ARCHITECTURE, LANGUAGE AND AI

Language, Attentional Generative Adversarial Networks (AttnGAN) and Architecture Design

MATIAS DEL CAMPO
[1] *Taubman College of Architecture and Urban Planning, University of Michigan*
[1] *mdelc@umich.edu*

Abstract. The motivation to explore Attentional Generative Adversarial Networks (AttnGAN) as a design technique in architecture can be found in the desire to interrogate an alternative design methodology that does not rely on images as starting point for architecture design, but language. Traditionally architecture design relies on visual language to initiate a design process, wither this be a napkin sketch or a quick doodle in a 3D modeling environment. AttnGAN explores the information space present in programmatic needs, expressed in written form, and transforms them into a visual output. The key results of this research are shown in this paper with a proof-of-concept project: the competition entry for the 24 Highschool in Shenzhen, China. This award-winning project demonstrated the ability of GraphCNN to serve as a successful design methodology for a complex architecture program. In the area of Neural Architecture, this technique allows to interrogate shape through language. An alternative design method that creates its own unique sensibility.

Keywords. Artificial Intelligence; Machine Learning; Artificial Neural Networks; Semiotics; Design Methodology.

Figure 1. Attentional Generative Adversarial Network - Sample of Result.

PROJECTIONS, Proceedings of the 26th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2021, Volume 1, 211-220. © 2021 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

1. Problem

When Walter Gropius, at the beginning of his career, started to work in Peter Behrens office, he kept a terrible secret: he could not draw (Wigley 1998). He struggled with this deficiency in the environment of the Behrens office but famously became proficient in dictating drawings to his collaborators, demonstrating the ability of language to describe the complex material and spatial relationships in an architectural project (Führ, 1997). Synonymously Sol LeWitt expanded the concept of description into an entire career, dedicated to the efficacy of language (as instructions) and foreshadowing the emergence of programming and scripting as a means of artistic and architectural expression (Miller, 2016). So, what is the relationship of language to architecture?

An answer to this question could fill tomes. Therefore, for the sake of an appropriate length of the paper, the authors would like to rely on an abbreviated discussion on language, semiotics, semantics, and the implementation of language in machine learning (ML) processes. On a simple level language is capable of producing both the symbolic and social dimension of spaces, aiding in the calcification of these dimensions in the form of the built environment. To give the reader more background please allow for a brief excursion into Semiotics and its meaning for the project presented in this paper. Breaking down language means observing its constituting parts first: signs. A sign can be defined as a physical subject that carries meaning and is formed by the combination of signifiers manifesting a phonological image of the world. In addition, it contains a conceptual image of the world which can be described as its *signification*. In order to understand the meaning of other signs, it is necessary to untangle its structural relationships (Peirce, 1977) - akin to the structural performance of a building.

This base gave rise to two distinct branches of interrogation: the semiotics of communication and the semiotics of signification (Xanthos, 2010) - both of which play a distinct role in the application of AttnGANs (fig.1) in general. The former -semiotics of communication- is concerned with the exploration of the signs within the domain of communication and thus centers around aspects of human language. The latter -semiotics of signification- considers that any phenomenon that signifies anything can be considered part of the domain of semiotics, in doing so it covers a far larger territory than language alone. A territory that includes symbolic cultures that expand into material cultures (de Landa, 2016): Brand Logos, Instagram Images, YouTube videos, TikTok, memes, paintings, sculptures, architectures. This has been exemplarily laid out by Roland Barthes in his writings on the semiotics of signification which discusses the plurality of meaning in language as well as concepts of denotation and connotation. With this background in mind, we can expand this idea to the nature of the dataset used in the example presented in this paper. The database, Microsoft COCO (Common Objects in Context) was used to generate the images that resulted in the basic 3D models of the project. The database was primarily created not with architecture in mind, but for object recognition. As discussed in the previous section -sign, signifier, signification (Barthes, 1968) - language contains the toolset for recognizing the world. With this in mind, we have to consider in addition that an object is part of a scene - and thus to achieve a better object recognition, scene recognition is key (Xu et al, 2018). COCO achieves this by collecting images of everyday scenes that contain common objects contextualized in their natural environment. These objects are then labeled using per-instance segmentation in order to be able to localize the object in the scene more precisely. Microsoft employed a crowd worker (Jäger et al, 2019) approach in order to create more than 2.5mio labeled instances in 328.000 images which included instance spotting and instance segmentation. Project 24 Highschool (Fig.2) used this database in an alternative fashion. Instead of using it for the detection of objects in an image - for example for facial recognition, traffic control, object sorting, or the likes, we used the database in a generative role, reversing the flow of information in order to create images rather than analyzing them with the help of the labeled database. In order to do so, we employed an AttnGAN.

Figure 2. The PE center and dorms of the 24 Highschool. This image shows the distinct hand of the GraphCNN in full color. The coloration created by the process was embraced and forms a distinct feature of the project.

2. Methodology

In order to describe our materials and methods, we will rely on the example of the proof-of-concept project 24 Highschool Shenzhen. It was important for the authors to demonstrate the efficacy of the approach by implementing it into a demanding design process. The practice SPAN (Matias del Campo & Sandra Manninger) was invited to participate in a competition for the realization of a large new Highschool with 110.000m2 providing space for around 3000 Students. The program included classrooms, multipurpose hall, laboratory building, fab lab, physical education building, dorms, canteens and outdoor sporting grounds. The structural engineering was done by Bollinger & Grohmann. A project of this size demands a considerate positioning within the urban texture. The site is located in the northwest outskirts of Shenzhen and is in close proximity to the sea. The wedge-shaped site has its highest point in the northwest corner, eight-meter above sea level, and slopes from there towards the southeast.

2.1. INITIAL MODELING

In a first step a series of raw massing models were generated in Maya, in order to explore various primary design directions. A decision was reached very quickly to position the masses of the school along a North/South axis, with a set of classroom clusters strewn along the south border of the site. Making use of the natural topography of the site helped to avoid building under the water table, thus implementing the requested underground parking without the costs of building under the water table. The natural slope of the site was accentuated by positioning the tallest volumes (the dorm towers) on the highest point of the site. This model provided the possibility to implement the competition rules, site boundary restrictions and height restrictions to a pragmatic model in order to make sure to comply with FAR rules, local building code and the competition program. In a future step plans are in place to apply this modeling step with the aid of a parametric model in order to increase the speed up the time needed to check the model for its viability. While one group in the office was working on the raw massing, another team was working on images created with the AtnnGAN.

2.2. ATTNGAN

Generating images automatically based on natural language descriptions is a fundamental problem in many applications, such as in computer-aided design and art generation, and serves as a propellant in multimodal learning research and inference across vision and language, as evidenced by the increased research activities in this area of inquiry (Zhang, 2017). Attentional Generative Adversarial Network (AttnGAN) allow attention driven, multi-stage refinement for fine-grained text-to-image generation. Attention in Neural Networks imitates the way that humans are able to concentrate on particular aspects of their sensory input and blend out the rest around them. The architecture of the AttnGAN consists of two components. component one is an attentional generative network (Xu, 2018) that contains an attention mechanism that draws different sub-regions of the image. The focus is on words that are most relevant to the sub-region being drawn. The generative network uses initially a global sentence vector to generate a low-resolution image. Subsequently, it uses the image vector in each sub-region to query word vectors by using an attention layer to form a word-context vector. The regional image vector is then combined with the corresponding word-context vector to form a multimodal context vector. These form the basis on which the model generates new image features in the surrounding sub-regions, which results in higher resolution pictures with more details at each stage. Component two in the AttnGAN is a Deep Attentional Multimodal Similarity Model, (DAMSM). Using both the global sentence-level information and the fine-grained word-level information, the DAMSM can compute the similarity between the generated image and the sentence. Therefore, an additional fine-grained image-text matching loss for training the generator is provided by the DAMSM. Co-author Alexa Carlson translated the AttnGAN algorithm to operate with the COCO database. The sentences used with the AttnGAN Algorithm were a combination of sentences derived from the program of the buildings and descriptions of the activities in these buildings. For example:

- Multipurpose hall: It can be used as a theater to hold lectures and events.

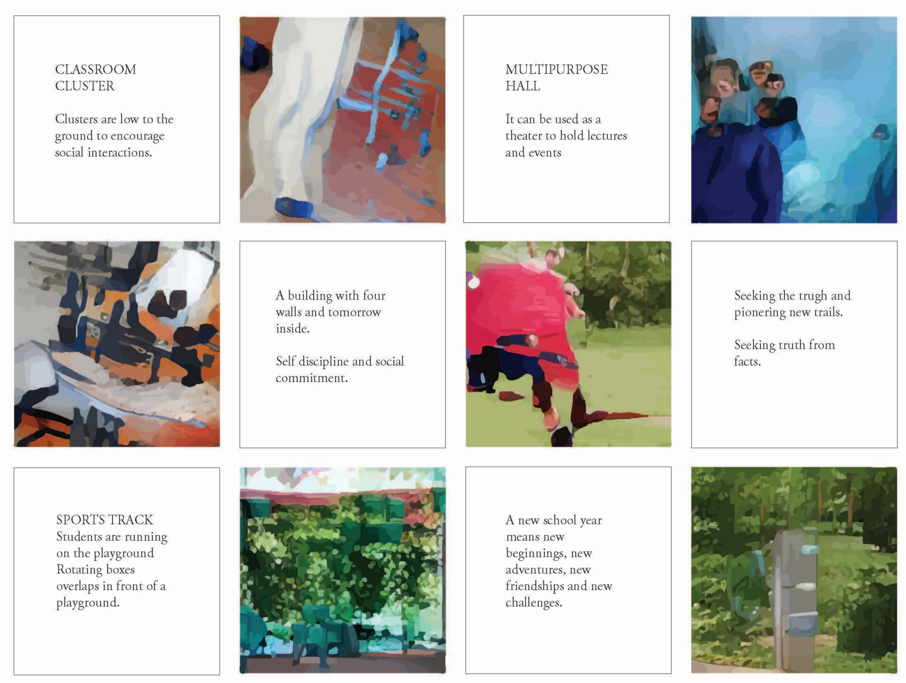

Figure 3. Examples of results of the AttnGAN.

2.3. GRASSHOPPER 3D

The results of this process can be seen in figure 3 these various images were used to inform the function and aesthetics of the buildings on the site. Using Grasshopper 3D the images were analyzed with the image mapper and subdivided into color patches. These color patches provided the blueprint to subdivide the larger 3D volumes that were created in the first round of modeling. An automated distribution pattern provided the vertical position and extrusion height of the volumes. Several passes of volume subdivision were employed in order to find the right balance between different scales and resolutions of the subdivision scheme. This step is certainly something that needs further exploration as it includes a portion of top-down design methodology, which contradicts the ambition to use an entirely emergent approach to design.

3. Discussion

3.1. CONTEXT OF THE DESIGN

The Northwest of Shenzhen is currently under rapid development and provides the larger population for this school project. The trapezoid shaped site has a longer side facing south and a short side facing north. The highest point of the site is in the Northwest corner. It falls from there towards south and east. The design team made an intentional decision to use the given topography of the site in the best possible way. It is best visible in the section (fig4) of the North South Axis:

Figure 4. Section .

The resulting design draws its qualities from the interplay of a well-balanced overall design, and the architectonic considerations put into each and every building. Although the building morphologies are distinctly different, and driven by its inherent programmatic needs, all of them together form a larger whole that provides students, faculty, administration and staff with a state-of-the-art school environment, that responds to the educational needs of this specific population. In addition, it provides a point of identification. Circling back to the conversation in the introduction of this paper it can be stated that this project complies with the semiotics of signification, which considers that any phenomenon that signifies anything can be considered part of the domain of semiotics, in doing so it covers a far larger territory than language alone. Thus the project demonstrates the ability of a Machine Learning processes to provide semiotic meaning to architecture through the use of language as design input.

Figure 5. Axonometry of Project.

4. Results

The use of AttnGAN (fig.6)as an architectural design technique is in its nascent state, thus it is difficult to position this project within a particular set of projects. Rare examples of computationally savvy architects testing the waters of AttnGAN exist, such as the works of Miloš Ilić (Silic, 2020) Much more common than the use in architecture is the application of AttnGAN in the arts (Amaya, 2020) and in fashion (Yi, 2019) though to describe them as common might be quite a stretch generally speaking. More successful would be an approach to contextualize the work along the lines of the use of language as a design environment. As a placeholder for a larger discussion on language and architecture, the authors will only mention Christopher Alexander - who is to this day notoriously unpopular with architects (Wright Steenson, 2017). This launches us directly back to the discussion on Semiotics in the introduction to this paper. Why is Semiotics of interest when discussing design methods based on Artificial Neural Networks? (For the record AttnGANs belong to the family of Artificial Neural Networks) Exactly because Semiotics can be considered a prime example of a theory of the artificial and thus occupies a central position in understanding phenomena of design. The differentiation between the analytic and the synthetic is akin to the liaison between design and semiotics, thus semiotics strives to talk about the environmental element of design which is so essential to narratives of architecture design based on symbolic cultures (signs, scripts, code, signifier, etc.). These conversations ultimately have to turn towards aesthetics, in order to conduct a survey of emerging milieus. In the frame of the conversation within this paper aesthetics is connotated (Berrio-Zapata et al, 2015) with the parallel mapping of the environment, the articulation of the environment, and the alternative nature of any design process - leading to the assumption that within Neural Architecture (fig.7), design and aesthetics take the same position and value. In conclusion, this also means that signs can be artificial, or rather that signs can be considered theories of constructed meaning, which includes design. There is a long history in the relationship of semiotics and design - not only in design methodologies but also in its pedagogy. This is for example evidenced in the way that Max Bill conducted his studios at Ulm (Betts, 1998), the pedagogical alternatives to the Beaux-Arts methodologies established in France and Italy post-1969 (Holt, 2017), the long-standing enamoration of Umberto Ecco with architecture and urbanism (Leach, 1997), Roland Barthes interrogation of fashion (Barthes, 2006), Anne Beyaer-Geslin's research on semiotics, design, and aesthetics (Beyaert-Geslin, 2012), and so on and so forth. The common denominator for a design theory gravitating around ideas emerging from Semiotics is the search for a language of design (Holt, 2017).

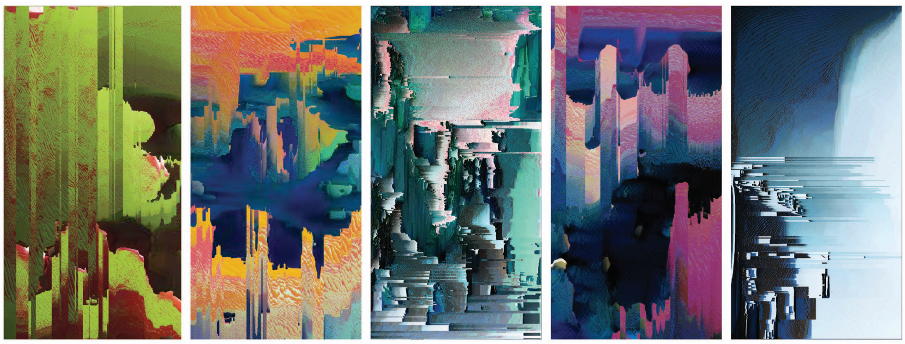

Figure 6. Result from the AttnGAN.

Figure 7. Translation into 3D models.

The results of the approach presented in this paper can be positioned exactly along this line of thinking. The search for a language of design based on the artificial creation of signs and signifiers. The major finding of this paper is -apart from the creation of a theoretical basis for the application of language in a NN driven design method - the proof of concept that this approach can produce viable architectural designs. The presented method still needs major work. For example: the design method primarily creates either exteriors or interiors, but not both at once and interdependently. Steps such as the initial massing modeling, the exact nature of the plans and details of the 3D models are introduced in a top-down method. It would be of interest to explore in further research the possibility to automatize more of the process, in order to interrogate the emergent properties of the artificial creation of signs. As stated above, Semiotics is a perfectly fitting thinking school in regard to its consideration of the artificial.

We started the paper with the story of Walther Gropius and his terrible secret: the inability to sketch anything. He overcame this problem by describing his ideas in a lingual fashion to others around him who converted his language into images. We presented in this paper a possibility to convert this terrible secret into a design weapon. One that oscillates between the interrogation of language, inquiries aspects of the artificial, and its role in contemporary design culture. One that creates images somewhere between abstraction and the surreal full of instances of estrangement and defamiliarization in an architecture that can be described as something different, alien, strange and wonderfully beautiful - maybe the first genuine 21st century architecture.

Figure 8. An example of Neural Architecture: Laboratory Building.

References

Amaya, N., Nieman, N., Prakash, R. and Rahman, F.: 2020, "Generating Album Artwork from Lyrics" . Available from <https://medium.com/@nickn9715/generating-album-artwork-fro

m-lyrics-699b7d57a92d> (accessed 13th December 2020).
Barthes, R.: 1968, *Elements of Semiology*, New York: Hill & Wang, New York.
Barthes, R. and Stafford, A.: 2006, *The Language of Fashion*, Berg, Oxford.
Berrio-Zapata, C., Moreira, F.M. and César Gonçalves, R.: 2015, Barthes' Rhetorical Machine: Mythology and Connotation in the Digital, *Bakhtiniana: Revista de Estudos do Discurso*, **10**(2), 135-157.
Betts, P.: 1998, Science, Semiotics and Society: The Ulm Hochschule Für Gestaltung in Retrospect., *Design Issues*, **14**(2), 67-82.
Beyaert-Geslin, A.: 2012, *Sémiotique du design*, Presses Universitaires de France, Paris.
Führ, E.: 1997, Werk Räume, *Wolke Magazin – Architektur Sprache*, **2.JG**(Heft 2), 56-67.
C. S. Hardwick and V. Welby Gregory (eds.): 1977, *Peirce, C. S., and Welby-Gregory, Victoria (Lady Welby), Semiotic and Significs: The Correspondence between C. S. Peirce and Victoria Lady Welby*, Indiana University Press, Bloomington and Indianapolis.
Holt, M.: 2017, Semiotics and design: Towards an aesthetics of the artificial, *The Design Journal*, **1**, 332-341.
Jäger, G., Zilian, L.S. and Hofer, C.: 2019, Crowdworking: working with or against the crowd?, *Proceedings of the J Econ Interact Coord 14*, 761–788 .
de Landa, M.: 2016, *Assemblage Theory*, Edinburgh University Press, Edinburgh, UK.
N. Leach (ed.): 1997, *Rethinking architecture : a reader in cultural theory*, Routledge, London.
Michela, D. and Zingale, S.: 2017, Semiotics in Design Education. Semiotics by Design,, *The Design Journal*, **20**(1), 1293-1303.
Miller, M.: 2016, Code Art Rooted in An Unlikely Era: The 1960's,, *Fast Company*, **Innovation by Design**, 43-56.
Reed, S., Akata, Z., Mohan, S., Tenka, S., Schiele, B. and Lee, H.: 2016, Learning what and where to draw., *ICML*.
Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B. and Lee, H.: 2016, Generative adversarial text-to-image synthesis., *ICML*.
Silic, M.: 2020, "AI-Architecture-AtnnGAN" . Available from <http://spacecollective.org/milosilic/9169/Ai-Architecture-AttnGAN> (accessed 13th December 2020).
Wright Steenson, M.: 2017, *Architectural Intelligence – How Designers and Architects Created the Digital Landscape*, MIT Press, Cambridge, Massachusetts.
Wigley, M.: 1998, Whatever Happened to Total Design?, *Harvard Design Magazine*, **2**(No.5), 52-63.
Xanthos, N.: 2012, Wittgenstein, *Signo*, **1**, 23-32.
Xu, T., Pengchuan, Z., Huang, Q., Zhang, H., Gan, Z., Huang, X. and He, X.: 2018, AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks, *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1316-1324.
YI, Z.: 2019, *Text-to-image Synthesis for Fashion Design*, Master's Thesis, KTH.
Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X. and Metaxas, D.: 2017, Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks, *ICCV*.