SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ DERGİSİ

Sakarya University Journal of Science
SAUJS

ISSN 1301-4048 | e-ISSN 2147-835X | Period Bimonthly | Founded: 1997 | Publisher Sakarya University | http://www.saujs.sakarya.edu.tr/

Title: Transforming Sketches into Realistic Images: Leveraging Machine Learning and Image Processing for Enhanced Architectural Visualization

Authors: İlker KARADAĞ

Recieved: 23.06.2023

Accepted: 7.08.2023

Article Type: Research Article

Volume: 27
Issue: 6
Month: December
Year: 2023
Pages: 1209-1216

How to cite
İlker KARADAĞ; (2023), Transforming Sketches into Realistic Images: Leveraging Machine Learning and Image Processing for Enhanced Architectural Visualization. Sakarya University Journal of Science, 27(6), 1209-1216, DOI: 10.16984/saufenbilder.1319166
Access link
https://dergipark.org.tr/en/pub/saufenbilder/issue/80994/1319166

New submission to SAUJS
http://dergipark.gov.tr/journal/1115/submission/start

Sakarya University Journal of Science 27(6), 1209-1216, 2023

Transforming Sketches into Realistic Images: Leveraging Machine Learning and Image Processing for Enhanced Architectural Visualization

İlker KARADAĞ [*1]

Absftract

This article presents a novel approach for transforming architectural sketches into realistic images through the utilization of machine learning and image processing techniques. The proposed method leverages the Stable Diffusion model, a deep learning framework specifically designed for text-to-image generation. By integrating image processing algorithms into the workflow, the model gains a better understanding of the input sketches, resulting in visually coherent and meaningful output images. The study explores the application of the Stable Diffusion model in the context of architectural design, showcasing its potential to enhance the visualization process and support designers in generating accurate and compelling representations. The efficacy of the method is evaluated through qualitative assessment, demonstrating its effectiveness in bridging the gap between initial sketches and photorealistic renderings. This research contributes to the growing body of knowledge on the integration of machine learning and image processing in architecture, providing insights and practical implications for architects, design professionals and researchers in the field.

Keywords: Architectural visualization, sketch-to-image transformation, machine learning, image processing, stable diffusion model.

1. INTRODUCTION

The field of architecture is characterized by constant evolution and transformation, driven by the inherent complexity and originality of the design process [1]. Recent advancements in technology, particularly in artificial intelligence (AI) and machine learning (ML), have opened up new possibilities in architecture, revolutionizing design processes and equipping designers with intelligent and efficient tools to achieve innovative and impactful outcomes [2, 3]. Recent digital approaches play a pivotal role in shaping contemporary architectural design and inspiring alternative solutions by encompassing transformative principles and philosophies [4].

This article aims to introduce a machine learning-based method called Stable Diffusion, which was introduced in 2022 as a deep learning framework for text-to-image generation. While its primary application is generating detailed images based on textual descriptions, it also holds potential for other tasks such as modifying or expanding the content of an image [5]. Notably, our approach incorporates image processing

* Corresponding author: ilker.karadag@cbu.edu.tr (İ. KARADAĞ)
[1] Manisa Celal Bayar University, Türkiye
ORCID: https://orcid.org/0000-0001-7534-2839
Content of this journal is licensed under a Creative Commons Attribution-Non Commercial No Derivatives 4.0 International License.

algorithms to scan sketches before the transformation process, enhancing the model's ability to comprehend the input sketches and produce visually coherent and meaningful outputs.

The Stable Diffusion model leverages the power of machine learning to capture visual details associated with textual descriptions, facilitating the generation of realistic images. This article provides a basic explanation of the Stable Diffusion model, delving into its underlying principles and the methodology employed to transform sketches into realistic images. Additionally, the potential applications of this model in architectural practice are discussed and the success rate achieved by the proposed approach is evaluated.

By shedding light on the integration of AI and ML technologies in architecture, this article serves as a valuable resource for academics, researchers, and architects interested in exploring the implications and opportunities arising from these advancements in the design process.

2. A BRIEF REVIEW OF AI GENERATED ARCHITECTURE

The realm of generating synthetic architectural images involves the utilization of generative models, which tackle image synthesis tasks by learning implicit statistical distribution. This can be achieved through techniques such as Generative Adversarial Networks (GANs) [6] or diffusion models [7, 8]. Diffusion models have gained prominence over adversarial networks due to their ability to address training convergence issues and produce satisfactory outcomes under proper guidance.

One notable open-source diffusion generative model is Stable Diffusion, which specifically focuses on text-to-image conversion. Developed with the collaboration of industry players like CompVis, Stability AI, and Runway ML, Stable Diffusion operates on LAION datasets. It offers fine-tuning capabilities and can be implemented through an advanced GUI (AUTOMATIC1111) with minimal computational requirements. Moreover, it supports local mode as well as cloud-based execution through collaborative tools like Google Colab.

From a user interface (UI) perspective, the generation of synthetic content typically involves two key elements: a text input (prompt) and the configuration of relevant parameters to facilitate probabilistic prediction and image generation.

The Stable Diffusion model possesses the capability to generate novel images starting from scratch by leveraging the mentioned text prompt that specifies the desired elements to be included or excluded from the resulting image [9]. This process is commonly referred to as "guided image synthesis" and involves the model's diffusion-denoising mechanism, which enables the incorporation of new elements into existing images based on the provided text prompt [10]. Furthermore, the model offers the functionality to partially modify existing images through techniques such as in-painting and out-painting, provided that the user interface supports these features. Numerous open-source implementations are available to facilitate the utilization of these functionalities [11].

In addition, Stable Diffusion incorporates a complementary sampling script called "img2img." This script takes as input a text prompt, the path to an existing image, and a strength value ranging from 0.0 to 1.0. The output of the script is a new image derived from the original image, which incorporates elements specified in the text prompt. The strength value determines the level of noise introduced into the resulting image. A higher strength value introduces greater variation within the image; however, it should be noted that this may lead to a decrease in semantic consistency between the generated image and the provided prompt [9].

ControlNet, as introduced by [12], is a neural network architecture specifically designed to augment diffusion models through the integration of additional conditions. This innovative approach involves duplicating the weights of neural network blocks, creating both a "locked" copy and a "trainable" copy. While the "trainable" copy learns the desired condition, the "locked" copy preserves the integrity of the original model. By employing this strategy, the training process with limited datasets of image pairs does not compromise the performance of production-ready diffusion models.

An essential component of ControlNet is the "zero convolution," which entails a 1x1 convolution with initialized weights and biases set to zero. Prior to training, all zero convolutions produce zero outputs, effectively preventing any distortion resulting from the application of ControlNet. Notably, no layer undergoes training from scratch; instead, the process involves fine-tuning to maintain the security and reliability of the original model. Furthermore, this method offers the advantage of enabling training on small-scale or even personal devices, ensuring its practical applicability across various computational environments.

3. MATERIAL AND METHOD

In this section, the process of transforming sketches into realistic images using a combination of machine learning and image processing techniques is presented. The proposed methodology involves a unique and tailored approach specifically designed for this purpose (Figure 1). The steps of the methodology depicted in the flow chart are also given correspondingly in the following part.

Preparation of the Datasets:

The first step in our methodology involves the compilation of a dataset of architectural sketches. It is important to acknowledge that the sketches were sourced from online platforms; however, these sources provided a diverse collection of architectural design concepts, encompassing a wide range of architectural elements and styles. Moreover, the methodology is not dependent on the chosen sketches.

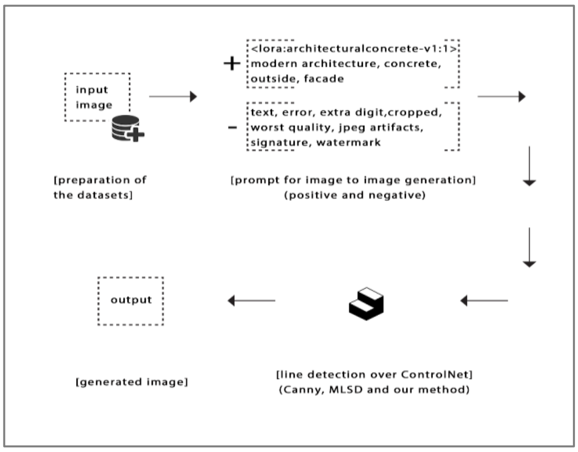

Figure 1 Flowchart of the methodology

Prompt for Image-to-Image Generation:

After the dataset is ready, prompts are included to help the image generating process. Two different prompt types are used in our methodology: positive prompts and negative prompts. Positive prompts provide precise directions for the desired outcome by describing design needs such as facade design and landscape elements. In contrast, undesirable characteristics including blurriness, poor quality, and the presence of watermarks are captured by negative prompts. The refinement of the image synthesis process made possible by these prompts allows for the creation of personalized and controlled outputs. After including prompts, the dataset processing begins. At this step, the power of deep learning and probabilistic inference are used to convert the sketches into photorealistic images that results in visually compelling outputs.

Line Detection over ControlNet:

A unique line recognition model that makes use of the ControlNet is used to maintain the key elements of the input sketches. This method confirms that the final images retain the original outlines and contours of the input sketches, improving the transformation's integrity. The performance of the model is thoroughly assessed in the results chapter by comparing it with well-known ControlNet models like Canny and MLSD. This comparative analysis offers insightful information about the performance and efficacy of our suggested approach in relation to architectural visualization.

Generated Image:

The proposed methodology makes it easier to convert sketches into realistic and visually convincing images that meet the specified design criteria by combining machine learning techniques, image processing algorithms, and the use of prompts. The end generated visuals not solely assist architectural visualization but also aid in the realm of architecture's knowledge and communication.

4. RESULTS AND DISCUSSION

In this section, the results and discussion of the study is presented to focus on the comparison between our model and the ControlNet models, namely Canny and MLSD. The outputs are provided in visual format, and the comparative findings for each input image are given.
The Canny model was used to process the first input image (Figure 2). While the Canny model can recognize fine details of lines, there are occasions where it misses the broad outlines and contours, losing the key elements.

The MLSD model, on the other hand, only does a superficial scan, which results in the loss of finer information. On the other hand, MLSD effectively catches the broad strokes as seen in the Figure 3.

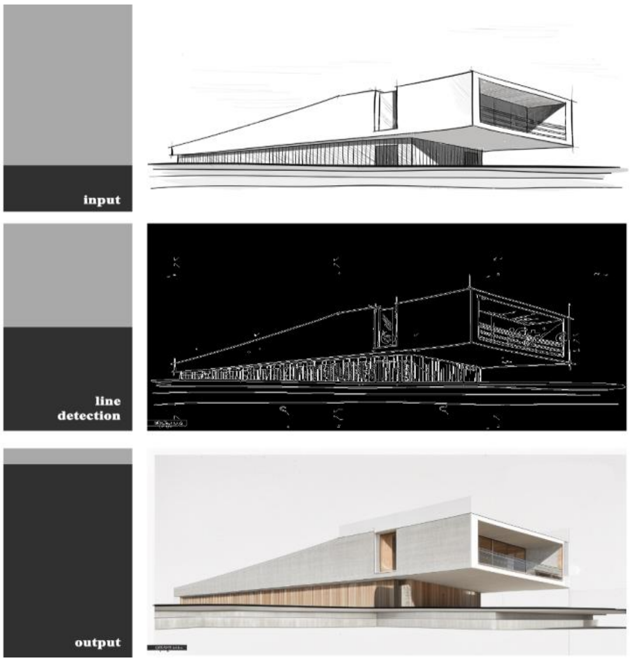

Figure 2 Output image generated by Canny Model

Figure 3 Output image generated by MLSD Model

Our model was created to find a balance between capturing small details and maintaining the key elements. It was inspired by the finest aspects of both models. High levels of coherence between the outputs produced by our model and the input image reflect enhanced alignment between the planned design requirements and the final visual representation (Figure 4).

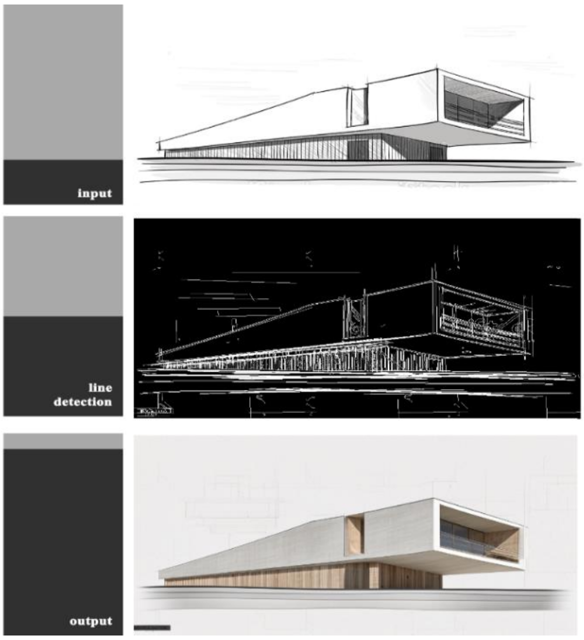

Figure 4 Output image generated by proposed model

For the second input image, the Canny model again displays its characteristic strength in detailed line recognition (Figure 5). However, it continues to struggle with accurately capturing the main outlines.

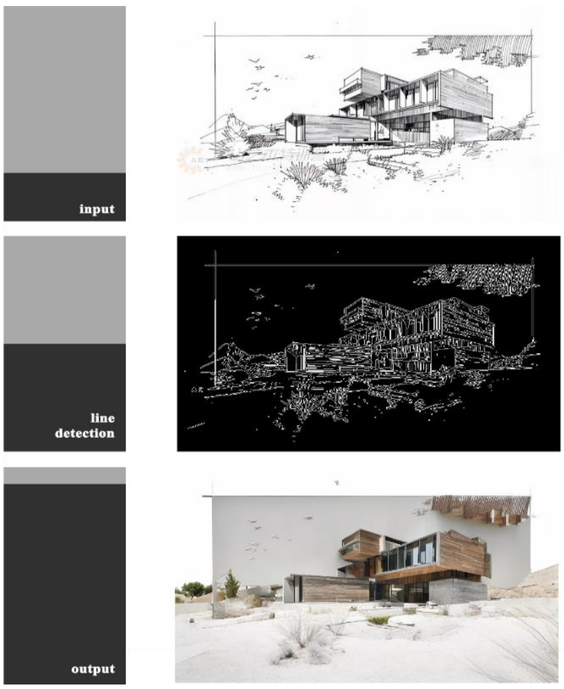

Figure 5 Output image generated by Canny Model

On the other hand, the MLSD model overlooks finer details, but manages to capture the main features. However, for this specific input, MLSD performed better than Canny Model (Figure 6).

Figure 6 Output image generated by MLSD Model

In contrast, our model combines the strengths of both models, resulting in outputs that exhibit an enhanced balance between detailed line recognition and accurate preservation of the main outlines (Figure 7). The outputs demonstrate a visually pleasing and semantically consistent representation that aligns well with the input image.

These comparative results highlight the benefits of the approach we suggest in producing a more well-rounded and satisfying result. Our model effectively captures both intricate details and the key aspects of the input sketches by combining the advantages of the Canny and MLSD models while addressing their drawbacks. With this method, the created images become more realistic, which improves architectural visualization and communication.

Figure 7 Output image generated by proposed model

The findings of our study highlight the importance of our unique model in converting sketches into lifelike visuals, so to sum up. Our methodology shows its effectiveness in producing visually appealing outputs that closely adhere to the specified design criteria by utilizing a combination of machine learning techniques, image processing algorithms, and the insertion of prompts. The results demonstrate our model's potential for increasing architectural visualization and encouraging enhanced comprehension and communication in the architectural community.

5. CONCLUSION

In this study, an alternative approach for rendering architectural visualization sketches into realistic visuals is presented. A methodology that combines the Stable Diffusion model with a novel line recognition model is built on ControlNet by utilizing the capabilities of machine learning and image processing techniques. The proposed method provides improved architectural visualization capabilities, enabling the creation of photorealistic images while maintaining the key elements of the original sketches.

The method to create new images from text prompts using the Stable Diffusion model is depicted, allowing for the insertion of certain design features and modifications while maintaining semantic consistency.

To further improve the integrity of the conversion, ControlNet was incorporated into the line identification process to guarantee that the final images accurately reproduced the original outlines and contours of the input sketches.

Prompts, both positive and negative, gave the image production process more control and personalization. Positive prompts might be used to explicitly specify design needs while negative prompts could be used to reject undesirable aspects. This made it possible to create customized, aesthetically pleasing outputs that adhered to predetermined design standards.

The comparison with popular ControlNet models like Canny and MLSD demonstrated the accuracy and efficiency of our suggested model. The results showed that our method performed ahead of the current models at producing realistic and visually appealing images from sketches. However, the scope of the study is limited to the field of architecture, therefore needs further validation to assess its success in other fields.
Overall, this study advances architectural visualization by offering a reliable and effective technique for converting sketches into realistic representations. Architects and designers can utilize this tool for visualizing design concepts and investigating alternate solutions thanks to the integration of machine learning, image processing methods, and prompts.
Further advancements can be made by enhancing the line recognition model, investigating other image processing techniques, and expanding the initial data set in future studies. It could also be interesting to

look into how our strategy might be applied to different fields of visual arts and design.

As a result, our work shows how machine learning and image processing can be used to improve architectural visualization, giving architects the capability to turn rough sketches into noticeable realistic representations.

The Declaration of Conflict of Interest/ Common Interest
"No conflict of interest or common interest has been declared by the authors".

The Declaration of Ethics Committee Approval
This study does not require ethics committee permission or any special permission.

The Declaration of Research and Publication Ethics
The authors of the paper declare that they comply with the scientific, ethical and quotation rules of SAUJS in all processes of the paper and that they do not make any falsification of the data collected. In addition, they declare that Sakarya University Journal of Science and its editorial board have no responsibility for any ethical violations that may be encountered and that this study has not been evaluated in any academic publication environment other than Sakarya University Journal of Science.

REFERENCES

[1] Frazer, J 1995, Evolutionary Architecture. London: Architectural Association, 1995.

[2] Chaillou, S, "AI + Architecture | Towards a New Approach," Master's Thesis, Dept. Arch., Harvard University. Cambridge, MA, 2019.

[3] I. Karadag, O. Z. Güzelci, S. Alaçam, "EDU-AI: a twofold machine learning model to support classroom layout generation," Construction Innovation, Sep. 2022.

[4] E. Kurucay, I. Karadag, "Computational Approaches in 21st Century Architectural Design: Defining Digital Representation Methods," Duzce University Journal of Science and Technology, vol. 10, no. 3, pp. 1201–1217, Jul. 2022.

[5] N. Dehouche, K. Dehouche, "What's in a text-to-image prompt? The potential of stable diffusion in visual arts education," Heliyon, vol. 9, no. 6, p. e16757, Jun. 2023.

[6] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio "Generative adversarial networks," Communications of the ACM, vol. 63, no. 11, pp. 139–144, Oct. 2020.

[7] Ho, J., Jain, A., Abbeel, P. "Advances in Neural Information Processing Systems", in Denoising Diffusion Probabilistic Models, Vancouver, Canada, 2023, pp. 6840–6851.

[8] P. A. Geroski, "Models of technology diffusion," Research Policy, vol. 29, no. 4–5, pp. 603–625, Apr. 2000.

[9] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer, "High-Resolution Image Synthesis with Latent Diffusion Models," 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2022.

[10] C. Meng, Y. He, Y. Song, J. Song, J. Wu, J. Zhu, S. Ermon, "SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations", arXiv [cs.CV]. 2022.

[11] AUTOMATIC1111. (2022, Nov. 10). Feature showcase for stable-diffusion-webui, GitHub [Online]. Available at: https://github.com/AUTOMATIC1111/stable-diffusion-webui-feature-showcase (Accessed: 12 June 2023).

[12] L. Zhang, M. Agrawala, "Adding Conditional Control to Text-to-Image Diffusion Models", arXiv [cs.CV]. 2023.