Semantic AI models for guiding ideation in architectural design courses

Emmanouil Vermisso
School of Architecture
Florida Atlantic University
111 East Las Olas Boulevard
Fort Lauderdale, FL 33301, USA
evermiss@fau.edu

Abstract

This paper discusses the combination of available artificial intelligence (AI) models, i.e. Neural Language Models (NLMs) with trained GANs and human interpretation to facilitate architectural ideation. The workflow identifies conceptual scenarios for a speculative design using semantic prompts. Results become visual references to complement revised semantic descriptions for guiding a VQGAN+CLIP model, leveraging control over the outcomes, which are then sorted using dimensionality reduction and further curated for training other models (GANs). The NLMs' interpretation of text input increases the possibility of spanning greater semantic distances, towards creative visual outcomes, while the nested workflow of AI-human steps enables an automated query of a larger solution space. Furthermore, it considers the problem of low-bandwidth, reductionist encoding of visual data (Hadamard, 1945) resulting from verbal-based (NLM) processing models (LeCun, 2021) that potentially constrain design agency.

1. Introduction

A reconsideration of the binary separation of intelligence as human and artificial, and the perception of intelligence as a "spectrum" (Bratton, 20 Feb.2022) may enhance human capacity to make decisions by introducing AI as an assistant for design creativity. This paper tries to propose a pedagogical structure where the agency of human designers is complemented by generative tools including language-based AI models (i.e.VQGAN+CLIP) for an architectural design studio course. The workflow involves a back-and-forth shift of "agency", trying to establish the type of framework where AI can make a contribution in relation to human-driven decision stages. Due to current limitations of Artificial Neural Networks (ANNs), and specifically Neural Language Models (NLMs), applying AI's potential in a heuristic manner seems more feasible than pursuing a clear problem within a deterministic design approach. Furthermore, NLMs' limitation when used in isolation warrants their integration with other ANNs (Bolojan, Vermisso, & Yousif, 2022).As a result, this project theme offers an open-ended, bottom-up way to craft a design agenda. The research involves the integration of computational generative tools with more analog design steps to re-imagine design scenarios for re-assembly of large, broken cargo ships and oil tankers into informal dwelling communities. The specific geographical and socio-cultural backdrop (Fig.1) is used to introduce NLMs and other ANNs during the early stages of ideation. In addition, we contemplate on further design tasks where AI may be useful, like the identification of ways to synthesize particular semantic features with each other. The project is inspired by the empirical creativity of ship-breaking communities in Bangladesh and operates on the fringe of the regular economy. It is viewed as an exercise that is guided by (human-computational) creativity, while being grounded by material and assembly constraints. Pedagogically, it offers a methodological exploration of design agency during a process of ideation and generation.

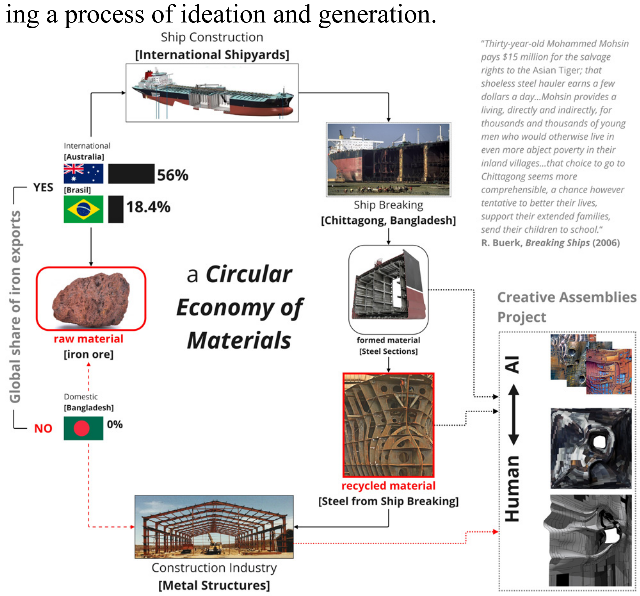

Figure 1 Iron supply for ship construction is recycled from unregulated ship-breaking (Bangladesh). The design project lives along the margins of this conceptual space, assisted by human and automated (AI) input.

Design Inquiry: The process of invention

The proposed workflow leverages generative models to operate in a heuristic fashion, allowing an expanded search in the solution space, for initial design conditions like visual inspiration. During this early stage, when a clear question or problem is absent, a flexible approach can trigger creative thinking. This argument has been widely discussed by experts including mathematician Jacques Hadamard, in his study of the cognitive mechanisms of mathematical invention, and the identification of a given problem: "*What* *shall we try to discover? What problem shall we try to solve?*" (Hadamard, 1945). Hadamard mentions neurologist Édouard Claparède's distinction between two kinds of invention: the former looks for the means to reach a particular goal (*question* to *solution*), while the latter imagines the usefulness of a fact that is discovered, possibly to be exploited much later (*answer before question*). Hadamard notes that human progress lies largely on the second kind. As Donald Rumsfeld famously remarked, besides what we know that we know and don't know, there exist "unknown unknowns", things which we would not even consider in our selection. (Zak, 2021) Adopting a flexible methodology opens the way to such unexpected discovery.

2. State-of-the-art: Architectural AI & NLM

Although Artificial Intelligence has begun to concern architects in the past 3-5 years for the most part, Makoto Sei Watanabe was one of the first architects to design using an AI interface in the late 1990s. His work used a number of inputs processed through an algorithm, which returned an output which was evaluated and scored by the designer, helping the algorithm to revise its performance. (Watanabe, 2005) Today, a number of architects use Generative Adversarial Networks (GANs) (Goodfellow, et al., 2014) because of their high quality results in computer vision, to perform data interpolation (StyleGAN) or domain transfer (CycleGAN) with impressive results (Bolojan, The Hitchhiker's Guide to Artificial Intelligence: AI and Architectural Design, 2021) Unfortunately, these kind of models can be computationally expensive, requiring precise curation of data and substantial computational resources, to reach very high resolutions. A type of network which has recently offered generative capacities through language are Neural Language Models, released in 2021. NLMs like CLIP, DALL-E and VQGAN+CLIP are pretrained on large datasets, so they are computationally cheap. (Rodrigues, Alzate-Martinez, Escobar, & Mistry, 2021) used VQGAN+CLIP for design inspiration, using text prompts in literal (analogy) and abstract (metaphorical) ways, combined with photographs and sketches. It is important to note that Watanabe mentioned, in his work, the separation of one architecture "condition" (i.e. form) from others, considering separate inputs within an AI-assisted workflow. This work aligns with this intuition, proposing to replace singular AI models with multiple ones which perform various tasks, including NLMs and GANs.

3. Design Methods

A design workflow is proposed, which includes manual curation of semantic descriptors, automated generation of visual spatial scenarios in 2D (Wombo "Dream" AI) and sorting of the visual outcomes (PixPlot), manual qualification of the results and a second layer of automated 2D scenarios generation (VQGAN+CLIP). The current state of this workflow (Fig.2) offers a robust method for generating an expanded search space of conceptual spatial references for further interrogation. Students used these methodologies as a foundation for developing a catalogue of design scenarios which were 3d modeled, implementing panelization strategies (GH+Ivy) to rationalize the surfaces and speculate how these could be constructed from reclaimed steel from ship-breaking. The discussion herewith will focus on the early stages of the catalogue material, and how this process can be refined using additional generative models beyond NLMs. Overall, this is a priori a speculative exercise to assess a process of varying agency, where the designer relinquishes control during certain parts of the process while he interferes elsewhere. Among the objectives is intuiting how to steer NLMs towards results which align with the design narrative, and identify how to use these automated steps in a meaningful way (i.e. producing nested "collage"-like drawings for concept generation.)

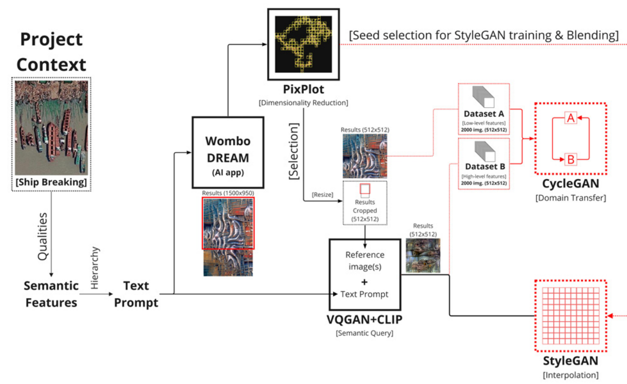

Figure 2. Workflow connecting NLMs with GANs.

NLMs for preliminary scenario generation

The "Dream" app by WomboAI is based on some internalized neural-language-model in the fashion of other existing language-based AI models like "DALL-e" or "VQGAN+CLIP", which contain a deep artificial neural network (i.e. GAN) for classification of visual data based on linguistic correlation and one for generation of visual results based on the classification. Such AI models can generate fairly complex visual results from text prompts of varying specificity. It is unquestionable that this type of results are seductive due to their visual complexity and speed of generation. However, it is important to question their significance for design decision making. As far as usefulness goes, the designer's agency is minimized, as it is difficult to intervene within the language-based AI model (in this case, Wombo's "Dream" app). Although we cannot accept these AI-generated outcomes as direct valid steps in architectural designing, they are likely references in the act of ideation, helping inspire and steer our design inquiry towards certain directions which may have otherwise remained latent. To efficiently manage the large number of results which can quickly accumulate from these automated processes, it is important to categorize the properties (spatial, tectonic, visual, formal etc.) which qualify a scenario as interesting, successful and/or optimal.

The creation of separate categories which relate to the various semantic descriptors is necessary to identify areas of interest in the visual results. Naturally, every outcome displays interesting high and low-level features. (Fig.3)

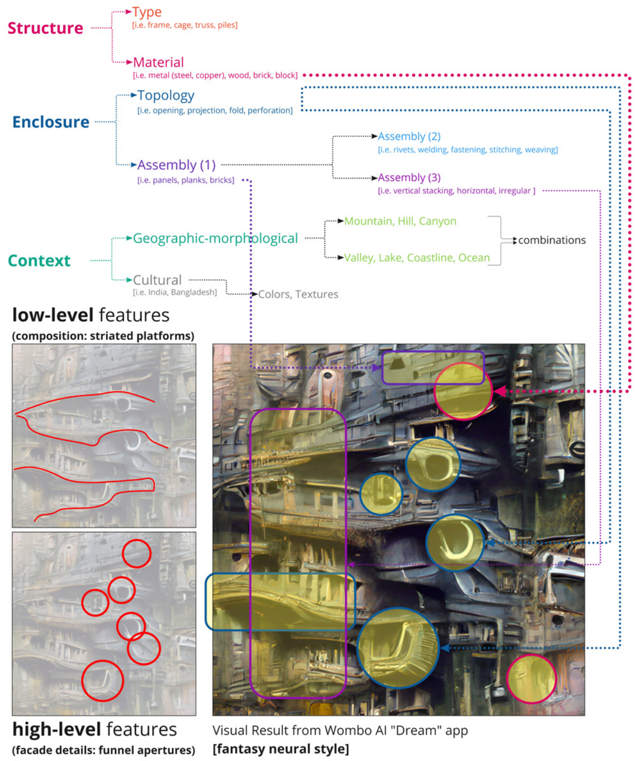

Figure 3. Schematic structure to identify correlation between design aspects and semantic feature description (text prompt); High and low-level image features.

It is not important to establish an ideal scenario, but to extrapolate qualities which can be translated into inspiring and applicable design features, i.e.strategies to create openings on a large surface by *perforating, folding, bulging, projecting*. Combining particular features should also be considered. Figure 3 gives an example of "high" and "low" level features. As a note, the same characteristic (i.e. a certain type of opening shape) can be read as a high-level feature in an image or a low-level feature in another, depending on its scale relative to the overall composition. In this paper, we will refer to properties like overall compositional structure as a "low-level" feature while "high-level" features will identify finer details inside a composition (openings, textures, individual structural members, etc). While humans are good at recognizing patterns, sorting through large samples of data with thousands of features requires another interface. In order to sort the AI-generated images form the NLMs (stage 1), we used dimensionality reduction (UMAP) and clustering algorithms (K-means) in PixPlot (Fig.4). The data comprised 861 images, grouped into 10 clusters based on their characteristics. It was clear that the network organized results from the same text prompt or same neural style (option in Wombo AI) together, picking high-level features like *woven configurations, bulging topologies, pleating,* as parameters for clustering, as well as other more visual ones like color distribution. Based on the PixPlot matrix, results with certain complex features were selected as reference images to guide a VQGAN+CLIP model. We are currently looking at selecting particular successful seeds from radically different clusters for blending through StyleGAN later, because blending seeds from different feature families may output interesting results.

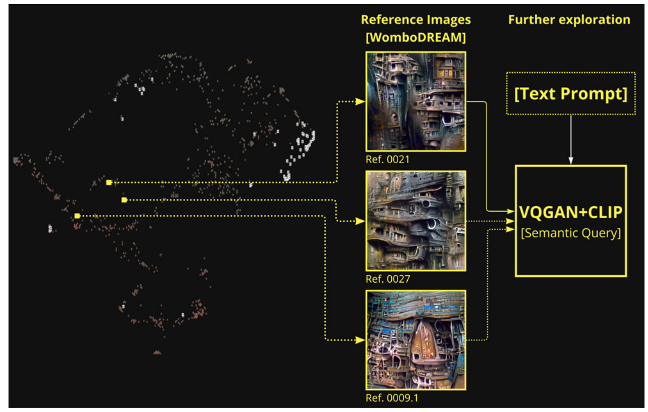

Figure 4. Visualization of 861 results (left) from the "Dream" app, sorted in 10 clusters, using PixPlot.

4. Results & Discussion: Encoding intentions

According to experts, words -like other aspects of our existence- are gradually "domesticated", entering a fully conscious state of acceptance via "methodical selection" and selection for regular use depending on our preference. (Dennett, 2017) A characteristic example of word types which are not "preferred", but necessary, are technical terms in a scientific field, which are commonly accepted to describe the field itself. In architecture, an array of semantic descriptors has been consciously adopted when referring to design properties or construction attributes. Selecting such semantic references to guide NLMs, i.e. Wombo "Dream" or VQGAN+CLIP is normal, leading to visual outcomes which are "expected", typically reflecting "realistic" qualities which the intended spatial configurations need to possess, to qualify as successful. If we are looking for something unique, creative, the choice of words needs to extend beyond the conscious selection of descriptors which are semantically close to each other and typical to the primary reference context (ship breaking), to ones which are not typically combined. We have tried to work with text prompts which supplement the 'expected' features (i.e. *metal ship hull; slum dwelling,* etc.) with ones which are unfamiliar and semantically distant (i.e. *fighter jet intake; woven exoskeleton*) (Fig.5) Increasing the semantic distance is an obvious idea; regarding the notion of "relevance", Boden mentioned that it is difficult to traverse large conceptual distances, but it is also more likely to reach novel concepts. (Boden, 2013) We tried to "lend" our human intuition to the network's search descriptors via unusual prompts, to inquire what results might be obtained.

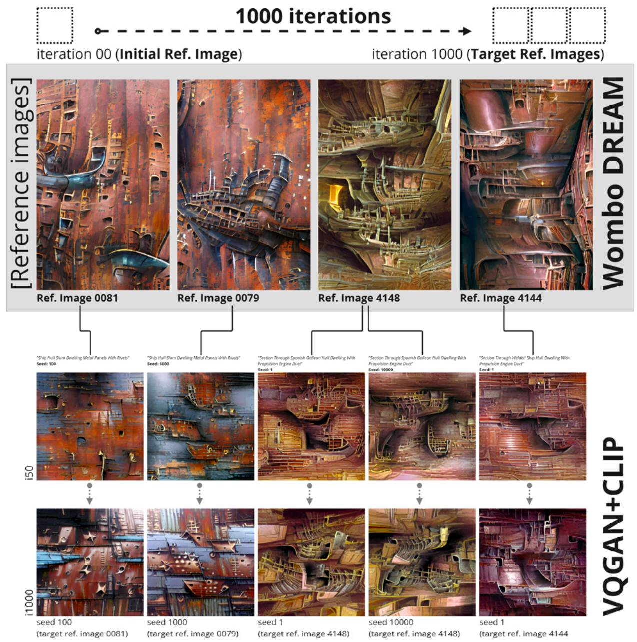

Figure 5. Wombo Dream results guide VQGAN+CLIP.

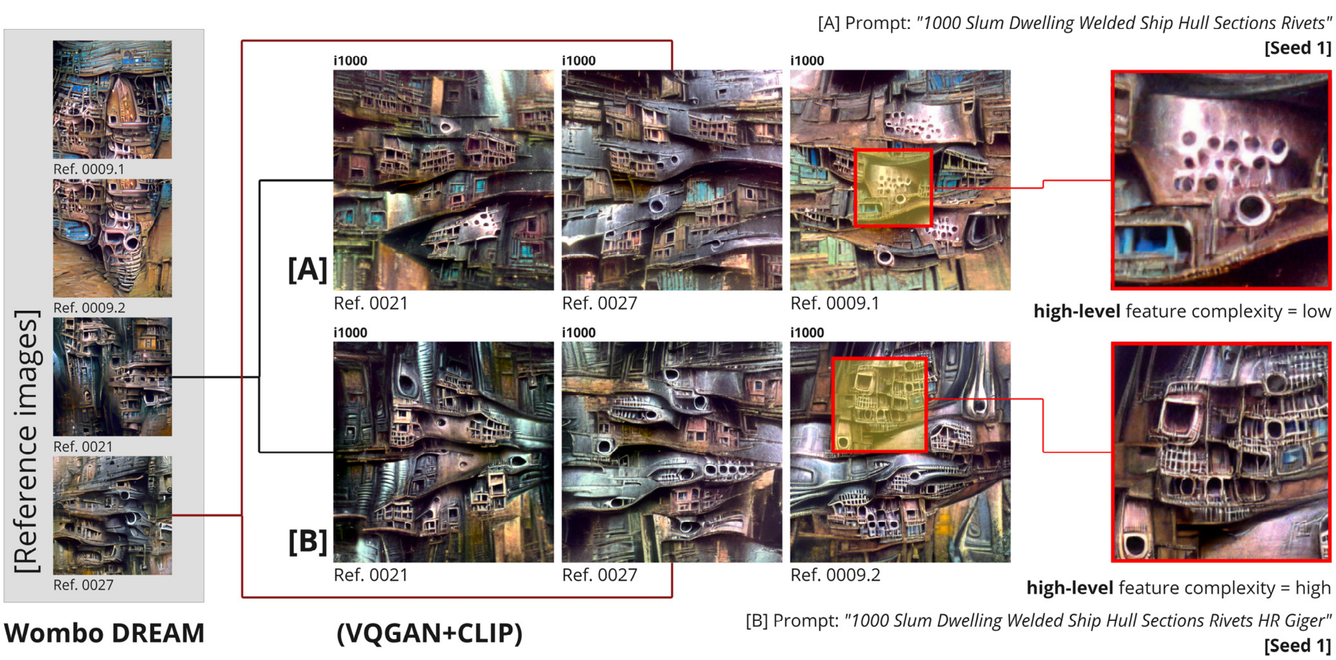

Figure 6. Scenarios generated with VQGAN+CLIP curate qualities consistent with the project theme: ship hull population with novel topological-tectonic features.

Fig.6 shows 6 results from a text prompt in VQGAN+CLIP with a target reference image; (one of four images -shown- which had been generated with Wombo was selected). The prompt for the first 3 trials was: "*1000 Slum Dwelling Welded Ship Hull Sections Rivets*" and the chosen seed was #1. As the reference images were generated in Wombo with a prompt including the phrase "HR Giger", we see the influence of that graphic language at a high-level feature is evident (but not exaggerated) in the results from the VQGAN+CLIP (A0021, A0027, A0009.2). We tried 3 more tests with the same prompt, adding "HR Giger" at the end of the prompt and keeping the seed to #1 ("*1000 Slum Dwelling Welded Ship Hull Sections Rivets HR Giger*"). It is clear that using the same semantic reference as the graphic reference in the target image is perhaps an exaggeration, because the high-level semantic features which relate to this formal language (the low-level arrangement does not seem to vary much) will be searched for both via language reference and image reference, resulting in overpowering visual results. (B0021, B0027, B0009.1) The image reference of this feature is sufficient to assign the preferable, as it works in a stronger sense than the text prompt, based on the results herewith shown (A/B009, A/B0021, A/B0027). However it should be noted that the results from the (B) trials demonstrate -despite their exaggerated "Giger" style, interesting details at the high-level resolution. Whether a semantic feature is pursued via both target image and prompt reference depends, therefore, on the specific scenario being considered.

Future Work: Feature Disentanglement in GANs

Due to the intrinsic limitations of language, introducing additional AI models is warranted to explore detailed semantic combinations further. We propose AI models which are good at interpolation, to combine known elements into new configurations (Boden, 2013). We are in the process of training a StyleGAN network, to later blend a number of qualified new seeds with selected features, as well as existing seeds from Wombo Dream. StyleGAN can perform feature disentanglement so these can be reconfigured into new organizations with emergent visual-spatial qualities.

Author Contributions

[EV] ideated and wrote the paper alone.

Acknowledgments

The author would like to thank his Architectural Design 10 students for their willingness to work with tools they were entirely unfamiliar with. The author would also like to thank his colleagues at FAU, Daniel Bolojan and Shermeen Yousif, for the broader research discussions which eventually led to this line of work. The author also acknowledges the research team at Yale Digital Humanities Lab, where PixPlot was developed, the people behind the WomboAI "Dream" app, and Katherine Crowson for making the VQGAN+CLIP model available.

References

Boden, M. (2013). Creativity as a Neuroscientific Mystery. In O. Vartanian, A. S. Bristol, & J. C. Kaufman, Neuroscience of Creativity (pp. 3-18). Cambridge, MA: MIT Press.

Bolojan, D. (2021, June 20). The Hitchhiker's Guide to Artificial Intelligence: AI and Architectural Design. Retrieved from www.digitalfutures.world: https://www.youtube.com/digitalfuturesworld/live

Bolojan, D., Vermisso, E., & Yousif, S. (2022). Is Language All We Need? A Query Into Architectural Semantics Using a Multimodal Generative Workflow. CAADRIA 2022: Post-Carbon. Sydney: CAADRIA.

Dennett, D. C. (2017). From Bacteria to Bach and Back: The Evolution of Minds. New York: Penguin Books.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., . . . Bengio, Y. (2014, June 10). Generative Adversarial Nets. Retrieved from https://arxiv.org: https://arxiv.org/pdf/1406.2661.pdf

Hadamard, J. (1945). The Mathematician's Mind: The Psychology of Invention in the Mathematical Field. Princeton: Princeton University Press.

Rodrigues, R. C., Alzate-Martinez, F. A., Escobar, D., & Mistry, M. (2021). Rendering Conceptual Design Ideas with Artificial Intelligence: A Combinatory Framework of Text, Images and Sketches. ACADIA 2021: Realignments: Towards Critical Computation.

Watanabe, M. S. (2005). Algorithmic Design/Induction Design: Three Kinds of Flow/Three Stations. Retrieved from https://www.makoto-architect.com: https://www.makoto-architect.com/kashiwanohaCSt.html

Zak, D. (2021). 'Nothing ever ends': Sorting through Rumsfeld's knowns and unknowns. Retrieved February 23, 2022, from https://www.washingtonpost.com/lifestyle/style/rumsfeld-dead-words-known-unknowns/2021/07/01/831175c2-d9df-11eb-bb9e-70fda8c37057_story.html