(IN)VISIBLE CITIES

What Generative Algorithms Tell Us about Our Collective Memory Schema

EDUARDO RICO CARRANZA[1]and SHENG-YANG HUANG[2] and JULIAN BESEMS[3]and WANQI GAO[4]
[1,2,4]Architectural Association School of Architecture
[1,2,3]Bartlett School of Architecture, University College London
[1]eduardoorico@aaschool.ac.uk, 0000-0002-1659-6166
[2]huang.sheng-yang@aaschool.ac.uk, 0000-0002-2748-5031
[3]j.besems.18@ucl.ac.uk, 0000-0003-2601-1173
[4]wanqi.gao@aaschool.ac.uk, 0000-0003-0717-0726

Abstract. The last decade has witnessed a turn in AI technologies working with differentiable neural network architectures learning the embedded functions between data points and performing generative operations synthesising unseen data. The move to a continuous and generative AI paradigm aligns with ideas in the field of cognition and psychology, where a growing body of authors are beginning to conceptualise memory and our representation of the past as a dynamic, malleable and ultimately generative field. So, how effective are generative algorithms in supporting and enabling this creative process of remembrance? To answer this research question, we propose an experiment on how the spatial movement and exploration of maps of real and imagined images can help our brain reconstruct its memories in a dynamic yet accurate manner. We develop an application allowing visitors to dynamically explore real and AI-generated images of a given site clustered by similarity in a virtual 3D space. Analysing visitor paths and observed images helps us understand visitors' perspectives on real and AI-generated data such as an increased preference for synthetic images by visitors familiarised with the site. We conclude with recommendations on how to approach visitor experience in generative AI-powered applications for engagement with historical and archival data.

Keywords. Collective Memory, Embedded Differentiable Functions, Latent Space, Spatial Cognition, StyleGAN2, Schema, Visitor Paths

1. Introduction

The past decade has seen a shift in AI technology from pattern recognition to creative and generative capabilities due to growing datasets and a focus on neural network architectures that learn differentiable functions. This allows for not only pattern detection and identification but also generation of new data "in-between" the information provided during training, providing a smooth flow through a continuously defined data field. The generative turn in AI has had a significant impact on creative work that involves producing novel data, such as synthesized images, 3D, texts, music, and short videos. This paradigm shift is reflected in other fields of cognitive science, where the concept of memory has moved from a fixed object/pattern to a differentiable/generative approach. Citizens are now seen as continuously constructing a collective schema of the city through their individual memories.

These parallel shifts in AI and memory can lead to the creation of new technologies for visualizing and describing memory. While AI has mainly been used for synthesizing new data, few studies have applied AI to exploring our perceptions of the past. To address this gap, we propose an experiment that creates an app for visitors to explore an image catalogue combining real and imagined data. We develop a framework for analysing visitor experience and test it with our tool. Our results show the relevance of this form of work and potential for further development in engaging with memory and data archiving. This article provides background and methods used, as well as results and discussion of main outcomes.

2. Generative Paradigm of Collective Memory

Collective memory is a term that bears a strong currency in design and planning disciplines since it is meant to encapsulate a common representation of the past of our environment that finds its way to planning and design in various manners. The understanding of the collective origin of our memory as well as the methods to conceptualise and "generate" it is, therefore an important task that practitioners have traditionally given attention to, which is currently under strong scrutiny as part of a wider debate on cities.

Collective memory is a widely used term derived from the concept of social hallucination conceived by French sociologist Maurice Halbwachs as distinct from individual memory (Halbwachs, 1992). As it needs to embody its ongoing connection to contemporary discourse and identity, it also differentiates itself from history and collective remembering for its constructive nature in cognition (Wertsch & Roediger III, 2008). The creation of urban space and the exploration of cognition by collective memory began in the era of neo-rationalism. Typology was used to explain the methodology between collective memory and the representation of urban space, which is the so-called architectural rationality. Aldo Rossi argues that architecture can be generalised in the accumulation of history into a variety of typologies with certain definite characteristics. Accordingly, he advocates the 'analogical thought' retrieving the "archaic, unexpressed, and practically inexpressible" thought in memory. Therefore, Rossi argues that the goal of architectural design is not self-expression, but that it should fit by creating that similarity The collective memory of the residents of the city.

Christine Boyer argues that urban representations are always mediated in perceived reality; they substitute for objective reality and do not imitate it (Boyer, 1996). So how is the generative force of urban representation forms formed? How does it work? The collective memory of the city is an ongoing process of construction.

Traditional paradigms of the definition of memories of cities were conceptualised as records of objective facts. Perceptual data were stored in the brain in the form of sequences of precepted representations. Memory was a set of fixed items, stored in a chronological manner that could be retrieved and represented accurately provided we had enough evidence stored. By the mid-twentieth century, however, there was ample scientific evidence that the form of memory was not a sequence of precepted representations but its characteristic patterns. Such beliefs gradually became widely accepted scientific beliefs in the ensuing half-century. On the other hand, Piaget's generative epistemology and its related schema construction theory established the explanatory foundation of collective memory in the rational practice of urban design. Schema is a mental model, a pattern of knowledge and experience, which is not only a cognitive structure existing in memory but also a construction scaffold on which memory and knowledge depend. According to Jean Piaget, a schema can be represented as a classification system that can organise, generalise, modify, and create object information. The object can be recognised by the subject only after the transformative processing by the subject's mental structure, and the degree of the subject's knowledge of the object depends entirely on what kind of cognitive schema the subject has. In this sense, the object structure is established by the subject. The cognition of the object also evolves with the development of the subject cognition schema, which becomes what Piaget calls the construction of the object. Therefore, the development of cognition relies on individual activities to trigger the interactions between the subject and the object, in which the dual construction of the subject and the object is carried out (Chelstowski, 1971; Piaget, 1970, 1971, 2003).

Collective memory is thus the fabricated output of a schema of urban representations based on urban objects. Such a simulacrum feature responds to Boyer's claimed differentiation between perceived representations and reality and rejects the Platonian form of cities.

In parallel to this paradigm shift, a similar transformation has been happening in computational urban studies, from symbolism based on a given schema toward connectionism based on data-driven, ever-developing schema construction. The field of AI has seen a strong development of generative algorithms, particularly in the field of computer vision and image synthesis. Many of these techniques compress the information contained in the image into a numeric representation in the shape of vectors that can be manipulated to give birth to new images by asking the algorithm to perform a reverse journey of data synthesis. The structure of these abstract representations, also known as latent spaces, is commonly referred to as differentiable, meaning that the computer is programmed to learn the smooth, continuous functions that connect, interpolate and hopefully extrapolate between the training data. Understanding the nature of this latent space, and how it relates to the final images produced, has been the subject of a substantial amount of work in research and visual arts, which has lately produced a wealth of image-generation algorithms which are now finding their way into commercial products.

Working more specifically with historical images or typological studies of architecture, projects such as Brutal Nature (Moullinex, 2020) allow the exploration of imagined brutalist architecture while for the generation of novel architectures represented in a 2D scattergraph produced by a dimensionality reduction technique

(Chen & Stouffs, 2021; Meng, 2021). These methodologies, such as t-SNE or other algorithms, help the representation and navigation of large image datasets, with implementations such as the PixPlot collection focusing on historical images (Duhaime et al., 2021). Working on a similar idea of spatialising data for navigation, Refik Anadol uses a t-SNE representation of the latent of GAN generated images of Gothic architectures (Anadol, 2019) to develop an explorer which brings up the image corresponding to the latent space linked to a particular 3D location of the camera. Similar attempts can also be seen in Immanuel Koh's 3D GAN Housing project. These cases all reveal the new, sub-symbolic, paradigm of spatial cognition and its construction (Koh, 2022).

3. Research Question

The exercises mentioned in the previous section focus strongly on the showcase of the creative capacity of these AI-powered algorithms, making an emphasis on the spatial qualities of the represented data and the emerging aesthetics in the images. There is less attention on the visitors themselves and the understanding of how these tools and the data produced are perceived as a form of documenting history or memory. There seems to be a gap in the research when studying how we can engage with this type of algorithm and how we react to the nature of the interpolated data in relationship with the real data. As a result, the following research question is formulated:

How do people observe datasets that combine AI-generated and real images of a given space? and how does their prior knowledge influence this experience?

Studying the engagement with these algorithms applied to historical datasets may help us understand how we approach a connectionist definition of the construction of heritage, memory and to archival material in general. The following sections describe an experiment where we begin setting up the methods of navigation of memory as well as forms of analysing our behaviour. We try to understand what the images that would typically call our attention are and what are how different visitors react to genuine and imagined AI-generated images.

4. Experiment Definition

Our experiment consists of the deployment of a 3D interactive app where visitors freely navigate and explore real and imagined pictures of a given place. By deploying the same app with different groups of participants who have different levels of familiarity with the site, we can develop comparisons that can help address the questions on the weight of prior knowledge on how we look at imagined and real images.

The space chosen is the Barrel Vault, in the Architectural Association School of Architecture, London. This is a narrow, elongated room with a distinct window system and a ceiling formed by a vault-like concrete and wood roof system (see Figure 1, left). The space has been traditionally used for pinups, exhibitions as well as teaching and lecturing and has a long tradition as a defining space within the school. The exploration of the images of the Barrel Vault, both real and imagined, is done through an app that is accessible through a link and is installed on the user's PC. The usage of the app is done on the computer screen, without the need for any particular tool and is carried out independently, without support or guidance from the research team. User instructions are briefly given via an introductory video supplied with the app link altogether with an explanation of the experiment as well as how data would be recorded and used. The interface records the movement of the visitor in this space as well as the index images being observed. This data is later processed to understand the visitor experience and, to an extent, their engagement with the images of the past.

An initial set of images of the Barrel Vault, which we shall denominate “real” images, come from several sources, both archival and current. We obtained 131 archive images from the AA historic database of school photographs, which contained photographs of events, presentations, performances and general school life dating back three decades. We then complemented them with 569 images obtained from 2 video shootings of the space carried out by the research team. The archive images are more likely to include a diversity of angles, themes and textures, while the ones coming from the video shootings provided a stronger quantity of information on spatial textures and overall structure. These “real” images consist of what we will call the “anchors” dataset.

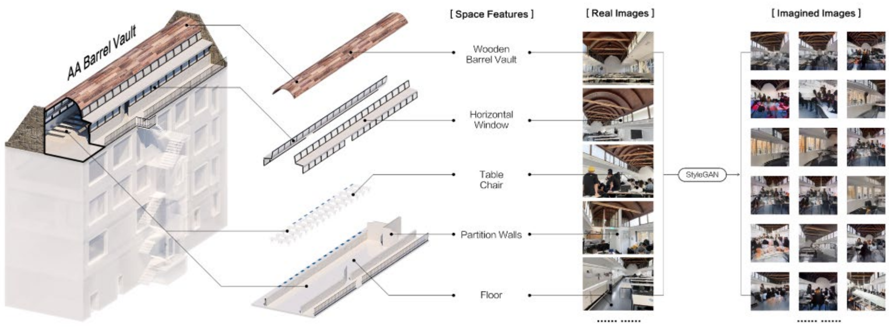

Figure 1. Site location in Barrel Vault, Architectural Association, London and real and generated images

The anchors are then fed to a StyleGAN2 algorithm (Karras et al., 2020) that is trained over 6000 Epochs and used to produce 20,000 “imaginary” pictures of the same space. Given the nature of the training, which in this case took place departing from an empty network, StyleGAN2 is likely to learn, and therefore replicate during inception, details and structures that are provided more often in the training dataset. These are likely to be some structural characteristics of the space, such as the ceiling, the repetitive nature of the roof beams and elongated windows. These appear in the majority of the images and are a distinctive feature of the room. The representation of people and activities as well as other details (objects, furniture or similar) are not so well detailed by the algorithm and would typically appear more blurred or unclear (Figure 1 top right). As a result, this second set of imaginary pictures is likely to have a distinct character when explored in detail, with a higher degree of vagueness or lack of definition.

The dataset composed of both "real images" (or anchors) and "imagined pictures" is then located in the virtual 3D space of the app that the visitor then can navigate freely (Figure 2 above) on their computer. The exact position of the images is estimated according to their visual similarity with other images forming clusters or coherent groups. This is done by feeding the images into an image classification neural network (VGG19) and extracting an intermediate abstract representation of these images as features. A dimensionality reduction algorithm is then run on these features with a t-SNE algorithm, which produces one 3D point in space per image. The algorithm generates the position of the points preserving similarity between images, hence clustering together images with common visual characteristics. When using the app, The visitor always enters the 3D virtual space in the same location within the 3D that contains all images (anchors and imagined). In the first instance, only the anchors (real images) are visible to the user. The user is led to focus more on the anchors close to the location of the camera thanks to a "mist" effect that fades distant objects in the background. At all moments, all images are rotated towards the camera, allowing the user to focus more on the immediate environment. The visitor can move through the environment using the mouse and keyboard and click on the existing anchors. By doing so, the app reveals all images in a nearing radius that "spawn" radially from the anchor and appear in their respective positions. The process can go on so that when clicking on any further image (anchor or imagined), more imaginary images are revealed (Figure 2 below). Clicking repeatedly on an image enlarges the radius until a certain maximum. The app records the path that is navigated by the visitor as well as the images clicked.

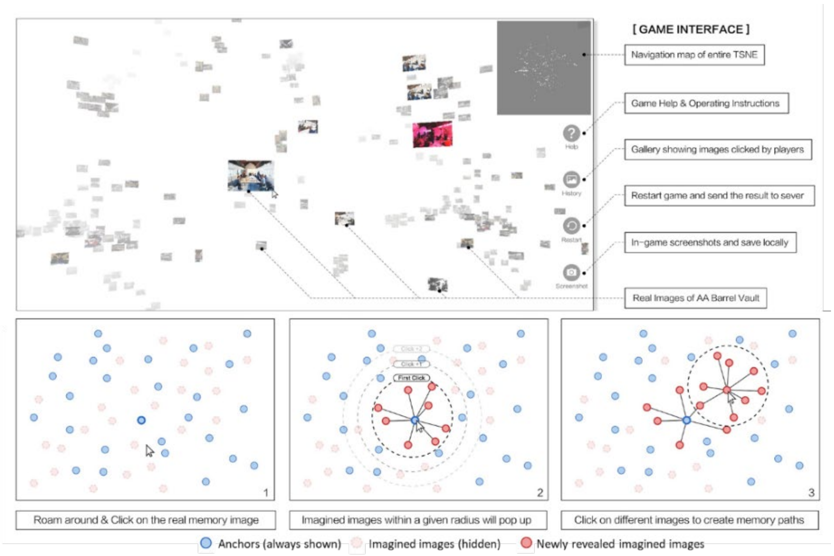

Figure 2. Site location in Barrel Vault, Architectural Association, London and real and generated images

The underlying assumption is that different knowledge and familiarity with the site being represented (Barrel Vault) may yield different forms of exploration. It is to be expected that people that already know the place may focus their attention on different types of images (real and imaginary) and that this can influence the ratio of the images visited. To evaluate these hypotheses data for each group previously mentioned is recorded separately and the results are used in a comparative study.

5. Experiment Results

The model was run a total of 59 times with 19 visitors familiar with the Barrel Vault site and 40 visitors that were not familiar. On each occasion, we gather data on the use of the tool, such as the time spent on the application, movement path and speed and images clicked for spawning. This last is used as a proxy for the attention given to a specific image. On average, the uses tended to be between one and 5 minutes, with some visitors taking much longer periods to evaluate the tool (Figure 3). We found cases where the tool was open for more than 2 hours, suggesting that several visitors were going through the exercise or that the tool was inactive for a large proportion of the time. These cases (three in total) were removed from the average analysis since they proved to be disorienting for the final results. These results seem to suggest a relatively good amount of time, especially considering that some of the visitors had no direct connection with the school or the research team. team.

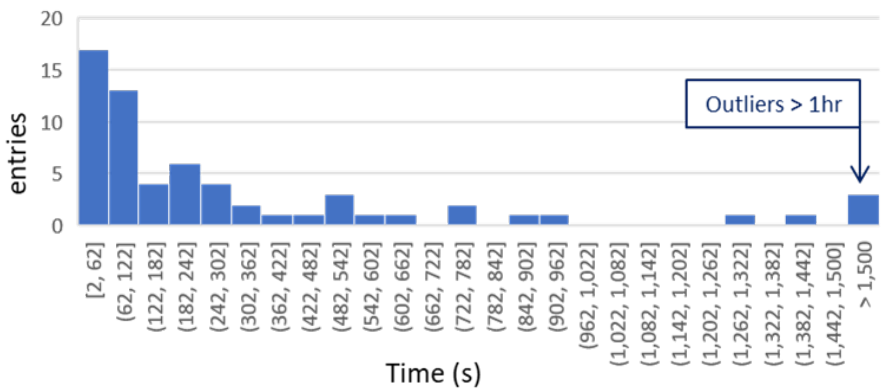

Figure 3. Distribution of total entries according to the time spent on the tool

Looking at a comparative analysis between visitor groups (Table 1 and 2), we saw that the average usage times for both groups are remarkably similar around 250 seconds, indicating a similar visitor engagement. An equal pattern can be observed with the number of images clicked, with both groups close to the aggregate average of 30 images clicked. On average, visitors click generated images more than twice as many as anchors (2.3 overall generated/real ratio). While there are many more generated images than anchors, only the last group is effectively shown when entering the interface. This means that the visitors kept unveiling newly synthesized images and explored them twice as often as they did with the real ones, hence suggesting a good degree of curiosity. The differences begin to appear when we look at the types of images clicked by these groups. The visitors that were familiar with the site did, on average 3.1 imaginary queries per real query, and those non-familiar with the site did substantially less (2.1). This is to be expected since people unfamiliar with the site are more likely to seek to familiarize themselves with the Barrel Vault before looking at the imaginary pictures that may appear harder to interpret since they have blurry or sketchy aspects to them due to their generative process. On the contrary, people that already know the place, entertain themselves longer by looking at interpolated or generated pictures of what they already know. Those familiar with the Barrel Vault which, by definition, will have more specific memory of the place, are likely to have a stronger engagement with abstract representations of the site. This could be linked to Piaget's schema construction mechanism. Strong prior knowledge of a location can lead to the enhancement of existing schemas or prompt the accommodation process to form updated spatial cognition through the introduction of heterogeneous visual stimuli from the generated images. If valid, Aldo Rossi's mental similarity hypothesis allows for this possibility to impact the collective memory.

		Average values across the group				
Metric	Valid entries	Total duration	Images clicked	Real images	Generated images	G/R
Unit	Number	Second	Number	Number	Number	Ratio
Total	59	250.7	30.1	9.0	21.0	2.3
Familiar	19	248.5	28.9	7.1	21.8	3.1
Unfamiliar	40	251.8	30.6	10.0	20.7	2.1

Table 1. Average visitor indicators per group.

		Variance across the group			
Metric	Valid entries	Total duration	Images clicked	Real images	Generated images
Unit	Number	N/A	N/A	N/A	N/A
Total	59	95574.0	3600.1	216.3	2227.4
Familiar	19	126582.5	1180.8	42.1	944.0
Unfamiliar	40	83709.5	4808.1	299.6	2876.4

Table 2. The variance of each indicator per group. Higher variance indicates less homogeneity in the observation.

We then turn to look at the images themselves and we produce an analysis of the images clicked by each visitor (Figure 4). We can see how some visitors click several times on a given anchor. Each time this happens, the app reveals a growing number of hidden images until these are exhausted. We can see how most of the clicks are repetitive as if the visitor was trying to “squeeze” as many images as possible around a given anchor before moving to the next. We equally carry out a study of the popularity of images by aggregating the number of clicks they attracted (Figure 5). Results from the exploration indicated that a larger proportion of the time spent by participants was exploring images with people and activities, rather than static objects. Equally, images with strong colours (exhibition or similar) disproportionally call the attention of the visitor when navigating through the t-SNE and are more frequently clicked. This also happens with images of people performing activities (pinups, designs or crits) which attract more attention than images of space.

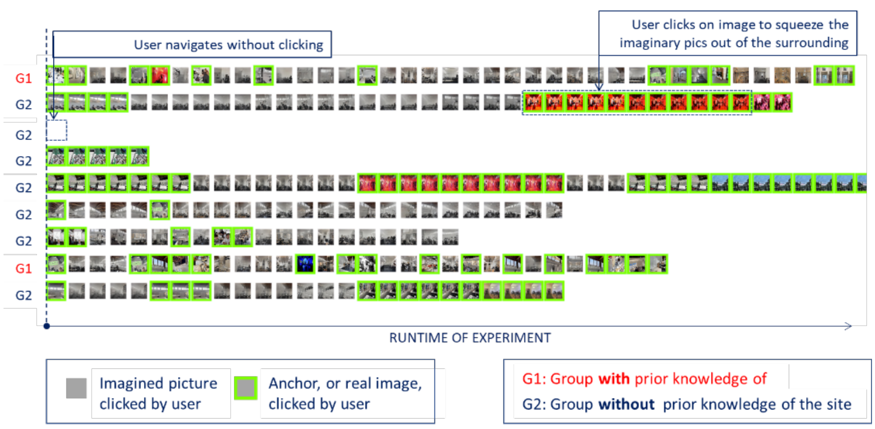

Figure 4. Examples of visitor exploration

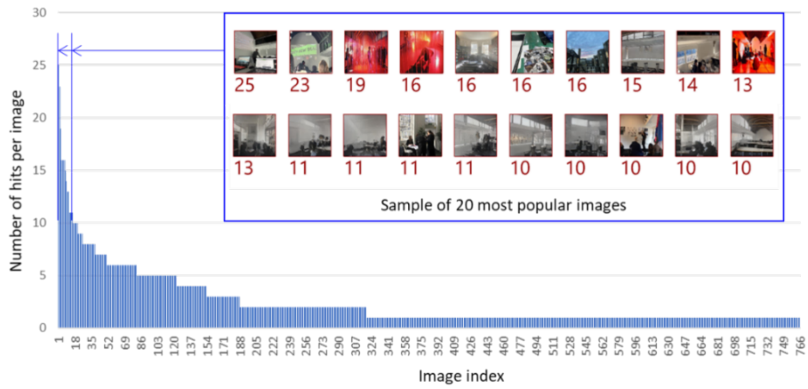

Figure 5. Popularity analysis and clicked figures

6. Conclusion

We have proposed an innovative form of using AI to study image-based archival material which stems from a continuist approach to coding as well as memory. We have tried to use algorithms that both organise images (t-SNE) as well as interpolate between them (StyleGAN2) and tried to study their deployment with visitors. Moreover, we have proposed methods of measuring the engagement of visitors with these tools and tried to extract conclusions on how the analytical data can relate to visitors' preferences and backgrounds. Our results indicate that those familiar with the site were more interested in the imagined pictures, but were more homogeneous in their use of the tool. This tendency may be linked to the behaviours of assimilation and accommodation in the collective schema-construction process of the group. This would suggest a different approach to questions of memory and representation, that, could be argued, relate to the terms of Piaget's schema. It reveals new tools for analysing the mechanisms by which the spatial cognitions of cities are collectively formed. While it is too soon to extract definitive conclusions, this experiment provides insight into how our relationship to images, memory and cities can be studied and approached via generative algorithms.

References

Anadol, R. (2019). *Refik Anadol | Architectural Photographic Memories. Parametric Architecture.* Retrieved December 6, 2022, from https://www.youtube.com/watch?v=yCC5GgakAeE.

Boyer, M. C. (1996). *The City of Collective Memory: Its Historical Imagery and Architectural Entertainments.* MIT Press.

Chelstowski, B. (1971). The Constructivistic Structuralism of J. Piaget. *Science and Society,* 35(4), 481-489.

Chen, J., & Stouffs, R. (2021). *From Exploration to Interpretation-Adopting Deep Representation Learning Models to Latent Space Interpretation of Architectural Design Alternatives,* CAADRIA 2021 (pp. 131-140). The Association for Computer-Aided Architectural Design Research in Asia (CAADRIA).

Duhaime, D., Leonard, P., Eskildsen, T., Choudhary, S., DeRose, C., Sanger, W., Arneja, M., Fish, E., Sorba, O., & Reagan, D. (2021). *PixPlot*. Yale Digital Humanities Lab. Retrieved 10 November from https://github.com/YaleDHLab/pix-plot

Halbwachs, M. (1992). *On Collective Memory.* University of Chicago Press.

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020). *Analyzing and Improving the Image Quality of StyleGAN, The IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 8110-8119). Institute of Electrical and Electronics Engineers

Koh, I. (2022). Architectural Plasticity: The Aesthetics of Neural Sampling. *Architectural Design,* 92(3), 86-93.

Meng, S. (2021). Exploring in the Latent Space of Design: A Method of Plausible Building Facades Images Generation, Properties Control and Model Explanation Base on Stylegan2. The International Conference on Computational Design and Robotic Fabrication (pp. 55-66). DigitalFUTURES.

Moullinex. (2020, 6 October 2020). *Brutal Nature.* https://www.facebook.com/moullinex/videos/342519493629041/?redirect=false

Piaget, J. (1970). *Genetic Epistemology.* Columbia University Press.

Piaget, J. (1971). *Biology and Knowledge: An Essay on the Relations between Organic Regulations and Cognitive Processes.* Edinburgh University Press.

Piaget, J. (2003). *The Psychology of Intelligence.* Routledge.

Wertsch, J. V., & Roediger III, H. L. (2008). Collective Memory: Conceptual Foundations and Theoretical Approaches. Memory, 16(3), 318-326.