AI IN ARCHITECTURE: ARE ARTIFICIAL INTELLIGENCE IMAGE GENERATORS USEFUL IN ARCHITECTURE?

DOI: https://doi.org/10.18485/arh_pt.2025.10.ch63

_ Jelena Stevanović

Teaching Associate and PhD student, Faculty of Architecture – University of Belgrade, Belgrade, Serbia, jelena.stevanovic@arh.bg.ac.rs, [0009-0004-9518-0485]

_ Nataša Bajc

Associate Professor, Faculty of Architecture - School of Visual Arts - Pasadena City College, Pasadena, CA, USA, natasha@natashabajc.com

_ Nataša Ćuković Ignjatović

PhD, Associate Professor, Faculty of Architecture – University of Belgrade, Belgrade, Serbia, natasa@arh.bg.ac.rs, [0000-0003-4934-3269]

ABSTRACT

Artificial intelligence is applied in architecture to enhance innovation in the building design process and minimize manual repetitive tasks. A wide variety of artificial intelligence models exist and they have different implementations. They make the design process more cost effective and time efficient while reducing resource requirements.

This research paper provides an overview of artificial intelligence models used for image generation and evaluates their utility in the domain of architecture. It examines potentials, limitations and overall impact of these models on this field, aiming to determine whether they can provide valuable contributions in advancing architectural building design processes and practices. It should be taken into consideration that specialized software developed for architects, such as AutoCAD and Rhinoceros, relies on vectors that are essential for constructing different geometric shapes and their components. Furthermore, there still exists the challenge to provide precise image segmentation and extraction of vectors from image components.

The paper introduces artificial intelligence in the context of image generation and explores selected generative modelling frameworks (Generative Adversarial Networks, Diffusion Models and Variational Autoencoders), analysing and highlighting their mutual similarities and differences. These distinctions provide a thorough understanding of their impact within the domain of architecture and assess whether these models for image generation can bring benefits to architects. It is shown that the provided new images have great potential to be used afterwards in the domain of architecture. They really contribute to the field, but its application requires a large set of input data, rasterization and vectorization preprocess and postprocess.

KEYWORDS _ *Artificial intelligence, Generative Adversarial Networks, Diffusion Models, Variational Autoencoders, architectural building design process*

INTRODUCTION

Artificial intelligence (AI) is implemented in many disciplines with the aim to automate processes and reduce time-consuming repetitive tasks and errors. Today, AI generative models are types of models widely used in the field of architecture. This research paper explores them and explains whether they can contribute to this field. The structure of the paper is as following. The first paragraph is a quick introduction to some related work. The second paragraph gives an overview of generative modelling frameworks, such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs) and Diffusion Models (DMs). The third paragraph provides their comparison. The fourth paragraph shows experiment results that represent implementation of Interpretable representation learning by information maximizing generative adversarial network (InfoGAN), based on apartment floorplan input dataset. The fifth paragraph is a discussion about architecture demands, positive and negative aspects of generative modelling frameworks and their potential impact in the domain of architecture. The sixth paragraph provides potential future development.

RELATED WORK

There are many architectural research projects dealing with automation of segmentation or classification processes. Kalervo et al. (2019) proposed a pipeline that implements convolutional neural network (CNN) for segmenting scanned drawings of apartment floorplans from Finland. CNNs are deep learning networks used for recognizing objects within image. Semantic segmentation is used for classifying pixels in image. The pipeline requires manual preparation of input data which consists of scanning and annotating floorplans. Annotation is done on vector based SVG format, ensuring the image keeps spatial relationships of all elements. Afterwards, they are split in training, validation and test input datasets. This research explains pipeline for transforming rasterized to vectorized data, resulting in vectorized representation of walls, windows and doors. Barreiro et al. (2023) in paper extended this research by implementing previous CNN segmentation and vectorization process to reconstruct 3D vectorized model from 2D scanned floorplans. This research contributes to the automation of digitalizing buildings constructed before existence of architecture and construction design software. Lu at al. (2020) introduced StructGAN, a specialized type of GAN for optimizing building design. StructGAN model makes buildings seismically resilient giving out different proposals where shear walls should be constructed. The input data also requires to be pre-processed by marking structural elements and augmenting through flipping and rotating floorplans. These studies have investigated correlation between AI segmentation and vectorization processes and architecture. This paper evaluates AI generative modelling frameworks working with vectorized data in domain of architecture.

ARTIFICIAL INTELLIGENCE IN CONTEXT OF IMAGE GENERATION

The usage of AI generative modelling frameworks in architecture can open new building design possibilities. They are able to create highly realistic images depending on previously prepared input dataset. The same as CNNs, these models are based on multiple-layer neural networks that encode and decode image information. They are unsupervised, which means that input dataset used for training is not labelled (O'Shea et al. 2015).

Generative Adversarial Networks

Generative Adversarial Networks are an unsupervised deep learning frameworks that generate photorealistic images. They are firstly introduced in 2014 by Ian Goodfellow and colleagues in a paper "Generative Adversarial Nets". Many GAN models are developed and have different implementations.

Some GANs can transform image into specific style or replace different objects. The structure of GANs is shown in Figure 1. GANs belong to adversarial models and consist of generator and discriminator. They are both trained and involved in Minimax game and try to outperform one another. Generator distributes new image while discriminator tries to determine whether that image came from training dataset or made by generator. Generator tries to make image more realistic and to fool discriminator. Even though GANs made great progress in computer vision, they require large input training dataset and huge amount of time used for training networks (Goodfellow et al. 2014).

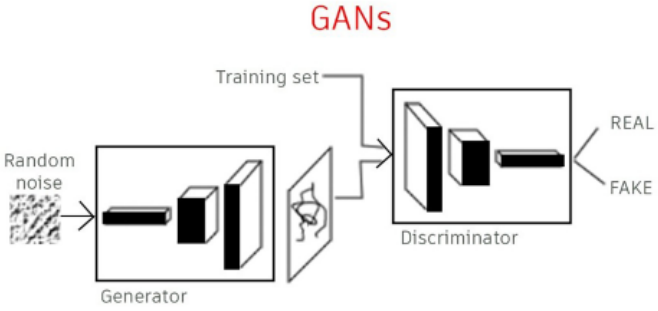

Figure 1: Generative adversarial network structure. Adapted from https://sthalles.github.io/intro-to-gans/

Variational Autoencoders

Variational Autoencoders are unsupervised generative models that generate image almost identical to the input image. VAEs are firstly introduced in 2013 by Diederik P. Kingma and Max Welling in a paper "Auto-Encoding Variational Bayes". The structure of VAEs is shown in Figure 2. They belong to probabilistic methods and consist of encoder and decoder. Encoder is a network consisting of layers and extracts latent variables from input image compressing it to lower-dimensional space called bottleneck. It represents compressed input data in the form of vector. It carries information about different attributes that define input image. Afterwards, latent space vector is used as an input for decoder. Latent space consists of mean and standard deviation. Mean represents the mean of values in the input dataset and standard deviation represents a square root of the variance that represent measure of how much the values vary from the mean value (Bergmann and Stryker 2024). Layers of encoder networks contain less nodes as they are closer to the bottleneck. Decoder reconstructs the input data and its network consists of layers where each following layer has more nodes. VAEs can be used for denoising images (Kingma and Welling 2013).

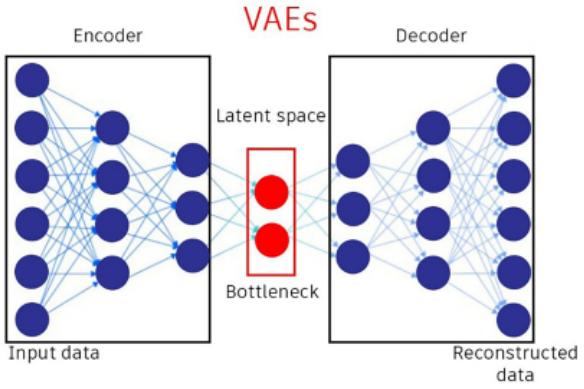

Figure 2: Variational autoencoders structure. Adapted from https://www.ibm.com/think/topics/variational-autoencoder

Diffusion Models

Diffusion models are generative models firstly introduced by Jascha Sohl-Dickstein and colleagues in 2015 paper "Deep Unsupervised Learning using Nonequilibrium Thermodynamics". DMs belong to the probabilistic models and consist of forward trajectory and reversed trajectory in which they add Gaussian noise to the input image and learn to remove the noise from the image respectively (Figure 3). These processes are done in diffusion steps and use Markov chains where current state depends on only a finite fixed number of previous states. As a result, DMs give high-quality images, but involves very slow sampling processes because of huge number of steps needed for image generation (Sohl-Dickstein et al. 2015).

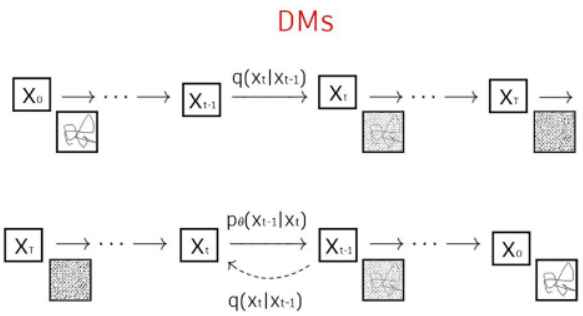

Figure 3: Diffusion models' forward and reversed trajectory. Adapted from https://www.superannotate.com/blog/diffusion-models

COMPARISON OF MODELS

AI generative modelling frameworks (GANs, VAEs and DMs) consist of generative neural networks. They differ in many aspects, such as learning type, mathematical foundation, loss function, training speed and quality of output data. GANs, VAEs and DMs are unsupervised models, but they can also be semi-supervised. That means that labelled data can be incorporated into training process. All three models use Gaussian distribution. DMs add Gaussian noise in forward diffusion process. GANs and VAEs use Gaussian noise for sampling latent space. Right dimension of latent space determines the efficiency of model. GANs contain generator and discriminator consisting of feedforward neural networks. VAEs consist of feedforward encoder and decoder. DMs work somewhat differently implementing forward trajectory for diffusion and reversed trajectory for denoising. GANs are adversarial networks and VAEs and DMs represent probabilistic neural networks. GANs and DMs result in high-quality output images. VAEs give out blurrier images, and that is why they have not gained the same attention as GANs or DMs. GANs are slow, because they rely on Minimax game. Generator and discriminator constantly try to outperform each other. VAEs use Variational Bayesian Inference and DMs use Markov chains. DMs are the slowest because they have huge number of diffusion and denoising steps. The fastest are VAEs models. All three models result in very diverse output and require a huge input set of image data.

Table 1: Comparison of GANs, VAEs and DMs

	GANs	**VAEs**	**DMs**
Learning type	adversarial	probabilistic	probabilistic
Output image quality	high-quality	blurrier	high-quality
Training speed	fast	fast	slow
Mathematical foundation	Minimax game	Variational Bayesian Inference	Markov chains
Output diversity	high	high	high

EXPERIMENTS

AI application for the automatic visualization of the objects elements in image is illustrated by Stevanovic (2022) as follows. InfoGAN is a type of GAN modelling framework that is able to disentangle different visual concepts. These concepts can recognize and generate various forms of lighting, pose, orientation, width or height of the object in the image. InfoGAN requires huge input dataset. This experiment evaluates InfoGAN implementation using apartment floorplans as input data. Collected floorplans were rotated 3 times for 90 degrees in order to increase input dataset, which resulted in approximately 5000 input images. The results from InfoGAN implementation are shown in Figure 4. Output image quality is not high due to still not a large enough input dataset. InfoGAN recognized and generated the position of walls and furniture. Apartment floorplans are changing their form by removing rooms from one side and adding them to another. That resulted in their transformation from horizontal to vertical orientation.

An additional AI application was developed by Deep Convolutional Generative Adversarial Network (DCGAN), a more stable type of GAN. Differently from InfoGAN, input dataset consists of grayscale images. The generated images are in the same tone. Some generated examples include floor hatches and furniture, because they are present in part of training dataset. DCGAN recognises different wall thickness. The input dataset is not large, but the results are more clear than those generated by InfoGAN.

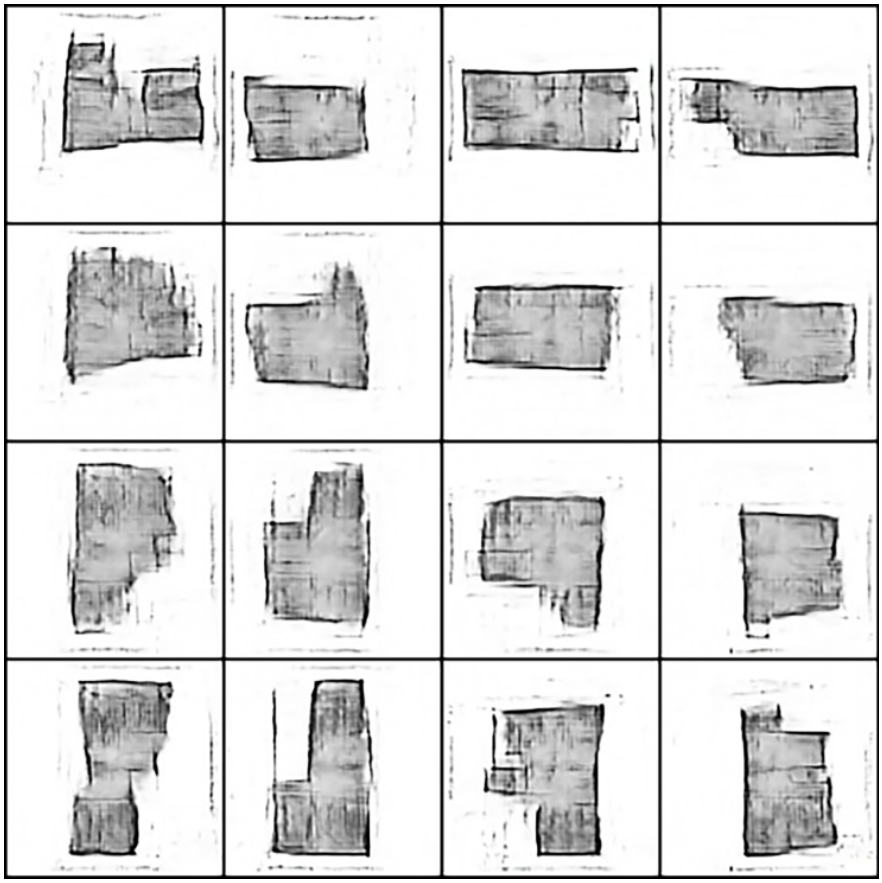

Figure 4: Result of InfoGAN implementation using different floorplans as input data (Stevanovic, 2022).



Figure 5: Result of DCGAN implementation using different floorplans as input data (Stevanovic, 2025).

DISCUSSION

Implementation of AI generative modelling frameworks in architectural design can give results in various inspirational proposals and help architects experiment in initial design stages that include concept design. Mostly used models are text-to-image diffusion models. They generate images based on the input textual description prompts of image content and artistic style. These are multimodal models, because they work with more input and output datatypes. At the beginning, model attempts to generate image that closely matches the description. Afterwards, the model focuses on improving the quality of the image (Zhang et al. 2023). Text-to-image models use natural language processing and try to understand the input data semantic. They manage to understand the meaning of the text and connect the words to the visual context. The text and image embeddings are made and concatenated (Gupta 2023). There is a great progress in DMs development. Some of them allow input text descriptions that will not be generated on the image.

Vector-based and latent-based environment

Architectural design software, such as AutoCAD, Revit or Rhinoceros, work in vector-based environment. They require spatial relationships between all elements and information about elements dimensions, orientation etc. The architectural project representation or documentation preparation

includes creating rasterized floorplan or section images after usage of architectural design software. The reversed process is still not completely automated. Automatization of floorplan generation using AI models requires work with vectorized data. On the other hand, state-of-the-art AI generative modelling frameworks are based on latent space that encodes pixelated data (Pham Van et al. 2021). They require additional preprocesses or postprocesses for preparing input or output data. Besides from rotation and mirroring operations for enlarging input dataset, preprocessing usually consists of labelling, segmentation or classification of floorplan elements. Postprocessing requires translation from rasterised to vectorised representation of results. Apart from manual work, these additional tasks could be performed with appropriate computer code. Forasmuch as they need huge input dataset for training, preparation is extremely time-consuming. The InfoGAN experiment from previous paragraph shows that generative modelling framework can recognize floorplans elements and give out various visual concept results. It is able to distinguish the space inside and outside the floorplan boundary and determine where furniture should be placed based on input image pixels. However, further application of these results in architecture design software as vectorized input data needs additional processing. It is essential to provide vector embeddings that AI generative models will learn to evaluate and generate. Then, preprocess and postprocess will not include rasterization or vectorization and will be much less time-consuming and error prone.

CONCLUSION AND FUTURE WORK

This paper reviews different types of AI generative modelling frameworks used for generating rasterized images and provides their potentials and restrictions in the domain of architectural design process. Also, the results from InfoGAN and DCGAN implementations based on apartment floorplan input datasets are introduced. It is emphasized that AI generative modelling frameworks do not work directly with vectorized data and still require rasterizing in preprocessing or vectorizing in postprocessing data. Faster development of floorplan could be achieved by connecting pixels and vectors, which diminishes manual repetitive work. Such an approach enables architects to easily test new floorplan outcome designs. In addition, architects could devote more time to different design processes, where vectorized results would instantly be available for further use. Future research could explore potential modifications or development of new generative modelling frameworks, in order to working directly with vector embeddings.

REFERENCES

- Barreiro, Aleixo Cambeiro, Mariusz Trzeciakiewicz, Anna Hilsmann, Fraunhofer HHI, Peter Eisert, and Humboldt University of Berlin. 2023. “Automatic Reconstruction of Semantic 3D Models from 2D Floor Plans.” *Proceedings of the 18th International Conference on Machine Vision and Applications (MVA)*, 1–5. IEEE, 2023.
- Bergmann, Dave and Cole Stryker. 2024. “*What Is a Variational Autoencoder?*” *IBM blog,* June 12. Accessed March 29, 2025. https://www.ibm.com/think/topics/variational-autoencoder.
- Goodfellow, Ian J., Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. “Generative Adversarial Networks.” *arXiv (Cornell University)*, January. https://doi.org/10.48550/arxiv.1406.2661.
- Gupta, Mehul. 2023. “How Does DALL-E, the Text-to-image Generator Work? - Data Science in Your Pocket - Medium.” *Medium blog*, February 2. Accessed March 29, 2025. https://medium.com/data-science-in-your-pocket/how-does-dall-e-the-text-to-image-generator-work-c2d9f4a0f26c.
- Kalervo, Ahti, Juha Ylioinas, Markus Häikiö, Antti Karhu, and Juho Kannala. 2019. “CubiCasa5K: A Dataset and an Improved Multi-Task Model for Floorplan Image Analysis.” *arXiv (Cornell University)*, January. https://doi.org/10.48550/arxiv.1904.01920.
- Kingma, Diederik P, and Max Welling. 2013. “Auto-Encoding Variational Bayes.” *arXiv (Cornell University)*, January. https://doi.org/10.48550/arxiv.1312.6114.
- Lu, Xinzheng, Wenjie Liao, Yuli Huang, Zhe Zheng, and Yuanqing Lin. 2020. “Automated Structural Design of Shear Wall Residential Buildings Using GAN-Based Machine Learning Algorithm.” *Research Square (Research*

Square), October. https://doi.org/10.21203/rs.3.rs-84863/v1.

- O'Shea, Keiron, and Ryan Nash. 2015. "An Introduction to Convolutional Neural Networks." *arXiv (Cornell University)*, January. https://doi.org/10.48550/arxiv.1511.08458.
- Pham Van, Toan, Tam Minh Nguyen, Ngoc N. Tran, Hoai Viet Nguyen, Linh Bao Doan, Huy Quang Dao, Thanh Ta Minh, Sun* Inc, and Le Quy Don Technical University. 2021. "Interpreting the Latent Space of Generative Adversarial Networks using Supervised Learning." Journal-article. *arXiv:2102.12139v1 [Cs.LG] 24 Feb 2021*. http://dx.doi.org/10.1109/ACOMP50827.2020.00015.
- Sohl-Dickstein, Jascha, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. 2015. "Deep Unsupervised Learning Using Nonequilibrium Thermodynamics." *arXiv (Cornell University)*, January. https://doi.org/10.48550/arxiv.1503.03585.
- Stevanovic, Jelena. 2022. " Generation of residential unit area with application of machine learning" MSc Work, University of Belgrade, Faculty of Electrical Engineering, Belgrade. (in Serbian)
- Zhang, Chenshuang, Chaoning Zhang, Mengchun Zhang, and In So Kweon. 2023. "Text-to-image Diffusion Models in Generative AI: A Survey." *arXiv (Cornell University)*, January. https://doi.org/10.48550/arxiv.2303.07909.