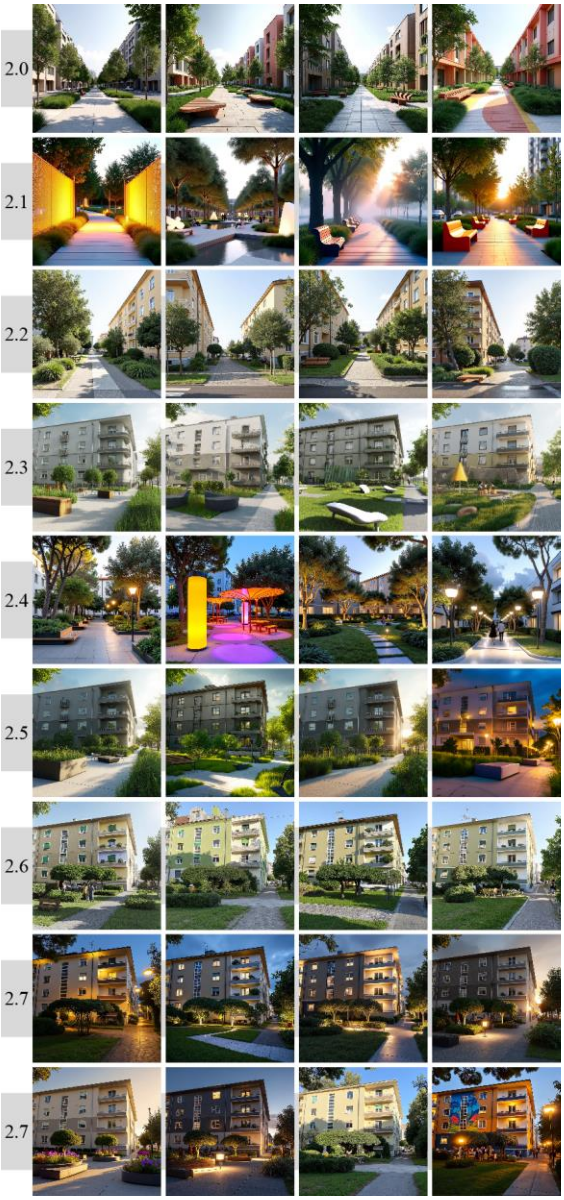


Figure 11: *Comparative analysis of generated visualizations for the Via Piave/Via Rindi suburban area, showing the progression from the base image through various conditioning stages to the final output (2.0 – Raw prompt; 2.1 – LLM; 2.2 – LoRA; 2.3 – ControlNet; 2.4 – LLM + LoRA; 2.5 – LLM + ControlNet; 2.6 – LoRA + ControlNet; 2.7 – LLM + LoRA + ControlNet).*

The generated images through the complete framework (LLM + LoRA + ControlNet) exhibited superior communicative efficacy, distinguishing themselves in their ability to represent complex project scenarios in an accessible manner while maintaining a convincing balance between realism and transformative potential. This characteristic is particularly valuable in participatory design, where communication between technical and non-technical stakeholders is key.

The overall analysis suggests that the integrated approach developed constitutes a fertile ground for further investigation in generating conditioned architectural visualizations [18], offering a suitable balance between technical accuracy and communicative accessibility.

7. CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

This research has explored the optimization of inference conditioning techniques for image generation in the context of participatory urban transformation, revealing promising results and outlining future development directions.

The primary limitations encountered are primarily related to the computational capabilities of the hardware used (NVIDIA GeForce RTX 4090 with 24GB GDDR6), despite this configuration enabling high-quality results. Access to greater computational capacity could enable the implementation of more advanced fine-tuning techniques, such as DreamBooth, or more complex base models.

It is essential to emphasize that, within these hardware limitations, the quality of images generated through the combined approach LLM + LoRA + ControlNet has reached notable accuracy and contextual coherence.

A primary direction for further development involves extending the methodology to a broader range of urban contexts, including rural areas, contemporary dense urban environments, and diverse public spaces. Applying the framework to these scenarios may reveal specific effectiveness trends and suggest contextual optimization of conditioning techniques. Developing sets of parameter-specific conditioning strategies for different urban contexts appears particularly promising.

The analysis has highlighted that success in applying these technologies is intrinsically linked to conditioning quality. While dataset availability constitutes a necessary foundation, the sophistication of conditioning techniques determines most of the resulting quality. Experience has demonstrated that approaches limiting control through textual prompts produce significantly inferior results compared to implementing multi-level conditioning strategies. This contributes significantly to architectural representation, suggesting a methodology-based paradigm focused on meticulous control of the generative process.

The potential demonstrated by these experimental tools suggests concrete possibilities for implementation through simplified and accessible interfaces. Translating these technologies into accessible tools could democratize architecture visualization, maintaining high-quality standards while leveraging advanced conditioning techniques.

The developed framework fits emerging AI regulations, including the EU AI Act [19]. In the specific context of participatory urban visualization, current regulations do not present significant obstacles to implementation, requiring primarily transparency in the process and clear communication of tool limitations.

The results obtained open up multiple directions for future research development. Quantitatively, extending to a broader range of case studies will enable the validation and refinement of conditioning techniques in diverse contexts. Qualitatively, integrating complementary strategies like inpainting and outpainting promises to increase control over the generative process further. Following the quick evolution of DMs, which continue to show significant progress in quality and control, will be fundamental.

Future research should focus not only on technical optimization but also on developing framework methodologies that facilitate the integration of these tools into professional practice. The emphasis on accessibility and user-centric design reflects a broader shift in architectural practice toward democratizing technological tools [20]. Particular attention must be paid to developing simplified interfaces to ensure advanced conditioning techniques are accessible without compromising efficacy, balancing accessibility, and comprehensive control over the generative process.

8. ACKNOWLEDGMENT

This paper is the result of a scientific research project by the authors. E. Pupi wrote paragraphs 2, 4, 4.1, 4.2, and 4.3; P. Rechichi wrote paragraphs 1 and 5. The authors wrote together parr. 3, 5.1, 5.2, 6 and 7.

9. REFERENCES

[1] Yigitcanlar, Tan, Kankanamge, Nayomi, & Vella, Karen: How are smart city concepts and technologies perceived and utilized? A systematic geo-Twitter analysis of smart cities in Australia. *Journal of Urban Technology*, vol. 28, no. 1–2, pp. 135–154, 2020.

[2] Koseki, Shin, Jameson, Shazade, Farnadi, Golnoosh, Rolnick, David, Régis, Catherine, & Denis, Jean-Louis: *AI & Cities: Risks, Applications and Governance*. UN-Habitat, 2022.

[3] Nelson, Ruth, Naing, Khant Min, & Beroche, Hubert: Generative AI for Urban Governance. *Urban AI*, 2023.

[4] Beroche, Hubert: *Urban AI*. Urban AI, Paris, France, 2021.

[5] United Nations: *Transforming Our World: The 2030 Agenda for Sustainable Development*. United Nations, Department of Economic and Social Affairs, 2015.

[6] He, Wei, & Chen, Mingze: Advancing Urban Life: A Systematic Review of Emerging Technologies and Artificial Intelligence in Urban Design and Planning. *Buildings*, vol. 14, no. 3, 2024.

[7] Quan, Steven Jige, Park, James, Economou, Athanassios, & Lee, Sugie: Artificial intelligence-aided design: Smart Design for sustainable city development. *Environment and Planning B: Urban Analytics and City Science*, vol. 46, no. 8, pp. 1581–1599, 2019.

[8] Chen, Jinliu, Wenquan, Gan, Ning, Liu, Pengcheng, Li, Haoqi, Wang, Xiaoxin, Zhao, & Yang, Di: Community Quality Evaluation for Socially Sustainable Regeneration: A Study Using Multi-Sourced Geospatial Data and AI-Based Image Semantic Segmentation. *ISPRS International Journal of Geo-Information*, vol. 13, no. 5, 2024.

[9] Zhou, Zhilun, Lin, Yuming, Jin, Depeng, & Li, Yong: Large Language Model for Participatory Urban Planning. Pre-print, *arXiv*, 2024.

[10] Wu, Yunfei, Liu, Qiqi, Hang, Tian, Yang, Yihong, Wang, Yijun, & Cao, Lei: Integrating restorative perception into urban street planning: A framework using street view images, deep learning, and space syntax. *Cities*, vol. 147, Article 104791, 2024.

[11] Guridi, Jose A., Cheyre, Cristobal, Goula, Maria, Santo, Duarte, Humphreys, Lee, Shankar, Aishwarya, & Souras, Achilleas: Image Generative AI to Design Public Spaces: A Reflection of How AI Could Improve Co-Design of Public Parks. *Digital Government: Research and Practice*, Association for Computing Machinery, New York, 2024.

[12] Nagenborg, Michael, Stone, Taylor, González Woge, Margoth, & Vermaas, Pieter E. (Eds.): *Technology and the City: Towards a Philosophy of Urban Technologies*. Springer, Cham, 2021.

[13] Zhang, Lvmin, Rao, Anyi, & Agrawala, Maneesh: Adding conditional control to text-to-image diffusion models. *In Proceedings of the IEEE/CVF International Conference on Computer Vision*, IEEE, pp. 3813–3824, 2023.

[14] Meng, Qian, Ge, Minyue, & Feng, Zhang: The integration of artificial intelligence in architectural visualization enhances augmented realism and interactivity. *Academic Journal of Science and Technology*, vol. 12, no. 2, pp. 7–12, 2024.

[15] Gan, Wei, Zhao, Zichen, Wang, Yuankai, Zou, Yixuan, Zhou, Shiqi, & Wu, Zhiqiang: UDGAN: A new urban design inspiration approach driven by using Generative Adversarial Networks. *Journal of Computational Design and Engineering*, vol. 11, pp. 305–324, 2024.

[16] Lian, Long, Li, Boyi, Yala, Adam, & Darrell, Trevor: LLM-grounded Diffusion: Enhancing Prompt Understanding of Text-to-Image Diffusion Models with Large Language Models. *Trans. Mach. Learn. Res.*, 2023.

[17] Hidalgo, Rafael, Salah, Nesreen, Chandra Jetty, Rajiv, Jetty, Anupama, & Varde, Aparna S.: Personalizing Text-to-Image Diffusion Models by Fine-Tuning Classification for AI Applications. In: K. Arai (ed.), *Intelligent Systems and Applications. IntelliSys 2023*, Lecture Notes in Networks and Systems, vol. 822, Springer, Cham, 2024.

[18] Shi, Mengnan, Seo, JoonOh, Cha, Seung Hyun, Xiao, Bo, & Chi, Hung-Lin: Generative AI-powered architectural exterior conceptual design based on the design intent. *Journal of Computational Design and Engineering*, vol. 11, no. 5, pp. 125–142, 2024.

[19] European Parliament: Legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of

the Council on laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206—C9-0146/2021—2021/0106(COD)). *European Parliament Legislative Resolutions*, 2024.

[20] Dehouche, Nassim, & Dehouche, Kullathida: What's in a text-to-image prompt? The potential of stable diffusion in visual arts education. *Heliyon*, vol. 9, e16757, 2023.