Serlio and Artificial Intelligence: Problematizing the Image-to-Object Workflow

Jean Jaminet¹(✉), Gabriel Esquivel², and Shane Bugni²

¹ Kent State University, 800 E. Summit St., Kent, OH 44242, USA
jjaminet@kent.edu

² Texas A&M University, College Station, USA

Abstract. Virtual design production demands that information be increasingly encoded and decoded with image compression technologies. Since the Renaissance, the discourses of language and drawing and their actuation by the classical disciplinary treatise have been fundamental to the production of knowledge within the building arts. These early forms of data compression provoke reflection on theory and technology as critical counterparts to perception and imagination unique to the discipline of architecture. This research examines the illustrated expositions of Sebastiano Serlio through the lens of artificial intelligence (AI). The mimetic powers of technological data storage and retrieval and Serlio's coded operations of orthographic projection drawing disclose other aesthetic and formal logics for architecture and its image that exist outside human perception. Examination of aesthetic communication theory provides a conceptual dimension of how architecture and artificial intelligent systems integrate both analog and digital modes of information processing. Tools and methods are reconsidered to propose alternative AI workflows that complicate normative and predictable linear design processes. The operative model presented demonstrates how augmenting and interpreting layered generative adversarial networks drive an integrated parametric process of three-dimensionalization. Concluding remarks contemplate the role of human design agency within these emerging modes of creative digital production.

Keywords: Serlio · Artificial intelligence · Language · Design agency

1 Influence of the Disciplinary Treatise

The classical disciplinary treatises of the Renaissance have become a technical-literary genre that today are considered an essential part of the historical development of architecture. From the time of its publication to the present, no treatise has been more influential than Sebastiano Serlio's *Tutte l'opere d'architettura et prospetiva*. Serlio's ambition did not rely solely on producing an encyclopedic treatise composed of seven volumes. One of his most important objectives was to provide copious illustrations to the first five volumes and the seventh. Each of these volumes has, in fact, abundant architectural drawings, all large woodcuts, sometimes full pages, which were a great challenge and achievement in the art of printing for the period. Certainly, Serlio's architectural treatise

was the first to provide a visual dimension to the study of architecture in print, something never seen before and in such a forceful way. Introducing the visual culture of architecture via drawings and diagrams is perhaps the most important achievement of Serlio's treatise.

Architectural language is typically understood through its coded operations. Serlio was instrumental in developing this code through his canonization of the five orders. These codes, or rules, are laid out in the earlier volumes of his treatise, then applied in the later volumes. What is fascinating about Serlio's experiments is that in applying the codes, he proceeds to vigilantly deviate from them. The results are sometimes defined by the code, where the code and the product are isomorphic—that is, a one-to-one relationship exists between the plan and the section. However, at other times architectural elements are organized or misaligned, which suggests that a latent diagrammatic operation other than the code is at work. Architectural language is not the code; the language emerges when the code is scrambled. Thus, in Serlio we find entrenched the architectural code (transposition) within its analogical modulation (transfiguration). These insights into the discordant pairing of the analog and the digital suggest alternative theoretical parallels between brains and computer as well as emerging modes of creative production regarding advancements in machine learning.

2 Analogical and Digital Flux

Language (the possibility of communication) cannot be separated in distinct categories—we do not have one language that is analogical (pictorial and continuous) and another that is digital (coded and discrete). According to Gilles Deleuze, “From one point of view, we think of ... analog and digital, as two completely opposite determinations. But from another point of view, we could say that every digital language and every code is deeply embedded in an analogical flux” [1]. For Deleuze, language is defined by the discordant pairing of both analogical and digital modes of communication. A rudimentary understanding of current digital display technology may help to clarify this enigmatic concept. When digital signals are received by a display, they are continuously decoded as a field of light pulses displayed as discrete points of color (pixels). The signal remains coded, but the screen becomes responsive or modulates as the digital code is transplanted into the analogical flux of the pictorial image.

This modulation is where Deleuze locates the function of the diagram in painting and in the aesthetic act itself. “The diagram, the agent of analogical language, does not act as a code, but as a modulator” [2]. According to Deleuze, the intention of the diagram is to remove any predetermined “figurative givens” or predetermined resemblances that might be implied on the canvas or in the artist's mind. Thus, “The diagram is ... the operative set of asignifying and nonrepresentative lines and zone, line-strokes and color-patches” [2]. Deleuze locates Francis Bacon's work somewhere between abstract painting (cubism) and abstract expressionism (art informel). The code is prevalent in the former—geometric shapes imply figurative resemblances (optical space of representation)—while the latter is all diagram; the modulating power of the diagram becomes inert as it is deployed across the entire canvas (tactile space of line and color). Deleuze notes, “The manual diagram produces an irruption like a scrambled or cleaned zone,

which overturns the optical coordinate as well as the tactile connection” [2]. This scrambling—the continual conversion between analog and digital, figure and figuration, optic and haptic—is the domain of the diagram and that which constitutes the possibility of the aesthetic act.

3 Analog-to-Digital Information Processing

The human brain receives and processes information through both analog and digital means. Cognition is understood as an integrated analog-to-digital conversion process. This prevailing model of information processing gained credibility when neuroscientists in the 1980s demonstrated that neurons exhibit properties of both the analog and the digital. As Shores observes, “[I]f we consider neurons in terms of their being in a ‘firing or non-firing state,’ then we are examining their digital operation, but if we emphasize the ‘ongoing chemical processes’ of the brain, then we are looking at their analogical functioning” [3]. In other words, our primary means of acquiring and processing information is a continuous stream of sensory perception (analog); however, to store and retrieve this information, these experiences are encoded into discrete units (digital).

Computational theorists are now developing AI systems that integrate both analog and digital modes of information-processing. The primary task of such systems is to make a computer that better models human intelligence. However, designers are also interested in how these new technologies not only simulate reality but also become creative tools of production, particularly in regard to generative adversarial networks (GANs).

Every GAN has two neural networks—a generator and a discriminator. The generator synthesizes new sample images from random noise, while the discriminator samples from both the initial dataset (input images) and the generator’s output. The generator’s output is compared to the initial dataset by the discriminator to determine whether the synthesized image can be considered real or fake. As the generator receives feedback from the discriminator, it learns to synthesize more images better resembling the input images. In addition, progressive training can improve detail and resolution with each successive training.

The primary intention of these image-based neural networks is to synthesize artificial images that are indistinguishable from authentic images. However, GANs can also operate diagrammatically. In this sense, the GAN creates an exchange between continuous analogical modulation and codification of discrete digital units. Analog information (image input) and the digital information (noise) are synthesized by the discriminator then fed back into the system as new inputs. This process creates a continuous feedback loop transferring code into the analog pictorial flow of the image in each successive training. In other words, the GAN creates the possibility of continuous modulation of the analog and the digital through pictorial flux and transplanted code.

This is particularly the case with poorly-trained AI models that produce *artifacts*—effects or residues made visible by their diagrammatic scrambling. Bacon might call these “involuntary free marks” and Deleuze might describe them as “*asignifying traits* that are devoid of any illustrative or narrative function” [2]. Figure, ground, and contour begin to lose their coherence in the synthesized image, allowing “a form of a completely

different nature to emerge from the diagram” [2]. Although the common or crude purpose of GANs can produce figurative resemblance, its novel intelligence may unlock new avenues for design and creative production. The power of the GAN is not to mislead but to modulate.

4 Problematising the Image-to-Object Workflow

This research is part of a larger design project that investigates the illustrated treatises of Serlio in parallel with discussions about aesthetics and advancements in artificial intelligence.¹ The intention of these experiments is not simply to synthesize new images that simulate Serlio’s illustrations but rather to modulate their qualities and problematize their 2D to 3D translation beyond the rules of representation and orthographic projection.

Dataset Curation. The image data sets collected for our initial investigations were retrieved from the Avery Architectural and Fine Arts Library’s extensive online holdings of the works of Sebastiano Serlio. Avery’s Digital Serlio Project [4] includes full-page digital scans of multiple published and unpublished editions of all seven of Serlio’s manuscripts and his *Extraordinario Libro* and several subsequently published manuscripts of collected works. To create our datasets, pages from the manuscripts were downloaded from Avery’s repository. Images of individual objects were cropped from these pages in 1024 px × 1024 px format to accommodate various image-based machine learning platforms.

Dataset images were collected based on broad categories of the illustrated objects—columns, porticos, plans, and facades. Although Serlio’s treatise is organized based on tectonics (geometry, perspective, orders) and typology (monuments, churches, domestic buildings), we chose instead to curate our datasets by object type. Since GANs require input based on superficial likeness between images, our datasets exploit the self-same repetition inherent to Serlio’s drawing tectonics. These broad groupings of drawings were necessary to establish a dialog between the coding of classical objects and the analog-to-digital modulation of image-based neural networks. The intention of this dataset curation and subsequent 2D and 3D experimentation was to explore the capacity of the image and its qualities to suggest alternative ideas about materiality and logics of assembly beyond the techniques of orthographic projection and its related narratives of language and representation.

Layered Generative Adversarial Networks. Following image curation, experiments were conducted to train the Serlio datasets using various GAN platforms. Again, the primary purpose of image-based GANs is to synthesize artificial images that maintain fidelity to the dataset. However, as a productive design tool, we were more interested in the latent image qualities that became evident during the training process.

¹ All design research was conducted during the 2021 spring semester at Texas A&M University College of Architecture under the instruction of Jean Jaminet and Gabriel Esquivel and with the assistance of Shane Bugni. Student contributors include Brenden Bjerke, Erin Carter, Nate Gonzalez, Kamryn Massey, Ana Rico, Luis Sanabria, John Scott, Dalton Turpin, Austin White, and Spenser Young.

In our initial styleGAN experiments, the Serlio datasets (input images) were trained against a pretrained model (generator input) by the discriminator. We chose pretrained models (generic datasets of faces, buildings, landscapes, etc.) for the generator input instead of random noise to expedite the training results. In cases where styleGAN platforms required a domain of images to feed into the neural network, other GANs—like sinGAN and style transfer—that only require a single image were deployed. The styleGAN trainings predominantly generated various image distortions. Conversely, sinGAN outputs seemed to break down individual images into smaller fragments. Style transfer output images were used later in the process as texture maps to enhance details of 3D objects.

Distortions that were produced reveal other shapes, profiles, and postures of the objects that move the image away from its original resemblance and semantic content. For example, a regulated facade becomes a cascading field of apertures, or a single arched opening becomes a winding surface of figural voids; both produce estrangement in silhouette and scale. These distortions are based on image values (color, contrast, saturation, etc.) rather than formal complexity and linguistic articulation (line, edge, plane, volume, etc.). Fragmentation, on the other hand, concerns pulling the image apart into components that are detached or incomplete. For example, incongruities in patterning of a segmented masonry arch become reassembled, suggesting a tectonic that is foreign to the initial construction system. This fragmentation of the Serlio objects reveals alternative logics of assembly beyond the classical tectonics suggested by their orthographic projection. These distortions and fragmentations are inherent to the way machine learning interprets data through its analog-to-digital conversion process. Of equal importance is how designers guide machine learning processes through direct manipulation of the code and visual interpretation of the output (Fig. 1).

Fig. 1. Fragmented pediment produced by sinGAN (left) and style transfer image projected onto morphed fragment (right). Image by Spencer Young.

Throughout the training processes, the degree of distortion and fragmentation can be controlled by adjusting various parameters of the GAN, including *truncation value* and *scale of manipulation*. The adjustment of these parameters allows designers to claim

agency in the machine learning process to control the fidelity to the initial input datasets and the uniqueness of the output. Lower truncation values synthesize images that are more self-similar to the initial dataset, whereas higher values can produce results that deviate significantly. Likewise, adjusting the scale of manipulation adjusts the size of the random noise fed into the GAN. Scaling up decreases the size of the noise and creates highly articulated images (changes the detail but not the figure of the image). Scaling down increases the size of the noise and creates chunky or fragmented images (reduces detail; however, creates figuration). The intention of these methods was not to synthesize columns, facades, or doorways that would be indistinguishable from Serlio's illustrations but to use the digital code of the GAN as a substitute for the architectural or orthographic code to scramble the analog pictorial signal of the image.

Integrated Parametric Three-Dimensionalization. Industry-standard 3D modeling software platforms use mathematical coordinate-based representations to simulate surface and mesh geometry. These models begin as primitive shapes (curve, cube, sphere, etc.) that gain complexity by augmenting the component geometry. Geometry is manually created and manipulated through direct access to its points, curves, surfaces, and polygons displayed as orthographic projection drawings in multiple default windows of the user interface. These platforms give designers the power to create any 3D object by drawing on skills of visual and tactile acuity and knowledge of conventional representational systems.

Other 3D modeling and animation software applications utilize procedural generation tools rather than coordinate-based geometry. *Procedural generation* “is a method of generating data or content algorithmically as opposed to manually that combines human-generated assets and algorithms with computer-generated randomness and processing power” [5]. The gaming industry uses procedural generation tools to create open world games that generate environments in real time, providing an immersive experience of limitless variation based on choices made by the player. Likewise, these powerful tools are utilized in movie editing to create fantasy landscapes and crowd swarms with unprecedented randomness [6].

We utilized these procedural generation tools to proceed with our three-dimensionalization of the GAN outputs. These tools can interpret GAN outputs through voxelization of the image content. Voxelization is similar to creating a heightmap that uses image qualities (color, black and white values, etc.) to determine the 3D coordinates of the 2D image data. However, whereas heightmaps extrude pixel information only along one coordinate axis, voxels can be projected in all directions, thereby creating more complex spatial geometry. It was important that our 3D process follow a similar logic of neural network image generation, wherein the computer procedurally generates the 3D model based on parametric constraints available in the code. However, direct access to the 3D model geometry is limited; instead, design agency is asserted by adjusting the parameters of the script. To accompany this more conceptual description, one example is outlined below that illustrates this image-to-object workflow in more technical detail.

5 Operative Model: Portico

One of Serlio's major contributions to architectural discourse is his canonization of the classical orders. These classical elements are exhibited and analyzed as fragments in his early volumes, then deployed in larger building configurations. Nowhere in his treatise is this mereology more evident than in the *Extraordinario Libro*, where Serlio creates prototype doorways (porticos) in which classical components (columns, arches, pediments, etc.) are deployed and combined with unprecedented variation. These part-to-whole relationships become the speculation of this operative AI model. Neural networks and procedural generation tools are used to speculate about latent assembly logics within the portico images that exist outside classical tectonics and traditional orthographic representation.

In this exploration, the portico undergoes a series of estrangements (neural network image generation), reassembly (procedural generation modeling), and stylization (texture mapping). First, our curated dataset of Serlio's porticos is processed by a styleGAN, the output of which inherits multiple features of individual porticos (image training). Due to the similarity between input images, the most productive synthesized images exhibit distortions of the architectural component (details), while the overall figure (silhouette) remain largely unchanged. Based on these characteristics, a single frame from the resultant latent walk was selected as the input for a succeeding sinGAN training. The intention of introducing the sinGAN was to fragment the portico image into multiple parts (fragmentation) (Fig. 2).

Fig. 2. Image-to-object workflow: estrangement (neural network image generation), reassembly (procedural generation modeling), and stylization (texture mapping). Image by Nate Gonzalez.

The sinGAN selects patches of elements it can recognize and repeat to create “diverse samples that carry the same visual content as the [input] image” [7]. This process can be useful in generating images that simulate the randomness in natural environments, cityscapes, and flocks and swarms. However, when trained on an image that depicts a singular object, the sinGAN produces unexpected configurations that have the effect of fragmenting rather than unifying the image. With the portico, the resulting fragmentation creates alternative parts loosely based on the textures of the illustration. The assembly logic of the synthesized image is based on image construction rather than drawing tectonics.

The subsequent 3D reassembly process combines both procedural generation tools and conventional 3D modeling. The fragmented images are voxelized by assigning depth

values to the range of black and white pixels (fragment three-dimensionalization). Two voxelized fragments are morphed to produce an object that inherits 3D information from both. This merged fragment is then combined with a 3D model of a recognizable portico fragment (fragment morphing). This novel morphological assembly (object output) simultaneously exhibits the figuration of voxelized fragments (scrambled digital code) and partial reconstruction of the familiar language of architecture (pictorial analog signal). An alternative assembly logic emerges that is a matter of data processing rather than tectonic joinery or orthographic projection.

The final step in the reassembly process deals with creating additional levels of detail. Utilizing a neural style transfer, the UV map of the morphological assembly is blended with a Serlio detail image then remapped to the object (style transfer). Detail is assigned rather than constructed through tectonic joinery (image mapping). This recursive process of estrangement, assembly, and stylization allows an alternative assembly logic to emerge that becomes a matter of image construction and data processing rather than tectonic joinery or orthographic projection.

5.1 Intelligence Beyond Serlio

Beyond the intentions to reinterpret historical artifacts and deploy novel design technologies, the focus of this research is to address the transformative potential of AI in architecture. Artificial intelligence is not simply a toolset to optimize building elements but rather emphasizes architecture's ability to serve as a cultural marker. Mario Carpo observes that in failing to recognize opportunities to expand architectural intelligence through technology, "the design professions seem to have flatly rejected a techno-cultural development that would weaken (or, in fact, recast) some of their traditional authorial privileges" [8]. These assertions radically challenge computational methodologies as tools of expediency and efficiency and more importantly embrace the possibility of using them as strategies of communication between the human mindset and alien intelligence. Serlio's treatise on architecture, although inherently analog, anticipates these contemporary technological circumstances. By cataloging a considerable variety of buildings, details, and typologies in a uniform manner, Serlio, in a sense, created his own data set 400 years before IBM coined the term. The ever-present problem of agency within the discipline necessitates a revisitation of these manuals through a hyper-digital lens.

By surrendering established roles of authorship, alternative design agency in the machine learning process is acquired through visual selection and interpretation, thus fostering a shared ownership between designer and machine. The machine is tasked with indiscriminately processing content through iteration, replication, and complication of data through feedback loops in the computational process. As such, human intervention is required at the level of mundane digital tasks and post-processing image manipulation. As Abrons and Fure note, "Designers can take up this challenge by critically considering the digital processes we take for granted, such as default render settings, photoshop filters, geometric primitives, 'pan' and 'zoom,' extrude commands, and so forth" [9]. These discerning maneuvers can be further integrated in varying degrees at different stages, rather than coming at the end of a more linear design process. Seemingly trivial operations become the primary form of mediation between human perception and machine learning.

Furthermore, the paradigm of drawing has undergone radical change since Serlio and no longer provides a stable reference for the discipline of architecture. What we think are drawings are actually pictures of drawings or simulations of lines on a digital interface. “Images are inherently dynamic, and our tendency to think of them as static or fixed is a result of the psychohistorical residue of drawings...” [10]. Likewise, the facility with which these images can be manipulated suggests that the drawing no longer constitutes an original act of creation. Problematizing the image-to-object workflow through image-based neural networks and procedural generation 3D modeling contests the hegemony of traditional drawing tectonics and assembly logics associated with orthography. These synthesized images and objects are fragmentary, which is a characteristic of the latent diagrammatic operation in Serlio’s drawings. However, this kind of machine learning can rework the way we present, learn, and teach architecture because it scrambles the orthographic codes or conventions that have defined architectural language since the Renaissance and have persisted through pedagogies established by the *École des Beaux-Arts* and the Bauhaus. The continuous modulation of analog and digital information processing defies linear design processes and dialectical translations from drawing to building—signaling a shift away from modern and postmodern notions of consistency, semantics, and representation toward a new paradigm of medium, communication, and agency—thereby creating the possibility for new languages to emerge.

Challenges to architecture in the twenty-first century demand a historical reflection on theory and technology as critical counterparts of architecture’s intelligence, particularly in regard to visual/spatial acuity unique to the discipline. Serlio’s illustrated exposition serves as a conduit to initiate these discussions about contemporary aesthetic communication and shared design agency that may allow architecture to gain disciplinary perspective on our technological circumstances and stimulate new modes of perception and creative digital production.

References

1. Deleuze, G.: Painting and the question of concepts/05. Available via The Deleuze Seminars (1981). <https://deleuze.cla.purdue.edu/seminars/painting-and-question-concepts/lecture-05>. Accessed 14 Mar 2021
2. Deleuze, G.: *Francis Bacon: the logic of sensation*, Smith D (trans) Continuum New York (2003)
3. Shores, C.: Deleuze’s analog and digital communication. Available via Pirates and Revolutionaries (2009). <http://piratesandrevolutionaries.blogspot.com/2009/01/deleuzes-analog-and-digital.html>. Accessed 14 Mar 2021
4. Avery Architectural and Fine Arts Library: Digital Serlio project. Available via Columbia University Libraries (2018). <https://library.columbia.edu/libraries/avery/digitalserio.html>. Accessed 15 Mar 2021
5. Wikipedia. s.v. Procedural generation. Available via Wikipedia. https://en.wikipedia.org/wiki/Procedural_generation. Accessed 14 Apr 2021
6. Sarawagi, S.: Procedural generation—a comprehensive guide put in simple words (2021). Available via 8bitmen. <https://www.8bitmen.com/procedural-generation-a-comprehensive-guide-in-simple-words/>. Accessed 13 Apr 2021
7. Shaham, T.R., Dekel, T., Michaeli, T.: SinGAN: learning a generative model from a single natural image (2019). [arXiv:1905.01164](https://arxiv.org/abs/1905.01164) [cs.CV]

8. Carpo, M.: The Second Digital Turn: Design Beyond Intelligence. MIT Press, Cambridge (2017)
9. Abrons, E., Fure, A.: Post-digital materiality. In: Borden, G.P., Meredith, M. (eds.) *Lineament: Material, Representation and the Physical Figure in Architectural Production*. Routledge, New York (2017)
10. May, J.: *Signal, Image, Architecture*. Columbia University Press, New York (2019)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.