FINAL PROJECT REPORT - DCP4300 - AI IN THE BUILT ENVIRONMENT

Text to Image Synthesis with VQGAN + CLIP:

The Future of Architecture and the Built Environment

Professors

Karla Saldaña Ochoa
Chaofeng Wang

Group Members

Natalie Bergeron
Mariana Capuchinho
Baichuan Liu
Merlina Operta
Stephanie Roberts

Research Questions

How can GAN simplify the architectural/built environment design process?

What makes GAN useful versus traditional creative design processes (e.g. physical modeling and mapping, sketching, collaging)? Further, what makes GAN supplemental to these processes?

What is the relevance of the current uses of GAN in architecture/built environment?

What are the potential/future uses of GAN in the architecture/built environment professions?

Abstract

Through an analysis of GANs related to Architecture and the Built Environment, this report documents the exploratory work of applying GANs in the profession and how it aids the creative process. The code on this project is a GAN: VQGAN + CLIP - the VQGAN being a Generative Neural Network, and CLIP, unlike VQGAN, is a trained model rather than a generative model. Both enable the generation of images using AI, thus opening new possibilities to develop designs in a quicker manner. The images from this code are generated in two different forms: one using only text to generate images, and the other combining text and images to generate output images; the text-to-image method is what is utilized in this project. The project began by collecting a database and describing the collection in such a way that proves relevant for architectural outputs. With the capabilities of VQGAN + CLIP, design intentions are then questioned. Those within the architectural and built environment fields should find it useful towards understanding what particular steps can be taken to influence generative designs, such as collecting a specific database with certain biases. All should come to understand how powerful of a tool this is for decision-making, and especially as a supplement for design professionals everywhere.

Introduction of GAN: VQGAN + CLIP

What is VQGAN?

VQGAN is one type of neural network architecture - Adversarial Generative Neural Network. After learning by giving it a graph library, VQGAN can be made to generate many images. It also combines Convolutional Neural Network and Transformer.[1]

The difference between VQGAN and previous adversarial neural networks is that it can generate high resolution images, capable of reaching megapixels.[2]

And the previous GAN also has the problem that the fitting interval is relatively small. For example, if the training set given to the GAN is all dogs, then it can only generate dogs. If there are many different individuals in a single image, then the generated image may become weird.

Because the training set of VQGAN is image net, this has many images in it. So, VQGAN will show a better fit than the previous GAN and can generate higher quality images.

However, the problem exists for VQGAN when it is trained from purely image data. If a person uses VQGAN individually, then they will not be able to get a specific image by VQGAN. In this case, it is necessary to find a tool that can connect text and image - CLIP.

What is CLIP?

CLIP, unlike VQGAN, is a trained model rather than a generative model. It is a supervised learning model that uses images and their corresponding textual descriptions, and after learning, it is able to determine the matching between the content of the images and the textual descriptions. It builds a bridge between text and images.

When these two are combined, they become a text-to-image model that generates variable-sized images given a set of textual cues (and some other parameters).

After one enters a text description into the model and adjusts the relevant parameters, CLIP can guide VQGAN to generate images based on the text description content.

How to Play with VQGAN + CLIP

Step 1: Google Colab - It provides access to dedicated GPUs that Google operates through the cloud.

Step 2: Setting Up the Notebook - Select Run GPU, create internal file storage, and install the code libraries.

Step 3: Models - Select and install the model of VQGAN, run “Load libraries and definitions” cell.

Step 4: Execution - Determine the parameters of the image and generate it.

The Parameters

`prompts:` These are text prompts that CLIP will convert into suggestions for VQGAN.

`width:`

`height:` The width and height of the generated image in pixels.

`model:` The model used by the machine.

`display_frequency:` It will show how many iterations the machine will run before printing something into the text box below the cell.

`initial_image:` An image for the machine to begin with in place of a noise sheet.

`target_images:` Target images are pictures that VQGAN will “aim for” when generating the image.

`seed:` It will determine the map of noise that VQGAN will use as its initial image.

`max_iterations:` The number of iterations the machine will run through before terminating the process.

One would then run the model by adjusting these relevant parameters to obtain the outputs. In this project, there was also a comparison of the outputs of the model run with different parameters.

Problems GAN Solves in Architecture and the Built Environment

Architecture and the built environment are two professions that require an extensive, and arguably exhaustive, creative process prior to completed work. This applies to buildings, landscaping, furniture, and more. According to Neil Leach, a professor at Florida International University, AI "can potentially offer...insights into how the mind works, and so too into how architects are trained to think."[3] The brains of designers has been pondered by Harvard psychiatrist in 1974, Arthur D. Colman, in terms of how "...there are unique processes operating in design and on design professionals which make it extremely difficult for them to look beyond the immediate pragmatic and practical variables affecting their work;" this continues to apply today.[4] Thus, this research proposes that GAN resolves not only practical problems of systematically sifting through new opportunities and ideas to transform the design practice, but also issues of temporality and mental capacity due to its digital format.

More overtly understood, GANs simplify, or rather resolve, the architectural and built environment design process for those involved. In the realm of machine learning, work is distributed to a computer, rather than solely on the human. The mental utilization required for and of creative thinking and decision-making is decreased on the designer as machines become a worker themselves; there is now a team effort towards an anticipated concept. Time is lessened in the overall design process since there is more opportunity to visualize an outcome thanks to machine learning's outputs. There is a recognizable speed of returned work that began at the turn of the 21st century which

revolutionized labor, and the impact is larger now more than ever for designers in applications discussed below.

There is controversy, however, with how decision-making is shared between people and computers. An article in the journal *Virginia Law Review* implores this; “Machine-learning tools are perceived to be eclipsing, even extinguishing, human agency in ways that compromise important individual interests.”[5] This poses the question of whether or not designers should feel a moral obligation, or even an objection, when gathering design ideas from computer-generated imagery, like with VQGAN + CLIP. Nonetheless, this may become an unnecessary concern in design when one recognizes how construction documentation still requires a level of precision that random, generated work cannot provide for a feasible and final solution.

To expand upon specifics in the design process, traditional design methods like modeling, mapping, collaging, and sketching are transformed and even supplemented with GAN. These analog forms of fabrication and compositional ways of generating design concepts are comparable to the results of GAN imagery. Yet both have enough differences to rely on one another, rather than only relying on one method of generation. To explain, depending on the parameters given on a GAN model like VQGAN + CLIP, there is no true systematic definition of inputs and representation of outputs, regardless of initial and target images. In other words, there is still an ambiguity hidden within the computer’s generation due to it essentially becoming a conglomeration of images from a trained database. This gives the designer another tool to work with; their personal interpretation of what was machine-generated provokes design concepts that were never the person’s initial thoughts.

Referencing the controversy prior, the designer becomes an integral part to the design process here with the basis that they discover the systematic quality needed to make the outcome work in addition to GAN. A model can be physically constructed, but GAN inspires particular depths, textures, and layers of materiality, for instance. Digital and/or physical maps and collages hold their own constraints on the basis of what is being analyzed or studied, yet GAN may pose useful

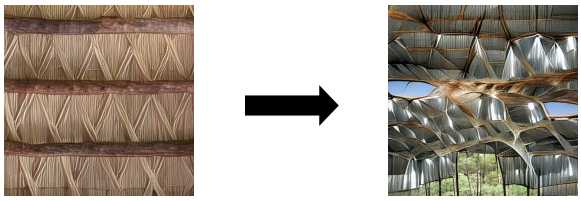
when deciding what compositional qualities are eventually wanted; this includes, again, textures, weights, colors, etc. Finally, sketching visions from one's brain can be more taxing than if GAN stimulates the designer with new opportunities for visions or architectural partis for a project. Therefore, GAN is a useful tool for simplifying and resolving the architectural and built environment professions in a plethora of ways.

Project: Utilizing GAN

The parameters for VQGAN + CLIP were tested in multiple ways. Listed here are three examples for how to explore the notebook.

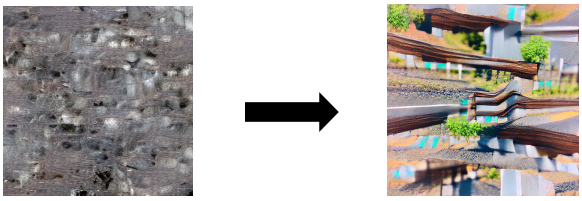
Using Initial Images

One way this code was implemented was inputting textures from the collected dataset as initial images, in addition to a text prompt. As previously mentioned, this eliminates the use of a noise sheet. By incorporating the dataset in this way, the resulting image is more related to the desired texture. Below is an example of this methodology, where the initial image of "nature material" is used, along with the prompt: "large span roof combines metal and braid pattern and natural and translucent materials."



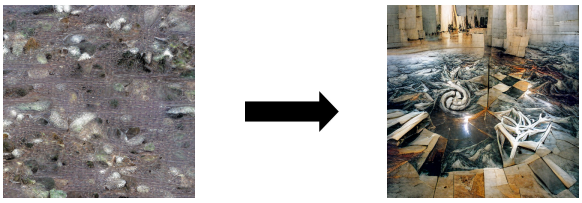
Using Only Text Prompts | Descriptive Words

Another approach for exploring this code is using only text prompts. Therefore, the code would utilize a noise sheet in the beginning to develop a resulting image. Below is an example of this with the prompt: "roads road rail rails railroads railing railings tracks track gravel metal wood repetition repeated pattern lines line linear bolted."



Using Only Text Prompts | Descriptive Sentences

Similar to the previous process,this method also uses a noise sheet and prompt. However, the text prompt is formatted more like a sentence, explaining the desired image. Below is an example of this with the prompt: “floor made of marble and metal.”



Project: Collecting a Personalized Database Towards Particular Results

The data collected was based on the goals of this project. The database consisted of textures and materials from rendering textures and photographs. The categories the data was sorted in was based on common architectural textures used in studios and projects. The main categories the textures were assigned to:

Concrete	Ornaments	Glass
Fabric	Plaster	Wood
Floors	Plastic	Paint
Ground	Road	Nature
Grunge	Roofing	Water
Stone	Rust	Sand
Metal	Soil	

Secondary categories details such as light, dark, tile, carved, were applied to those within the files.

As a result the image output would be a mix between materials or concepts like “atmospheric stone tile” or “colorful urban landscape.” This could then inspire one to create the same or similar material, a space, or concept.

Database Collection for Project.

The textures in the figure above were downloaded from large image databases such as texture.com. Then they were sorted into the categories with the main categories and 3-5 sub-categories.

While running the VQGAN + CLIP the chosen dataset came from an already established database, sflckr.

Datasets you can use for commercial purposes:

faceshq: ☐

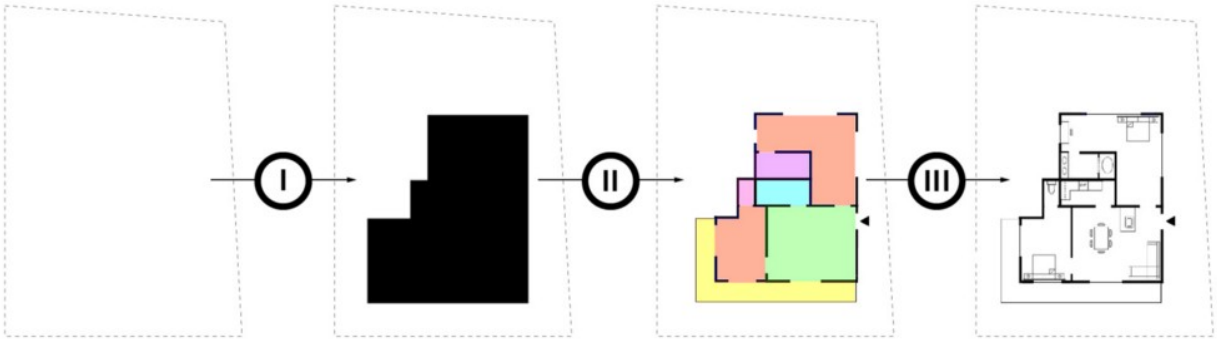
sflckr: ☑

The required database needed to be within the hundreds of thousands or even millions of photos. The bias in the database would be that the individuals determining how “grunge” or “concrete” a texture may be is based on appearance. This would differ from person to person depending on their personal bias and opinion on materials and textures.

Proposed Application of GAN for Designers

Existing Uses

Understanding the potential importance of GANs in architecture and the built environment requires analysis of the existing uses. One example of current research that utilizes GAN in the built environment is done by ArchiGAN, which studies and develops architectural floor plans through “building footprint massing,” “program repartition,” and “furniture layout.”[6] This includes inputting a specific outline of land in order to output an overall parti, which then is given options for programmatic and interior plans for the residential spaces.[7] This process can be seen in the diagram below.



Stanislas Chaillo, *ArchiGAN: a Generative Stack for Apartment Building Design*, diagram, Nvidia Developer, July 17, 2019, https://developer.nvidia.com/blog/archigan-generative-stack-apartment-building-design/.

Potential Uses

A similar process could be applied to a building once the overall form is developed. For instance, in order to explore possible materiality and surface conditions, designers could implement the GAN outlined in this report to invent the details of a building. A potential first step could be deciding what materials the proposed design will utilize. The collected dataset in this project included textures within the following categories:

Concrete	Ornaments	Glass
Fabric	Plaster	Wood
Floors	Plastic	Paint
Ground	Road	Nature
Grunge	Roofing	Water
Stone	Rust	Sand
Metal	Soil	

These textures are common building materials and were used as inspiration for defining the desire images within the following building categories:

Roof	Walls
Facade	Floors
Detailing	Landscape/Environment

To explicate these components further, using GAN for roof designs expands the opportunities for what a roof may become in a final design; questions of materiality and joint connections could be derived from the eventual outputs. Similarly with facade and detailing of architecture, materiality and texture is understood through GAN, as well as what role this may play as an exterior boundary towards a site’s contextual surroundings. As for walls and floors of finished products, GAN becomes a useful tool towards the way space is shaped. After all, generated images may result in particular compositions and arrangements that may inspire post-modernism’s continued fluidity of the floorplan. With the landscaping and environmental conditions

generated by GAN models, the current exterior limits of landscape architecture are questioned as GAN changes the traditional view of what landscaping currently is; the generations are merged into a new creation. Thus, generally this process provides new opportunities for textures, organizations, and joints at various scales within the project’s design.

Conclusion

The generation of images using VQGAN + CLIP have created interesting and promising outputs for future designers and their intents. The resulting work from this project proves that the creative process is supplemented with this tool to generate unique, unprecedented, and never-before-thought ideas in an unorthodox manner. To summarize, this exploration culminated in a balance between the relationship in design of people and machine learning; the diversity of generative design is just the beginning.

References

1. *"Introduction to VQGAN+CLIP,"* Accessed April 23, 2022, https://docs.google.com/document/d/1Lu7XPRKlNhBQjcKr8k8qRzUzbBW7kzxb5Vu72GMRn2E/edit.
2. Alexa Steinbrück, *VQGAN+CLIP – How does it work?* Accessed April 23, 2022, https://alexasteinbruck.medium.com/vqgan-clip-how-does-it-work-210a5dca5e52
3. Neil Leach, "Architectural Hallucinations: What Can AI Tell Us About the Mind of an Architect?," *Special Issue: Machine Hallucinations: Architecture and Artificial Intelligence* 92, no. 3 (May/June 2022): 66-71, https://doi.org/10.1002/ad.2815.
4. Arthur D. Colman, "Notes on the Design Process: A Psychiatrist Looks at Architecture," *Journal of Architectural Education* 27, no. 2/3 (June 1974): 19-26, 55, https://www.jstor.org/stable/1423860.
5. Aziz Z. Huq, "A RIGHT TO A HUMAN DECISION," *Virginia Law Review* 106, no. 3 (May 2020): 611-688, https://www.jstor.org/stable/27074704.
6. Stanislas Chaillou, "ArchiGAN: a Generative Stack for Apartment Building Design," Nvidia Developer, July 17, 2019, https://developer.nvidia.com/blog/archigan-generative-stack-apartment-building-design/.
7. Ibid.