Original Research Article

International Journal of Architectural Computing
2025, Vol. 0(0) 1–15

DOI: 10.1177/14780771251346584
journals.sagepub.com/home/jac
Sage

Revisiting architectural generation with structural and interpretive image-to-image AI approaches

Xavier Marsault

Abstract

The paper presents a new way of obtaining original representations of architectures well integrated into their environment, far from the generation mode supervised by textual semantics. It illustrates the compositional nature of deep learning, a perfect example of the organized complexity theories popularized by Christopher Alexander and Nikos Salingaros. Initially, the user assembles a collection of images chosen according to his or her preferences for style, structure, environment and rendering type. The challenge was to take these properties into account in a single generative chain, without using textual descriptors of the architecture. The concept developed combines two complementary approaches of Generative AI: structural (framework) and interpretive (detail and finish). It replaces and complements the classic notion of style transfer (global, often from a single image, with a fairly fixed rendering) with organic interpretation (multi-scale parts being associated to find a local hidden style). The generative abundance enabled by this coupling is illustrated on a small learned dataset of 4300 images. It reinforces the promising impression of this type of exercise, which, in our opinion, has been neglected too much since the appearance of GAN in 2014.

Keywords

Architecture, detail, generative AI, GAN, image-to-image, interpretation, representation, structure

Introduction

An architect's creativity in the ideation phase

The multidisciplinary profession of architect combines design, production and project management. It involves a variety of tasks ranging from the production of technical and descriptive texts, detailed plans and images, to building analysis, evaluation and simulation, as well as site supervision. Creativity comes into play when solving a problem to find a new and original solution. It often follows the ideation process, which

URM MAP, Aria Lab., ENSA de Lyon, Lyon, France

Corresponding author:
Xavier Marsault, ENSA de Lyon, MAP-Aria lab., 3 rue Maurice Audin, 69512 Vaulx en Velin, Lyon, France.
Email: xavier.marsault@lyon.archi.fr

mobilizes language and visual artefacts to give a concrete expression to project intentions. In this phase, the architect defines concepts, aspirations and project attributes, then associates them with visual elements to illustrate possible ambience and solutions.

The creative process in architecture embodies the richness and diversity of thought, where the project emerges slowly through exploration, reflection and experimentation. Negrotti suggested that AI could treat the intelligence and creativity of machines as an interest in itself, rather than as a way of understanding those of human activity.[1] In this sense, G-creativity, adapted to AI, is based on the power to generate valuable novelty, and is distinct from the knowledge and problem-solving process.[2] Generative AI has become a complementary tool to the reference searches phase at the start of the project design. By playing with learned graphic, structural and stylistic archetypes, it produces original expressions of architectural form (Figures 1–4) and opens up new creative, aesthetic and plastic spaces capable of stimulating architects' creativity.

Architectural representations

Architectural artefacts are useful not only for designers of buildings and urban spaces (architects, urban planners) but also increasingly in the visual arts, from illustration to the creation of games and films. In this age of foundation models – large language models, semantic image and video generators – it is undeniable that a new source of creativity has emerged from the stimulating exploration of certain Generative AI models that have learned, within hundreds of millions of annotated images, multimodal elements determining many styles (a style is a notion grouping graphic, structural and compositional characteristics).

Guided by semantics and visual conditioning, these AIs show an ability to semi-autonomously produce original, attractive and sometimes ingenious (but not necessarily realistic) expressions of architectural form, in the mode of representation, combining styles and processing architecture and landscapes as a harmonious whole. It should be noted here that these are orphaned (uninformed) representations of architecture, not projects defined by a coordinated and coherent set of documents (plans, sections, façades, perspectives, annotations, etc.).

But the main question we raise in this article is: what generative power is hidden within a simple collection of a few thousand targeted images, which can be the reference library of a student, a teacher or an architectural firm? If one image is worth more than several words,[3] what about thousands of images?

Objectives and method

We propose a generative concept combining only unsupervised image models (no textual semantics), aimed at producing rich architectural representations. It differs from the usual semantic generation (based on textual prompts, which requires foundation models with millions of images) by taking as input a limited collection of

Figure 1. Typical architectural representations generated by our method. The overall structure is produced by a GAN, while details are added with an interpretive and upscaling tool, resulting in high-quality pictures. A thousand results are available at: https://www.flickr.com/photos/201879342@N08/albums.

Figure 2. Example of a generated urban perspective: (left) with StyleGAN3 (512x512 pixels), (right) interpretive with KREA (1024x1024 pixels). While StyleGAN3 clearly outlines the main building, street and vegetal structures, KREA provides convincing detail at each level, adding pillars, doors, windows, tree trunks and leaves, decorative elements and even shadows, respecting the colour range of the native image.

images. Indeed, most architects don't have access to massive datasets or the training capacity needed to create original Generative AI models.

Our method combines a structuralist approach and an adaptive approach to deal with scales of detail through interpretive enrichment. The first starts with a limited amount of personal data to obtain an initial I2I[1] model for personalized learning via a *state-of-the-art* GAN: StyleGAN3. Users compose their own database of images that make sense to them (style, structure, natural and built environment, rendering type). Only then do they call up an I2I foundation model to enrich the former's productions. This is the adaptive approach, which opens the door to multiple levels of interpretation in terms of visual semantics and the addition of realistic details, particularly with tools such as KREA AI Enhancer, Magnific or *Leonardo AI Upscaler*.

Contributions to the scientific community

We've already published an online tutorial simplifying the installation of StyleGAN2-ada/3 under Ubuntu (20.04, 22.04 LTS): https://medium.com/@xmartialm/installing-stylegan3-on-ubuntu-9ddeb11fc6cb. It is far more stable and straightforward than the official guide provided by Nvidia (https://github.com/NVlabs/stylegan3).

Moreover, we provide on-line (https://huggingface.co/xmarsault):

(1) A dataset of 4300 images at 512 × 512 pixel resolution, based in 2021 on 700 illustrations from the work of architecture students, then progressively enriched with architectural perspectives located mainly in urban areas.
(2) The best-trained model (minimum FID checkpoint file) for each StyleGAN version, according to Table 1, directly usable for inference.

The results of more recent experiments on an expanded dataset will be available online, our huggingface repository, in 2025.

Figure 3. Example of a generated urban perspective: (left) with StyleGAN3 (512 × 512 pixels), (right) interpretive with KREA (1024 × 1024 pixels). While StyleGAN3 clearly outlines the main building, street and vegetal structures, KREA provides convincing detail at each level, adding pillars, doors, windows, tree trunks and leaves, decorative elements and even shadows, respecting the colour range of the native image.

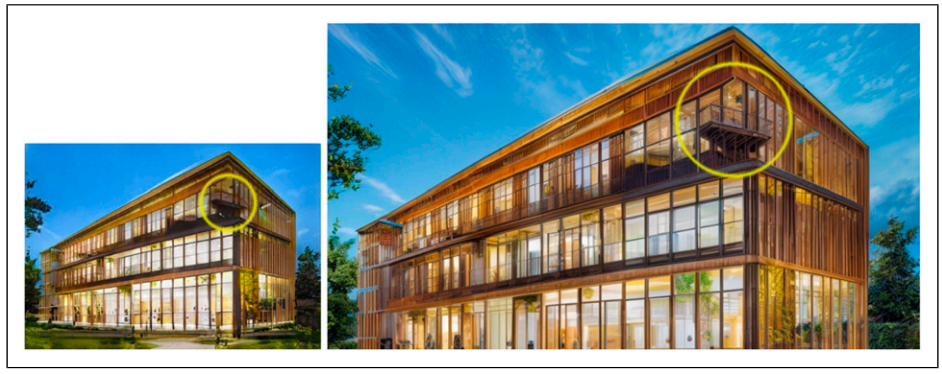

Figure 4. An example of detail generation with surprise: KREA's interpretation of a sloping balcony with its structural support well integrated into the facades, a feature that emerges in the generative process.

Background

Strengths and weaknesses of text-to-image embedding (T2I)

General AIs such as Stable Diffusion (SD) and Midjourney are among the most widely used applications in the architectural community (Dall.e.3, Imagen, Kandisky and Leonardo are also available). Coupled with SD, non-architecture-specific tools such as ControlNet help generate images from sketches, photographs of models or 3D renderings. Specialized AIs such as Veras and Lookx generate renderings directly from 3D models created in architecture-specific modellers (Sketchup, Revit, Vectorworks, Rhinoceros, Forma). What these applications have in common is their reliance on T2I diffusion models, where a semantic instruction, called a prompt, guides the generator. This prompt defines linguistic and/or pictorial contexts conditioning image production.

On the one hand, advances like *latent diffusion models* (LDM) have greatly improved text-to-image (T2I) generation. LDM,[4] which forms the basis of Stable Diffusion, performs the diffusion process in the latent

space with textual information injected into the underlying network (UNET) through a cross-attention mechanism, bringing reduced computational complexity and improved generation fidelity. But LDM still struggles to balance fine-grained precision with high-level control.[5]

On the other hand, even if T2I models have made huge progress since 2021, it is important not to focus visual content generation solely on language, for several reasons that are often forgotten or underestimated:

(1) The enormous difference in dimensionality between the possible space of text and the possible space of images leads to necessarily reduced generation capacities. This weakness of T2I generators is highlighted in a recent video by Le Cun and Friedman.[6] Text embedding in a 2D space is insufficient in terms of information to address the entire image space.
(2) The underlying question of the ergodicity of generative processes, especially as there is no universal generator of shape,[7] nor images, probably.
(3) The cognitive bias induced by the imposed human annotation canvas or automatic captioning (CLIP-type). Numerous semantic notions, at several levels, escape the relatively standardized exercise.
(4) Finally, prompt conveys a strong impression of control, in the very human mode of command or injunction. But since 2021, prompt engineering – which aims at efficient editing of the textual initiator – is a difficult art, and its sequential conditioning steps are rather complicated to implement. It remains the subject of much research. The dimensionality principle recalled earlier makes it rather illusory to claim to obtain the desired result, especially with the consideration of multiple constraints other than text, which will filter potentialities in directions unknown to the user. Practitioners of generative architecture in T2I mode often find themselves in a posture of premature acceptance of results progressively conveyed by the tool,[8] rather than with real adapted control and real editing possibilities.

Organized complexity and compositionality

Le Cun has explained the success of deep learning on the basis that the world is complex and compositional, with a hierarchy of level.[9] Its networks act as immense sponges for complexity (captured at all scales, in myriad natural and man-made objects). They give it unrivalled generative power: more relevant correlations, finer models, and more faithful generated data.

To better understand the conceptual background of architectural complexity, and show how we draw on it in our article, let's take a look back at two world-renowned researchers, Alexander and Salingaros, who have devoted a long period to understanding the organized complexity of nature and human artefacts, particularly in architecture.[10,11] Their research has crossed, over the years, many analytical points of view on living structures and human constructions, revealing an uncommon scientific quality and depth. Dealing especially with hierarchical complexity in architectural form, both have shown that the methods for generating organized complexity adapted to human physiology and sensitivity are to be found in adaptive design techniques that can be implemented in algorithms.

Alexander's theory of organized complexity[10] culminated in an analytical grid based on 15 fundamental properties in nature and art.[10] The perception of a certain geometry formed by this set of properties characterizes all-natural and artificial structures, including architecture. Salingaros's theory of adaptive complexity[11] completed Alexander's work with a methodological guide based on seven structuralist rules of design.[12] He revealed that everything you need to do to design adaptively, even simple images, is to organize and enrich the emergent complexity generated at each stage, paying attention to all scales and taking care not to introduce disconnection.[12] In the field of generative architecture, following these rules can lead to a strong structural approach, without which it seems difficult to innovate in morphogenesis.[13]

Our methodology may be considered as an intermediate between their theory on adaptive structures and an analogic approach. Although neither Alexander nor Salingaros dealt with the generative aspects of neural networks, their theoretical framework has received considerable attention since the advent of deep learning and perfectly applies to multi-step image generations. To be more explicit, among Salingaros's rules, the following are particularly relevant to the generation of architectural forms and scenes using deep neural networks: (1) connecting parts of a system, a structure, by various geometric means, often with multiple connections; (2) implementing spatial correlations using distance similarities and scale symmetry; (3) building a system or structure using a sequence of adaptive steps, where organized complexity results from an evolutionary process. Our pipeline illustrates and combines each of these three rules, fostering internal connectivity networks, with no break between scales, but with distinct explicit processing steps for low and high frequencies.

Unsupervised image generation with generative adversarial networks (GANs)

Before the advent of T2I diffusion models in 2017, GANs have been widely used for non-supervised image-to-image (I2I) generation.[14–17] Invented in 2014 by Goodfellow et al.,[15] they have long defined a reference framework for unsupervised, hyperrealistic and expressive I2I generative models. From 2014 to 2019, the scientific literature around GANs and their applications has grown copiously, as instability, convergence and memory issues (particularly for their use on GPUs) have gradually been resolved, and their generative capabilities (quality, diversity, resolution) have increased. GAN-based workflows have provided powerful tools for image generation and editing, enabling a wide range of creative and practical applications, including design, art, data augmentation and image-to-image translation. The field of generative architecture has seen very little investment in GANs, starting with the first experiments by Hasey et al.[2] in 2019,[18] the PhD works on plan generation of Chaillou in 2019[19] and the paper of Nguyen et al. in 2022.[20]

Training a GAN (Figure 5 left) involves a game-theory-type scenario, with competition between a generative model {G} (which captures the distribution of a collection) and a discriminative model {D} (which estimates the probability that a generated image comes from the training data rather than from G). G and D thus have opposing objectives: G tries to deceive D by producing increasingly realistic images, thanks to the corrections made by D's supervision, and D tries to unmask G. The models are trained simultaneously by two multi-layer neural networks.

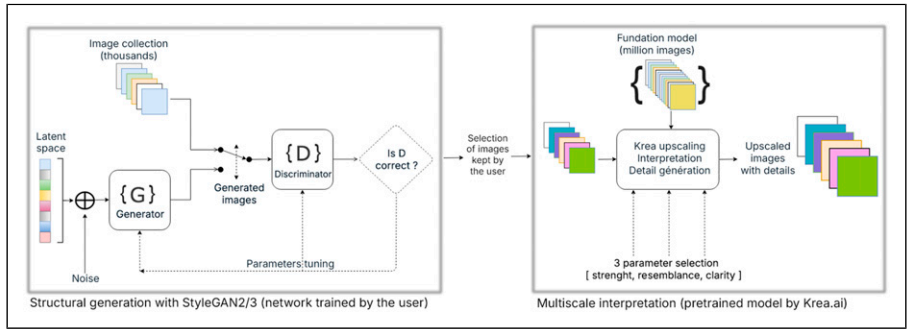

Figure 5. Coupling diagram for the I2I generative models used in this article. The output of the first network is fed as input to the second. The user only trains a StyleGAN2 or StyleGAN3 network, for which he owns the dataset and over which he has full control of the learning process. The second network is a pre-trained foundation model, developed by Krea.ai, over which he has no learning control. Each one deals with a specific range of scales within images.

Borrowed from the style transfer literature, StyleGAN2-ada/3 made a major leap into the GAN world between 2019 and 2021, offering much broader and more intuitive control of image synthesis through modifications or blends of different styles and scales, and a more informative latent space.[14,16] Notably, they provided: (1) a generative architecture G that leads to an unsupervised and automatically learned separation of high-level attributes and stochastic variation in generated images; (2) a normalization of G to encourage better conditioning in the mapping of latent codes to images; (3) an adaptive rate-of-increase mechanism (hyperparameter ada) for the D discriminator, enabling the StyleGAN2-ada code[14] to perform almost as well with only a few thousand images for the training base (a key success factor in architecture, where it is difficult to obtain massive amounts of data).

When trained on limited data, generative models – including GANs – tend to suffer from overfitting and difficulty in converging,[21] which can affect the fidelity and diversity of the solutions generated. Data augmentation[14] and loss regularization[22] were introduced early on to prevent these risks. Moreover, the natural tendency of deep neural networks is to capture low frequencies first in the weight adjustment process, then high frequencies, but for the latter the problem becomes more complicated when dealing with datasets of restricted size,[23] which is the case with a few thousand images.

Traditional machine learning workflows often involve explicit training, validation, and test sets. However, in StyleGAN2-ada and StyleGAN3, the approach is slightly different and can be considered as an implicit handling of validation and testing:

(1) They don't use a separate validation set in the traditional sense. The adaptive augmentation mechanism acts as a form of implicit validation. By monitoring the discriminator's performance and adjusting augmentation accordingly, the model effectively validates its generalization ability during training itself. At last, they cleverly incorporate adaptive data augmentation to mitigate overfitting, which makes a separate validation set less critical.
(2) While there's no separate test set used during training, the trained model is evaluated using established metrics to assess its performance including Frechet/Kernel Inception Distances, or qualitative assessment of generated images by human observers.

Visual prompting for amplification and detail generation

Introduced with the first CNNs, style transfer (the global adaptation of source content to a destination) is one of the most creative concepts in deep learning. It has highlighted the capabilities of its internal representations and given rise to numerous applications. Beyond the transfer of style, the question of an architect's creativity in the ideation phase arises above all in terms of his ability to (re)interpret solutions derived from Generative AI, whether they are fuzzy or complete. An architect knows how to be imaginative in the use of the digital resources at his disposal: he makes use of different resources to come up with something original.

In the I2I field of image upscaling and (re)interpretation, three applications can move in this direction and that seems to be attracting architects: KREA AI Enhancer, Magnific and Leonardo AI Upscaler. They are designed to scale images or enhance low-quality content, such as old photographs or low-resolution renderings, transforming them with little interaction and parameters. While the first two are used predominantly by users, our experiments have shown that KREA is much better suited than Magnific to architects' rendering modes, favouring results that are more natural, smoother and, above all, plausible.

KREA uses an automatic multi-scale analysis of the input image content, which is screened by progressive filters acting differently according to the selection of three model parameters (strength, resemblance, clarity). If desired, users can simplify the setting step by calling up a preset which applies a specific type of

enhancement (standard, flat, strong, reinterpretation, oil painting, digital art). Technically, KREA is based on one of Stable Diffusion's foundation models, coupling *Low-Rank Adaptation* models (LoRA) and *latent coherence models* (LCM), which considerably reduce the number of steps required in the inference process (image generation in a matter of seconds). A real-time version of LCM-LoRA[24] is currently used by KREA.

Proposed method: coupling structural and interpretive models

Motivations

Our ambition is to integrate questions of structure and detail in the same generative chain, without using textual descriptors of the architecture. Generated images unsupervised by text are compositions that benefit from the many degrees of freedom present in I2I models. Less constrained, they can express surprising nuances and reveal the hidden face of a collection, whose multiple levels of coherence are learned in depth.

The coupling of two different and complementary approaches – one structuralist (framework) and the other interpretive (detail addition, super-resolution and finishing) – has enabled us to propose adaptive image production using Generative AI, revisiting the notion of style transfer through organic interpretation, as defined above. The first model generates the structure of the image and its main components, while the second deals with multi-scale details, but with global coherence (very important). The first one is based on training a GAN with a small dataset (4300 images). The second model has been pre-trained on hundreds of millions of images, making it possible to add plausible detail to blurred or uncertain areas, and above all to reinterpret the content of GAN productions subtly.

But in the context of limited data, why use a GAN instead of a LoRA under Stable Diffusion, and why StyleGAN ? We have capitalized on the fact that a pre-trained StyleGAN2-ada/3 model is a considerable resource for local structure, even if it seems much more limited in terms of generality than recent T2I diffusion models. Researchers have observed that in the presence of limited data, the generative richness of a GAN often exceeds that of T2I diffusion models.[3] Above all, the unfinished state of a GAN model trained on a reduced dataset produces images with high structural diversity, but low-frequency content (average FID between 13 and 20), making them still capable of enrichment. Finally, StyleGAN has less of a tendency than other GANs to replicate dataset images. On the contrary, its diversity and compositional richness make it a good candidate for ideation explorations.

Bringing enhanced content from foundation models

The low frequencies generated by StyleGAN2-ada/3 correspond to the morphological and graphic structures learned, while interpretation and enhancements are brought about by a broader model such as KREA (Figure 5 right), which provides the high frequencies specific to scales of detail. Nevertheless, the separation between the joint action of the two models (coarse to fine) is not so clear-cut, since enhancement is not just a local interpretation problem, but also an upscaling one, and depends on several parameters (see next section).

While StyleGAN enables stimulating generation at the structural scales that provide the essential morphology of images and their style, KREA's upscaling engine through content analysis and reinterpretation demonstrates an inventive capacity at fine scales, coordinated at global levels, capable of provoking and stimulating architects where they least expect it. A form of serendipity can thus manifest itself at all compositional scales, from structure to style. This is the strength of deep learning, whose two phases proposed in this paper, based on two truly distinct Generative AI processes in the tradition of Alexander's and Salingaros's theories, give pride of place to the neuro-mimetic imagination and achieve a compromise capable of reconciling structuralists and artists.

Experiments and results

Settings: Parameters and scales

Results shown in this article and above all on our drive (https://www.flickr.com/photos/201879342@N08/albums/) are derived from the coupling of StyleGAN2-ada/3 and KREA, for which we provide in Table 1 and the following command lines all the parameters required for reproduction. GANs being limited by quadratic training times with image resolution, we opted early on to process a 512 × 512-pixel dataset.

Table 1. We summarize here the various settings used in our experiments. StyleGAN's most important hyperparameters, carefully fine-tuned for this type of dataset[3], are regularization (gamma) and augmentation rate (hard-coded in the file training_loop.py, like generator/discriminator learning rates G_lr and D_lr). Dozens of experiments were performed to achieve FID[4] of 13.6 after 1742 training epochs (Figure 6).

		StyleGAN parameters and FID							KREA parameters			
Model	Dataset size	Configuration	G_lr	D_lr	Gamma	Aug_rate	Batch	FID	Best preset	Strength	Resemblance	Clarity
StyleGAN2_ada	4300	stylegan2 mirror	0.002	0.002	15.5	5500	8	13.62	flat	[0.9 ; 1.0]	[0.65 ; 0.7]	[0.07 ; 0.12]
StyleGAN3-t	4300	stylegan3t mirror	0.0025	0.002	35	10000	8	13.67	strong	[0.78 ; 0.82]	[0.1 ; 0.2]	[0.6 ; 0.75]

The command lines used for training styleGAN models are:

```
python train.py --outdir=training-runs --cfg=stylegan2 --data=../datasets/pers_4300 --gpus=1 --batch=8 --gamma=15.5 --mirror=1 --snap=16 --metrics=fid50k_full

python train.py --outdir=training-runs --cfg=stylegan3-t --data=../datasets/pers_4300 --gpus=1 --batch=8 --gamma=35 --mirror=1 --snap=16 --metrics=fid50k_full
```

KREA implements an upscaling factor ranging from 2x to 8x. In our experiments, we have only tested 2x and 4x outputs, up to 2048 × 2048 pixels, but our main results are shown in 1024 × 1024 pixels. Cascading 2X upscaling has also been tested to improve enhancements, and it appears to give better results than direct 4x upscaling. For convincing architectural results, we have found tight range settings for the three KREA parameters (Table 1) of its *legacy model*: *strength* [0.9; 1.0], *resemblance* [0.65; 0.7] and *clarity* [0.07; 0.12] are well suited for StyleGAN2, whereas *strength* [0.78; 0.82], *resemblance* [0.1; 0.2] and *clarity* [0.6; 0.75] are more appropriate for StyleGAN3. Please refer to KREA presets for other value behaviour and resulting styles.

Finally, we wondered at what FID threshold the coupling can be used with meaningful results. The most common metric for assessing the performance of generative models is undoubtedly the FID.[17] It quantifies the difference between the synthesized distribution and the observed distribution. A low FID indicates that these distributions share similar visual statistics. Figure 7 clearly shows that below the FID = 30 threshold (i.e. around 220 StyleGAN calculation epochs), KREA begins to produce credible results. Nevertheless, lower FID values (below 15) are preferable for better ones (Figure 7).

Collections

Having left behind the formatted framework of T2I, the very structure of the objects and the graphic and interpretive content of the resulting images are likely to surprise architects. Deep learning composes something new by linking and cross-referencing multi-scale elements learned in depth from a collection of images. Volumes, shapes, landscapes, decorative elements, styles and rendering modes are combined and reused in a multitude of ways (Figures 1–4). But each model or tool has its signature. A different representation of architecture emerges, with the enrichment of details in blurred areas, coherent transformation in

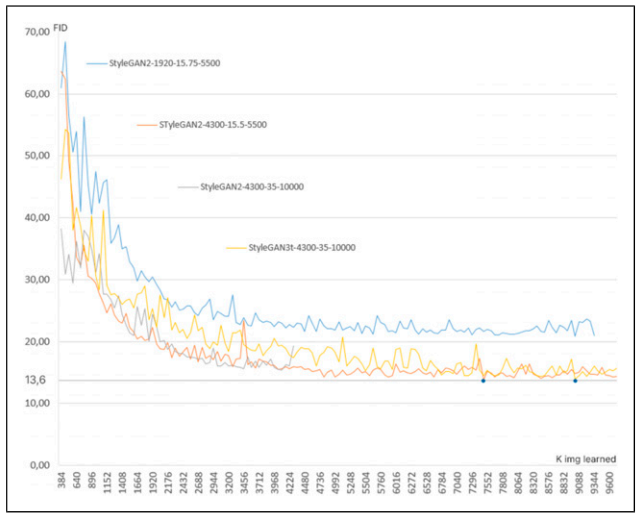

Figure 6. Graph showing the evolution of the Frechet Inception Distance (FID) measure as a function of the evolution of our dataset size (1920 to 4300) and the number of learning epochs (~K-img) during the training step. This enabled us to select the best model (Table 1) for each version of StyleGAN with its main hyperparameters (augmentation rate, regularization = gamma).

context, and a source of surprise and poetry. This can be seen online on the drive, divided into 17 collections taken from the generated catalogue: alignment, watercolour, bold, cosy, domus, legend, neo-classical, neon, oblique, gable, square, point, pyramid, point roof, tower, rounded tower, vegetated. These collections demonstrate the diversity and generative richness achieved with our two-phase process. They also reveal many points of interest inside the GAN's latent space, grouping clusters with structural and morphological identity and a wealth of variation. They were previewed at an exhibition in the School of Architecture of Lyon in France, in autumn 2024.

Beauty and quality

The results presented here reveal an immense and little-known potential for the architectural community, which has been gradually investing in GenAI tools since 2022. The representations generated are plausible, visually pleasing, sometimes disturbing, and often stimulating. They exhibit organic and vegetal characteristics (according to, of course, to the dataset design) that meet the expectations of architects in search of biophilia in an emerging post-industrial world. We strongly advise the reader to spend some time on our repository and take an in-depth, comparative look at the results grouped into 17 classes. Although limited to 1K resolution, these images are of sufficient quality to demonstrate the generative power and originality of the method described in this paper.

It's worth mentioning that images produced by a GAN often contain imperfections, strange little details called blobs, a kind of visual hallucination well-known in the literature. Their reinterpretation based on local coherence sometimes leads to significantly integrated artifacts, since context is always taken into account, as it is for all compositional aspects of the image.

Architects and teachers were invited to assess the results of the finishing stage on several hundred validation models, thus contributing to the final choice of the range of three significant parameters in the Krea model (Table 1). In addition, they observed that the results produced by StyleGAN3, while providing better

Figure 7. Starting from the same random seed, this series shows the progression of StyleGAN2-ada quality and KREA interpretation, as a function of the number of training epochs (epk) and the associated FID. Clearly, below the FID=30 threshold, results improve markedly and become credible.

looks with preset « strong », appear less diversified than those of StyleGAN2, even though its training consumes 50% more time.

Hardware requirement, training and power consumption

For training StyleGAN2/3, a Nvidia RTX card is necessary, equipped with at least 16 GB of RAM. Without aiming for the still-expensive 6000 series, the RTX 5000 series offers sufficient computing power for a mid-range price. StyleGAN2-ada takes around a week to train a thousand epochs of our dataset on the RTX 5000-ada GPU, generating a power consumption of 30 kWh. StyleGAN3, on the other hand, requires 50% more energy. This cost can be considered very reasonable for an architect or an agency, compared to the cost of training a broadcast model on millions of images. KREA inference just takes a few seconds per image. The amount of energy resulting from all the models learning calculations implemented in this research (with an expected sober approach) was roughly 1200 kWh over 3 years.

Conclusion and discussion

This research is a further step towards understanding how AI-guided generative approaches can support architectural design processes while balancing structural and interpretive transformations. « *The key is to consider the potential offered by AI not as an end in itself, but as an aid within a process that needs to be rethought* »[5]. In this vein, we have proposed an algorithmic approach and process exploiting a little neglected facet of Generative AI, adapted to users: (1) wishing to observe innovative architectural morphologies well integrated into a plausible built and landscape environment; (2) having time to build datasets that make sense to them.

Benefits

The generative capacity contained within a relatively limited number of images is a significant asset in the project ideation phase, where draughts and sketches are often the starting point, but also during the design phase, when a significant amount of architectural reference material is created. We have also highlighted its growing interest in the visual arts. We've just seen that a *state-of-the-art* GAN such as StyleGAN2-ada/3 generates architectural object structures that are still rich in interpretive potential, making them ideal for ideation and sketching. Another advantage lies in the ease of local editing within latent spaces. Indeed, StyleGAN2-ada/3 produces local variations that can be manipulated intuitively, something that is far more complicated to achieve in T2I mode. We also observed the multi-scale nature of interpretive composition, which can surprise us by generating fine details linked to contexts (the example of the gable balcony in Figure 4 is significant in this respect).

Mixed reviews

Some of the architects who have observed our results confirm that the interpretive power of the AI is greatest at the end of phase 1, which is still unfinished, while phase 2, although visually stimulating and capable of proposing highly original representations, is much less of a project. This can be linked to the possibility of a design fixation effect, that is the limitation of the designer's ability to deviate from the AI's proposal due to the photorealistic nature of the images generated.[8] What architects are now realizing is that Generative AI, by calculating too fast, easily short-circuits their rather slow processes of interpretation and deliberation, the basis of human creativity, even though it proposes many solutions with variants. Yet these processes are open to something fundamental for human beings: meaning and effort to reach the goal, the essential difference between production and art.

Finally, we note that this type of multi-scale detailing approach is quite likely to lead to the production of non-standardized, rather lush architecture, the cost of which does not fall within the scope of sustainable development. This rich aesthetic would therefore be reserved for exceptional architectural projects. But it will also delight visual artists.

Future work

Comparison with other generative models

Current Generative AI is mainly based on pre-trained foundation models that have captured the internal distribution of hundreds of millions of objects. These heavy models are then declined or adapted into lighter versions, either by knowledge distillation or by adaptation to local data of limited size, using fine-tuning techniques: *Retrieval Augmented Generation* (RAG, essentially for LLM), *Low Rank Adaptation* (LoRA) and *Few Shot Generation* (FSG), the latter becoming increasingly important when collecting large-scale training samples is not feasible.[21] LoRA is a *Parameter-Efficient Fine Tuning* learning technique increasingly used in the professional world in the context of style transfer from a collection. It involves fine-tuning the parameters of a pre-trained SD model using a set of data specific to a small target domain, to better align with specific needs.[25] The models produced are often shared free of charge online, but architects can also build models from their productions or references.

Therefore, three possible avenues will be explored. (1) A tempting idea is to replace StyleGAN with a diffusion model relying only on a few visual inputs, like the prompt-free diffusion-based approach,[26] known to outperform prior exemplar-based generators. (2) Test other sketching/reinterpretation/editing tools, where each phase can be guided by a different tool. (3) Finally, train one or more specific LoRAs (from the 17 selected collections) in Stable Diffusion, and analyse the differences in rendering with our method.

Improving models

Our first objective is to increase the size of the dataset to reach the 10K-image threshold, and to continue testing the generative enrichment this brings (diversity, lower FID). In addition, it is now common practice to enrich datasets based on AI-generated solutions. This is already the case, for example, when training a model on synthetic data. Here, the pipeline could consist of a cascade (StyleGAN + KREA) → (StyleGAN + KREA), justified by the novelty of the crossovers obtained on different data learned differently.

Latent spaces

Perhaps the most exciting aspect of GANs is the way their latent space is organized. It has been shown that GANs tend to organize their latent space smoothly, that is in such a way that nearby regions in the latent space represent similar data.[27] A pre-trained StyleGAN model is therefore a considerable resource for local structure and disentanglement. Moreover, it is possible to inject new concepts into such a model, without disturbing the structure of the information already learned or even increasing the size of the model. To demonstrate this, Nitzan et al. developed a domain expansion method based on the capabilities of StyleGAN.[28] They found that the generator contains significant pre-trained latent space that offers dormant, unused directions that do not affect the output and can easily be reallocated to learn several new domains without constraint.

Finally, exploring the 512-parameter latent spaces of StyleGAN2-ada/3 reveals a multi-scale structuralist approach, capable of producing a wealth of representations exceeding – in intuitive generation – the limits of the parametric paradigm ubiquitous among architects. The in-depth study of these spaces, admittedly carried out on larger datasets (10K to 100K), could bring to light unprecedented archetypal structures of interest to the

field of architectural analysis, and consequently its generation. This would add an important analytical dimension to our initially purely generative work.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest concerning the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Xavier Marsault https://orcid.org/0009-0005-4009-7024

Data Availability Statement

The authors share their research data in relevant public repositories.

Notes

1. First appeared with conditional GAN,[29] image-to-image (I2I) is the task of transforming an input image through a variety of possible manipulations and enhancements, such as style transfer, upscaling, inpainting, outpainting, colourization, editing and more.
2. Michael Hasey has worked at the intersection of GANs and the built environment. His work with engineer Elliot, Gan-Itecture, included experiments with GANs that have shown potential for producing new images of architecture, notably in the style of Zaha Hadid, despite initial results of poor quality.
3. For a discussion of the widespread use of these settings, see https://github.com/NVlabs/stylegan3/ blob/main/docs/configs.md.
4. Frechet Inception Distance.[17]
5. Laurent Lescop, professor at ENSA Nantes, School of Architecture, France.

References

1. Negrotti M. *Alternative intelligence. Understanding the artificial: on the future shape of artificial intelligence*. London: Springer-Verlag, 1991, vol 55–75.
2. Still A and D'Inverno M. A history of creativity for future AI research. *Proceedings of the seventh international conference on computational creativity*, 2016.
3. Gal R, Alaluf Y, Atzmon Y, et al. An image is worth one word: personalizing text-to-image generation using textual inversion. *ICLR* 2023.
4. Rombach R, Blattmann A, Lorenz D, et al. Highresolution image synthesis with latent diffusion models. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 10684–10695.
5. Navard P, Karimi Monsefi A, Zhou M, et al. *KnobGen: controlling the sophistication of artwork in sketch-based diffusion models*, 2024, arXiv:2410.01595v1.
6. Le Cun Y and Fridman L. *Podcast #416*, 2024. https://www.youtube.com/watch?v=5t1vTLU7s40
7. Gielis J. A generic geometric transformation that unifies a wide range of natural and abstract shapes. *Am J Bot* 2003; 90(3): 333–338.
8. Dortheimer J, Schubert G, Dalach A, et al. Think AI-side the box! Exploring the usability of text-to-image generators for architecture. *ECAADe 2023: Digital Design Reconsidered*, 2023, pp. 567–576.

9. Le Cun Y. *L'apprentissage profond: une révolution en intelligence artificielle*. Conférence au Collège de France, 2016.
10. Alexander C. *The nature of order, book 1: the phenomenon of life, 2001; book 2: the process of creating life, 2002; book 3: a vision of a living world, 2005*. Berkeley CA: Center for Environmental Structure, 2005.
11. Salingaros NA and Masden K. *The science of intelligent architecture*, Teka komisji urbanistyki i architektury pan oddział w Krakowie 2017.
12. Salingaros NA. Adaptive vs. Random complexity, parts 1 and 2. *ArchNewsNow*, Chapter4 A & B. *Architecture's New Scientific Foundations*, 2015.
13. Marsault X. *Inspired generative architecture. Habilitation to supervise research*. INSA: University of Lyon, 2023.
14. Karras T, Aittala M, Hellsten J, et al. Training generative adversarial networks with limited data. *Adv Neural Inf Process Syst* 2020; 12: 104–114.
15. Goodfellow I, Jean P-A, Mehdi M, et al. Generative adversarial networks. *Adv Neural Inf Process Syst* 2014; 27.
16. Karras T, Aittala M, Laine S, et al. *Alias-free generative adversarial networks*, 2021, arXiv:2106.12423v1.
17. Heusel M, Ramsauer H, Unterthiner T et al. GANs trained by a two time-scale update rule converge to a local nash equilibrium. 2018.
18. Hasey M, Elliott S and Rhee J. Archi-base: automated dataset construction of architectural imagery for deep neural networks. Proceedings of CAADFutures' 2021.
19. Chaillou S *AI + architecture - towards a new approach*. 2019, thesis of Harvard University.
20. Marsault X and Nguyen HMC. Les GANs : stimulateurs de créativité en phase d'idéation. *SHS Web Conf* 2022; 147: 06003. DOI: 10.1051/shsconf/202214706003.
21. Yang M and Wang Z. *Image synthesis under limited data: a survey and taxonomy.* 2023. arXiv:2307.16879v1, 31.
22. Li Z, Usman M, Tao R, et al. A systematic survey of regularization and normalization in GANs. *ACM Comput Surv* 2023; 55(1): 1–37.
23. Xu Z-QJ, Zhang Y, Luo T, et al. *Frequency principle: fourier analysis sheds light on deep neural networks* 2019, arXiv:1901.06523.
24. Luo S, Tan Y, Patil S, Gu D, Von Platen P, Huang L, Li J and Zhao H. *LCM-LORA: a universal stable-diffusion acceleration module. Nov 2023*, arXiv:2311.05556v1. https://latent-consistency-models.github.io/
25. Moon T, Choi M, Lee G, et al. Fine-tuning diffusion models with limited data. *Adv. Neural Inform.Process. Syst. Workshop*, 2022.
26. Xu X, Guo J, Wang Z, et al. *Prompt-free diffusion: taking text out of text-to-image diffusion models*. 2023, arXiv: 2305.16223v2.
27. Bermano AH, Gal R, Alaluf Y, et al. *State-of-the-Art in the architecture, methods and applications of StyleGAN*. 2022, arXiv:2202.14020v1.
28. Nitzan Y, Gharbi M, Zhang R, et al. Domain expansion of image generators. 2023.
29. Isola P, Zhu JY, Tinghui Zhou T, et al. Conditional image-to-image translation with conditional adversarial networks. *Proc. CVPR* 2017.