A New Soul of Architecture

AI-directed drawing

Fabio Bianconi[1], Marco Filippucci[2], Andrea Migliosi[3], Chiara Mommi[4]
[1,2,3,4] University of Perugia
[1,2]{fabio.bianconi|marco.filippucci}@unipg.it
[3,4]{andrea.migliosi|chiara.mommi}@dottorandi.unipg.it

This study explores the integration of generative artificial intelligence into architectural creative processes, examining how its inherent stochasticity can be oriented and structured within defined design trajectories. Through the adoption of nodal algorithms, the research develops a methodology to balance computational autonomy with design control, establishing the algorithmic curator as a critical mediator between human intentionality and machine-generated potential. Therefore, nodal algorithms are examined not merely as instruments of automation or stylistic emulation, but as operational frameworks that foreground co-authorship between human and machine, emphasizing control strategies rooted in semantic and procedural constraints. Within the ComfyUI framework, the nodal algorithm architecture organizes the creative process into phases that are discrete yet dynamically linked, with each node serving multiple functions as a control point for design parameters, a filter for semantic and stylistic coherence, and a bridge between human input and machine interpretation. The system incorporates DreamBooth-trained models fine-tuned on architectural datasets, allowing for specialized generation capabilities while maintaining the flexibility needed for creative exploration. Applied to the complex case study of historic urban centers, this approach demonstrates how AI-assisted design can operate within culturally sensitive contexts. The methodology transforms architectural representation into an iterative dialogue between designer and algorithm, where human expertise guides the system's stochastic processes toward meaningful design solutions. Through this co-creative process, the research redefines digital-age authorship while preserving the expressive richness of AI generation, ensuring alignment with design intent through dynamic constraints, and producing hybrid design languages that meaningfully bridge historical context with contemporary innovation.

Keywords: *Algorithmic Design, Architectural Representation, ComfyUI, Style Transfer*

INTRODUCTION

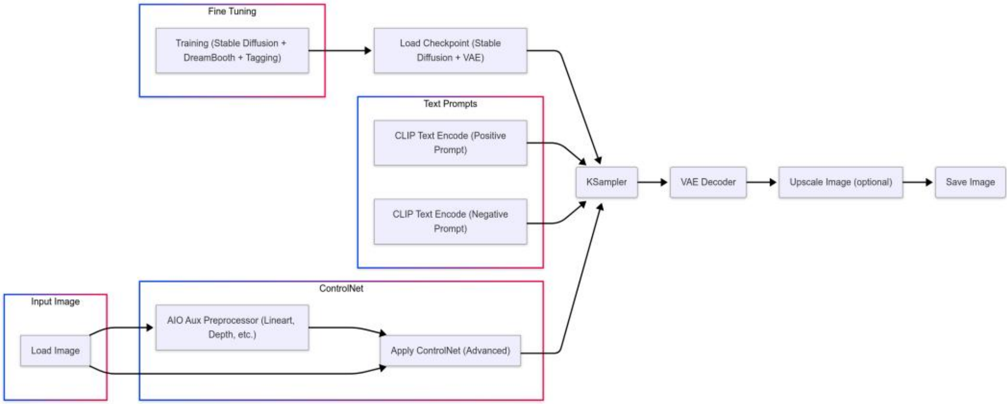
The primary objective of this research lies in investigating how generative artificial intelligence technologies, when integrated into the creative process (Hertzmann 2018; McCormack et al. 2024), can be guided toward meaningful design outcomes within a dialectical equilibrium where algorithmic unpredictability is not suppressed but channelled through control strategies capable of transforming stochastic behavior into a heuristic resource. This approach redefines the boundaries of architectural creativity in an era where the design act increasingly configures itself as a hybridized process shaped by the continuous interplay between human intentionality and computational autonomy.

To achieve this goal, it becomes necessary to identify those essential factors that enable the user, a figure who must synthesize within themselves the roles of designer, critic, and cultural mediator, to harmonize the generative potential of AI with the necessary project coherence, in a process where the intrinsic stochasticity of neural systems is not denied but channelled through a series of dynamic constraints acting as semantic and stylistic filters, ensuring that outputs maintain an organic connection to the project's original intent (Del Campo and Carpo 2022) without relinquishing the combinatorial richness typical of generative models. A possible response to this need has been identified in nodal algorithms, which, thanks to their procedural structure, configure themselves as critical interfaces capable of articulating the generative flow into a series of discrete yet interconnected steps, where each node represents a moment of verification and reorientation of the creative process, thereby preserving the complexity of reality without succumbing to the representational anarchy that often characterizes purely generative approaches. Through this methodological lens, it becomes possible to explore the ways in which the relationship between designer and artificial intelligence can be redefined (Bölek, Tutal, and Özbaşaran, 2023) in light of the potential offered by nodal algorithms (Yeming, Song, and Keiman Yip 2024), which transform the traditional input-output paradigm into a complex relational system where creativity emerges from the dynamic interaction between human and machine (Akcay Kavakoglu et al., 2022), in a continuous exchange between control and unpredictability that redraws the very boundaries of authorship in the context of contemporary architectural production (Carpo 2023). This process defines the figure of the algorithmic curator (Bianconi et al., 2024), who goes far beyond merely issuing commands to the system, actively negotiating with its generative potential and guiding results through a series of conscious choices that intervene at nodal points in the process, where stochasticity is channeled toward outcomes consistent with a precise design vision while remaining open to the heuristic surprises that only interaction with a non-deterministic system can offer. From this perspective, the nodal algorithm (Hyung 2024) transcends its role as a technical tool to become a philosophical device, embodying a vision of architectural design as a hybrid practice where manual tradition and computational innovation do not oppose but mutually enrich one another. This generates a language capable of reconciling the historical depth of the discipline with the infinite possibilities of the digital, within a cultural horizon where creativity is no longer the exclusive prerogative of humans nor a mere product of algorithms, but rather the result of a complex and ongoing dialogue between these two polarities, mediated by the curator as a critical interpreter of a process that transforms both machine and human, thereby opening new scenarios for the representation and conception of architectural space in the age of artificial intelligence.

METHODOLOGY

The development of the workflow for image generation and transformation was driven by the need to combine accessible neural models with modular open-source tools, prioritizing ease of customization and transparency over proprietary complexity. This choice was made after evaluating alternative frameworks, opting for a configuration that could flexibly manage the complexity of visual processing while remaining replicable in standard academic settings. In this perspective, the adoption of ComfyUI as a node-based development environment represents a strategic choice, as it enables a graphical representation of the interconnections between the various operational modules, ensuring not only a clear visualization of the processing flow but also a dynamic management of dependencies among system components. This modular approach was found to be particularly advantageous during iterative refinements of the pipeline, as it allowed for targeted substitution and reconfiguration of nodes in response to training inconsistencies or visual incoherencies observed in early testing phases (Figure 1). This configuration ensures a clear functional separation between the various modules, each optimized for a specific task yet integrated into a unified execution flow that facilitates reproducibility of results and system adaptability to different application scenari. The combined use of pre-trained models, transfer learning techniques, and advanced conditional controls establishes an innovative paradigm for visual content generation across domains ranging from industrial design to architectural representation and digital media production (Pandy , Pugazhenthi and Murugan, 2025). The structural flexibility of the architecture, combined with fine-tuned regulation of operational parameters, enables the handling of a wide range of design contexts, anticipating future developments in the integration of multimodal models and increasingly advanced machine learning techniques. A crucial element of the workflow is the customization of the model through DreamBooth, a fine-tuning technique that allows the adaptation of a pre-existing model to a specific dataset. This methodology is based on the progressive optimization of model weights through incremental training, where dataset images are correlated with unique identifiers, allowing the system to learn highly specialized latent representations. The effectiveness of DreamBooth is further amplified by the integration of semantic annotations derived from preprocessing phases, guiding the model in associating visual features with textual descriptions, thereby improving the alignment of results with design intentions. The training process is conducted locally via a dedicated script, enabling autonomous management of the entire pipeline without reliance on external infrastructures. This approach integrates a tagging system directly within DreamBooth, eliminating the need for external annotation tools such as the WD14 tagger.

Figure 1
Flowchart of the workflow.



As a result, workflow efficiency is enhanced, consistency between metadata and model parameters is increased, and potential semantic discrepancies between training datasets and generated outputs are minimized. The success of the fine-tuning phase is highly dependent on the quality and coherence of the input dataset, which must consist of images selected according to strict criteria regarding relevance to the application domain, stylistic uniformity. The standardization of dataset dimensions is not merely a technical precaution but rather a fundamental methodological choice to ensure computational efficiency during training, preventing the emergence of artifacts or distortions caused by inconsistent image size management. Dataset preprocessing also includes essential operations such as the removal of extraneous elements, color correction, and stylistic harmonization, aiming to create a homogeneous data corpus that facilitates model learning and enhances its generalization capabilities. At the core of the proposed pipeline is Stable Diffusion 1.5, an image generation model based on diffusion architectures (Saharia et al., 2022), selected for its proven stability and its ability to produce high-quality visual outputs through an iterative denoising process guided by textual instructions. The choice of this specific version is motivated both by the model's maturity, derived from training on extensive and diversified datasets, and by its compatibility with a broad range of auxiliary tools, which facilitate its implementation within complex workflows. The architecture of Stable Diffusion is distinguished by its robustness in image generation and its predisposition for subsequent fine-tuning phases, enabling targeted specialization for specific application domains without compromising the structural coherence of the generated outputs.

A key element of the pipeline is the integration of ControlNet (Liu et al., 2024), an advanced framework that allows deterministic control over the generative process, constraining the output to structural and stylistic properties defined by a reference image. ControlNet is applied through the "Apply ControlNet (Advanced)" node, which operates in synergy with the "AIO Aux Preprocessor" module to extract salient features from the guide image—such as contours, depth maps, or semantic segmentations—and inject them into the generative process as binding conditions. This mechanism enables the separation of content generation from the underlying composition, making it possible, for instance, to maintain the architectural geometry of an image while modifying its visual style. The fine-tuning of ControlNet's influence (Zavadski, Feiden and Rother, 2024) is managed by specific parameters, such as strength, which determines the intensity of the binding effect, and start/end percentage, which defines the temporal interval of the diffusion process in which the conditions are active, allowing for an optimal balance between adherence to imposed constraints and creative freedom.

Finally, image generation is carried out through the KSampler node, responsible for iterating the diffusion process according to a highly customizable parametric configuration. Key parameters include steps, which determine the number of denoising iterations applied, directly influencing the quality and definition of the final image; cfg (Classifier-Free Guidance), which regulates the trade-off between adherence to textual instructions and output variability; and denoise, which modulates the degree of removal of the initial noise, with direct consequences on fidelity to input data. The image generated in the latent space is finally decoded into the visual space through the VAE Decoder module, which transforms the compressed representation into a high-resolution image, potentially subjected to upscaling techniques to enhance definition without introducing visual artifacts.

CASE STUDY

The investigation of training modalities and the specificity of the object on which training is conducted emerges as an essential operation to understand how generative neural networks can absorb and reproduce distinctive stylistic patterns.

This issue is particularly relevant in architectural representation, where line quality, compositional logic, and graphic conventions serve both as distinctive elements of authorship and as cognitive structures underlying specific ways of conceiving and translating space into two-dimensional signs. To this end, the study was structured through the training of three distinct models based on the Stable Diffusion 1.5 architecture, each trained on a specific dataset with the goal of evaluating the AI's ability to assimilate, reinterpret, and interpolate graphical sets characterized by different expressive rules. The analysis focused on the system's ability to recognize recurring structural patterns and how these are synthesized in subsequent generations, requiring careful balancing of training methodologies to prevent overfitting or incorrect stylistic attribution.

The entire process was implemented on a hardware configuration that, while not optimal for large-scale training, was adapted to computational needs through advanced execution parameter tuning. The setup included an RTX 3070 GPU with 8GB of dedicated memory and 16GB of RAM, necessitating targeted VRAM management through specific adjustments in the training pipeline configuration, and particularly the definition of a reduced batch size and the use of gradient accumulation strategies to mitigate memory capacity limitations.

These measures enabled the execution of backpropagation operations, ensuring stable model convergence without compromising weight update quality.

The next phase of the process involved the careful selection of the training dataset, a crucial step to ensure output coherence. The first model was trained exclusively on hand-drawn sketches by masters of the Roman school of architecture, a corpus characterized by strong internal consistency due to the shared graphic lexicon - an expression of a representational method in which line construction follows principles of volumetric synthesis and perspectival regularity (Figure 2).

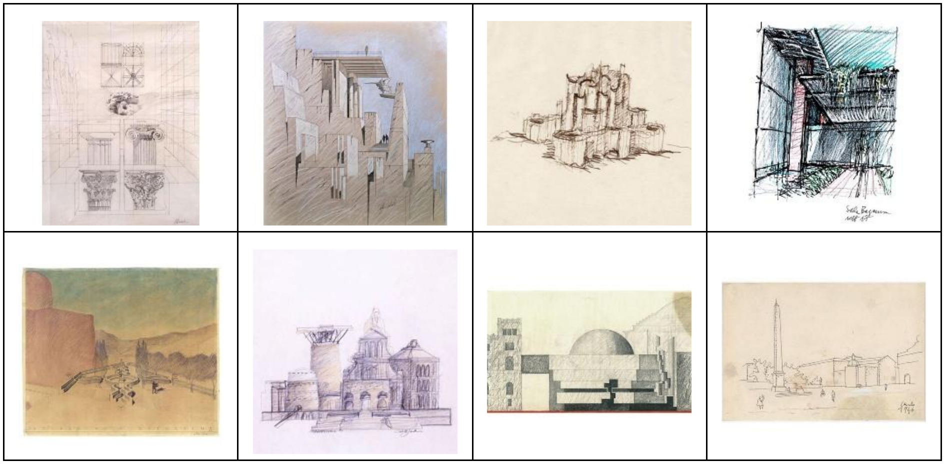


Figure 2
Some images used in the training for the Roman School Checkpoint.

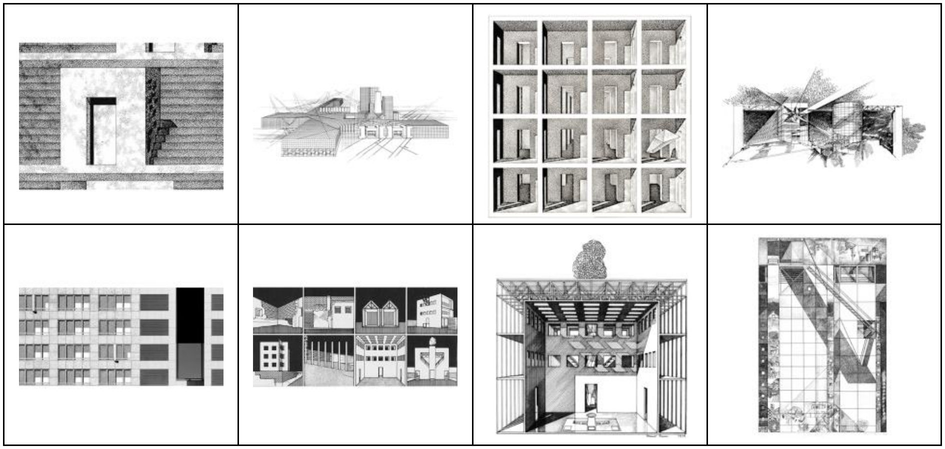


Figure 3
Some images used in the training for the Franco Purini Checkpoint.

The second model underwent training exclusively on drawings by Franco Purini, whose stylistic signature is distinguished by a different approach to linework - more analytical and marked by a spatial decomposition tendency that fragments the image into structured fields according to a modular principle (Figure 3). This contrasts with the more homogeneous fluidity of the previous corpus, making the comparison between the two training sets particularly significant for analyzing the AI's behavior in recognizing and reinterpreting different graphic grammars. Finally, the third model was trained on both datasets to examine whether the system would produce a hybrid language resulting from the fusion of the two reference matrices or if it would maintain a clear distinction between the two drawing modalities. This question is central to understanding how the latent model of Stable Diffusion organizes learned information and structures image generation based on stylistic categories that, from a human perception perspective, are distinguishable but that, for a neural network, might represent mere weight variations within the same latent space. This aspect was further investigated through a series of comparative tests conducted on successive generations, monitoring variations in latent vectors in response to identical prompts submitted to the three models. The objective was to verify the extent to which differences in training datasets were reflected in the AI's graphic production. A first experiment was set up by excluding the use of ControlNet, keeping the standard K-Sampler parameters unchanged and limiting the prompt to the single use of the term "drawing" to evoke the tagging inherent to the dataset. This configuration was adopted with the intent of probing the model's capacity to assimilate the fine-tuning process. An initial examination of the outputs indicates that the model inspired by the Roman School reproduces, albeit not in strict accordance with the full stylistic complexity, the essential elements of that tradition, in a process that evokes rather than directly imitates, through a subtle reminiscence of the original. In the case of the model trained with Franco Purini's drawings, however, emerges an immediate awareness of the expressive necessity of black and white: the lines, clear and decisive, intertwine with a hint of pattern, suggesting a dialectic between rigor and formal suggestion. Ultimately, the hybrid model, resulting from the convergence of the two stylistic sources, does not

Figure 4
First style transfer tests, left to right: Roman architecture referen ce, Franco Purini-inspired aesthetic, hybrid composition merging both.

achieve a homogeneous fusion but rather manifests a plurality of zones in which the specificities of Purini and the Roman School coexist, forming a complex mosaic that recalls the dialectical tension between difference and unification (Figure 4). The input image selected for the subsequent experimental phase is the relief of the historic center of Perugia, a dense and layered urban fabric whose morphological structure presents itself as a palimpsest of signs accumulated over time (Figure 5). The aim of this experimental phase is to test the effectiveness of style transfer by evaluating how well the model manages to reinterpret the input image through the assimilation of expressive codes inferred from the training dataset. ControlNet was used in the process, configured with a calibrated setting to detect contour lines and estimate scene depth. However, in order not to overly constrain the model to a rigidly structured reproduction, the parameters were intentionally maintained at moderately values (1.5), allowing for a margin of variation that reveals an interplay between the structural fidelity of the image and the degree of interpretative freedom granted to the neural network. This methodological approach aims to verify not only the algorithm's capacity to preserve the spatial framework of the original, but also the extent to which the introduced transformations can constitute an added value, generating a re-elaboration that transcends a mere stylistic transposition and instead functions as a critical device capable of rewriting the visual data through a new graphic syntax. In this perspective, the application of style transfer assumes a value that transcends the mere overlapping of graphic codes, serving instead as a tool for re-elaboration and interrogation of established visual grammars.

The first output image obtained, evocative of the Roman School tradition, is configured as a reinterpretation of an expressive language imbued with historical memory and painterly sensitivity. The color washes do not simply reproduce the essence of the style; rather, they distill its evocative character, transforming the algorithmic transposition into a critical act of synthesis and reformulation. The apparent static nature of the composition conceals a semantic vitality, wherein the tension between technique and intuition dissolves the boundary between the human gesture and computational logic (Figure 6). Otherwise, in the second image, the training with Franco Purini's drawings introduces a dialectic between rigor and fragmentation, between the precision of the stroke and the density of the spatial grid. Here, the generative intervention adopts an analytical stance: the lines become clear, rigorous, almost imperious, articulating a system of modular relationships that echoes Purini's approach to architectural representation.

In this case, the algorithm does not merely translate a graphic code, but rather amplifies the intrinsic tension between mathematical abstraction and expressive depth (Figure 7). At last, in the third output image, the experiment reaches its point of greatest complexity, taking shape as a dialectical hybridization between the two stylistic instances. Here, the fusion between the interpretive fluidity of the Roman School and Purinian modular rigidity is resolved, as in the first test, into a juxtaposition, but despite this, a more subtle and dynamic interweaving is found. The third image is distinguished by a clear prevalence of the typical sensibility of the Roman School, evidenced by the ability of the gestural sign to overlap and reconfigure itself within the composition.

This outcome takes the form of the emergence of a structure in which aesthetic tensions are dynamically articulated: the expressive flexibility of the Roman tradition succeeds in maintaining and modulating tensions in a hybrid context, while the rigidity of the Purinian syntax, with its modular character, limits flexibility and imposes itself less incisively in the process of stylistic fusion (Figure 8).

CONCLUSIONS

The analysis conducted through the application of style transfer in this specific experimental framework suggests a promising potential for generative processes in architectural representation, posing questions that transcend the mere technical dimension to invest the theoretical and disciplinary domain. While artificial intelligence demonstrates that it can assimilate and reinterpret established graphic codes, its work is not reduced to passive reproduction, but takes the form of a true act of rewriting visual language.

In this investigation, the relationship between architectural input and algorithmic processing proved to be nonlinear, but rather dialectical: the model did not limit itself to faithfully transposing the stylistic features of the Roman School or the Purinian

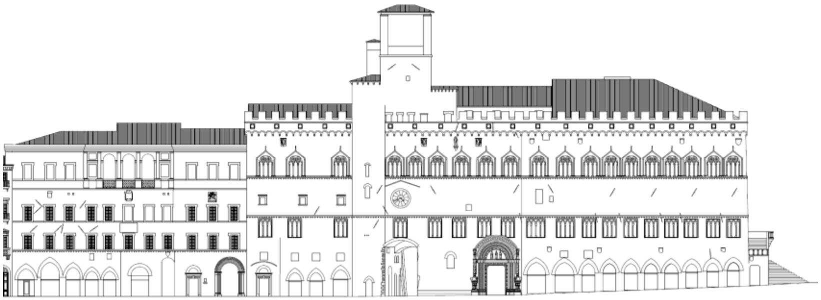


Figure 5
Image of the survey of Perugia's historic center used as input for the style transfer.

Figure 6
Reinterpretation of the urban survey using the graphic language of the Roman School.

style trait, but rather activated processes of synthesis, hybridization, and sometimes unexpected distortion. The result is an intermediate expressive field, in which traditional visual grammars open up to new modes of interpretation, generating images that reside in a borderline dimension between memory and innovation. However, the results obtained from the model trained on both styles simultaneously suggest that stylistic unity constitutes a surplus value. The hybrid image does not achieve the expressive clarity observed in the singularly trained outputs, indicating that contamination does not necessarily generate a qualitatively superior synthesis. Instead of producing a cohesive and balanced reinterpretation, the combination of styles results in an ambiguous and less effective visual outcome, reaffirming the importance of a clear and coherent stylistic framework. This suggests that the use of AI in architectural design and representation should not be considered a mere automation tool, but rather a heuristic device capable of expanding the spectrum of design possibilities.

The algorithm does not replace the author, but becomes its active interlocutor, stimulating reflections on the epistemological value of the sign, the translatability of styles, and the very nature of drawing as a critical operation.

At a time in history when the boundary between technique and creativity appears increasingly blurred, these experiments open up a research perspective in which AI does not merely reinterpret the past, but becomes a platform for the exploration of new forms of expression. The findings indicate that rather than seeking hybridization as an end in itself, future research should explore the potential of AI to refine and reinforce distinct stylistic identities. These findings suggest educational potential, where AI-driven stylistic negotiation can support the exploration of authorship and representation within studio-based learning contexts, fostering critical reflection on stylistic formation and disciplinary frameworks.

REFERENCES

Akçay Kavakoğlu, A., Özgen, H., Barlas, A., Kömürcü, E. and Karagöz, B. (2022) 'AI Driven Creativity in

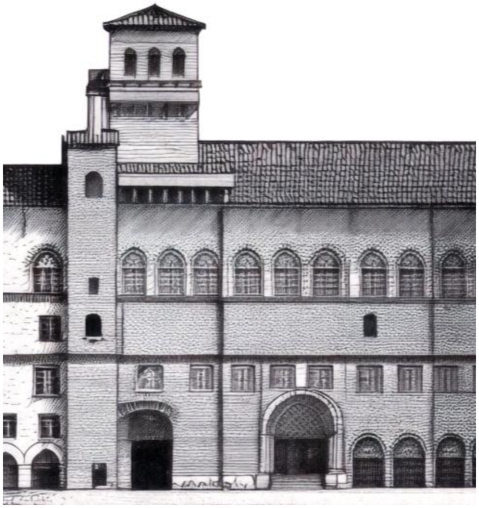


Figure 7
On the left, reinterpretation of the urban survey using the language of Franco Purini.

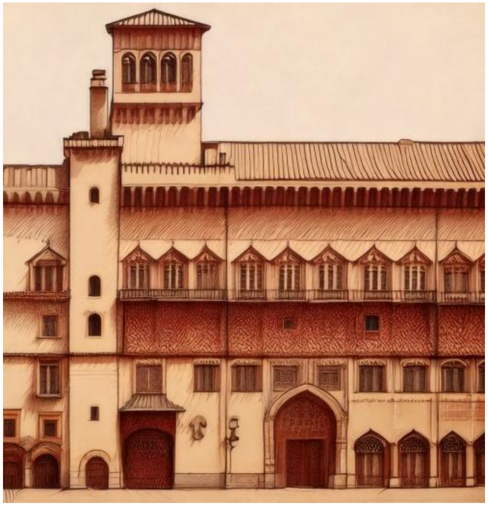


Figure 8
On the right, reinterpretation of the urban survey using the Hybrid language of the Roman School and Franco Purini.

Early Design Education - A pedagogical approach in the age of Industry 5.0', Pak, B., Wurzer, G., Stouffd, R., (eds) *Co-creating the Future: Inclusion in and through Design - Proceedings of the 40th eCAADe 2022 Conference,* 13-16 September 2022, Ghent, Belgium, pp. 133-142. Available at: https://doi.org/10.52842/conf.ecaade.2022.1.133 (Accessed: 8 Febraury 2025).
Bianconi, F., Filippucci, M., Migliosi, A. and Mommi, C. (2024) 'AI-D. Algorithmic Experiments with AI to Aid Drawing', *SCIRES-IT-SCIentific RESearch and Information Technology*, 14(2), pp. 31–48. Available at: http://dx.doi.org/10.2423/i22394303v14n2p31 (Accessed: 15 January 2025).
Bölek, B., Tutal, O. and Özbaşaran, H. (2023) 'A systematic review on artificial intelligence applications in architecture', *Journal of Design for Resilience in Architecture and Planning*, 4(1), pp. 91–104. Available at: https://doi.org/10.47818/DRArch.2023.v4i1085 (Accessed: 10 March 2025).
Del Campo, M. and Carpo, M. (2022) *Neural Architecture: Design and Artificial Intelligence*. Novato: Applied Research & Design Publishing.
Carpo, M. (2023) *Beyond digital: design and automation at the end of modernity*. Cambridge, Massachusetts: MIT Press.
Hertzmann, A. (2018) 'Can Computers Create Art?', *Arts*, 7(2), pp. 18. Available at: https://doi.org/10.3390/arts7020018 (Accessed: 15 January 2025).
Hyung, K.J. (2024) 'A Study on Visual Art Production through Collaborative Image Generation Using Artificial Intelligence and Unreal Engine', *Korean Society For Computer Game,* 37(3), pp. 85–91. Available at: https://www.earticle.net/Article/A456066 (Accessed: 10 March 2025).
Liu, X., Wei, Y., Liu, M., Lin, X., Ren, P., Xie, X. and Zuo, W. (2024) 'SmartControl: Enhancing ControlNet for Handling Rough Visual Conditions', in *Computer Vision - ECCV 2024*, 29 September - 4 October 2024, Milan, Italy, pp. 1-17. Milan: Springer.
McCormack, J., Llano, M.T., Krol, S.J. and Rajcic, N. (2024) 'No longer trending on artstation: Prompt analysis of generative ai art', in *International Conference on Computational Intelligence in Music, Sound, Art and Design*. 3-5 April 2024, Aberystwyth, United Kingdom, pp. 279–295. Aviable at: https://doi.org/10.1007/978-3-031-56992-0_18 (Accessed: 8 Febraury 2025).
Pandy, G., Pugazhenthi, V.J. and Murugan, A. (2025) 'Generative AI: Transforming the Landscape of Creativity and Automation', *International Journal of Computer Applications*, 186(63), pp. 7–13. Available at: 10.5120/ijca2025924392 (Accessed: 15 January 2025).
Saharia, C. et al. (2022) 'Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding', in : S. Koyejo and S. Mohamed and A. Agarwal and D. Belgrave and K. Cho and A. Oh (eds*) Advances in Neural Inform*ation *Processing Systems 35 (NeurIPS 2022),* 28 November 2022- 9 December 2022, New Orleans, USA, pp. 36479 - 36494. New Orleans: Neural Information Processing Systems Foundation, Inc. (NeurIPS)
Yeming, L., Song, J. and Keiman Yip, D. (2024) 'AI-Assisted Content Creation of Naked-Eye 3D Effects on Curved LED Screen: Enhancing Artistic Expression and Creativity', in *2024 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*. pp. 1–5. Available at: 10.1109/ICMEW63481.2024.10645467 (Accessed: 10 March 2025).
Zavadski, D., Feiden, J.-F. and Rother, C. (2024) 'ControlNet-XS: Rethinking the Control of Text-to-Image Diffusion Models as Feedback-Control Systems', in *Computer Vision - ECCV 2024,* 29 September – 4 October 2024, Milan, Italy. Springer. https://doi.org/10.1007/978-3-031-73223-2_20