Latent Space: Granularity and Architectural Imaging/Imagining

Jono Coles

The contemporary architect translates meaning through digital images, operating (often unknowingly) at the granular level of the individual electric signal. Through the latent automation under the hoods of our software, vast sedimentary information is manipulated and perceived not at the particulate scale of the pixel but on the comparatively boulder-like size of the signifier.[1] There is an inescapable discrepancy between the granular level at which designers assemble architectural images and the level at which the images convey architectural meaning. This discrepancy exists at the border between human and digital cognition, and it has problematized architectural production by obscuring the fundamentals of architectural thought.

The concept of the atom is fundamental. The earliest recovered definition—a discrete unit of matter of which all things are formed—is derived from ancient writings on physics and was popularized in Enlightenment thinking by John Dalton.[2] However, the atomic theory has come to transcend scientific thought; the unit now defines the approach of disciplines of representational media. Impressionism, for instance, was founded on the brushstroke as a unit, with figures, landscapes, objects, and spaces loosely implied by collections of them. Collages are amalgamations of several discrete images, focusing instead on the relationships between them.[3] While words are composed of characters, writing is assembled and perceived based on individual readings. In the case of digital computation, the individual electric signal is its founding component, the finest unit of sediment.[4]

Architects and theorists contest architecture's basic unit of measurement. The dichotomy between the building and its representations, established by Italian Renaissance author Leon Battista Alberti, is founded upon a well-challenged Vitruvian platitude: buildings are assemblages of components.[5] The corollary to Alberti's sixteen-thousand-year-old reference is that architecture is the graphic assembly not of the components of a building themselves but rather the lines on paper that signify them—the architectural medium is the image of the boulder, not the boulder itself. The cognition required to meaningfully produce these images of boulders is currently considered architectural *thought*.

Moving 500 years from Alberti to the present, where the computer is the site of architectural labor, those same signifiers are now pixelated. The contemporary architect no longer works with pen and paper to produce signifiers of an architectural assembly but indirectly manipulates latent automated processes of vast sedimentary information to create an image that flickers between recognizable and microscopic. Digital and physical craft are similar;[6] architects still play with boulders but now, in doing so, they are simultaneously playing with grains of sand.

[1] John May, Signal. *Image. Architecture: Everything Is Already an Image* (New York: Columbia Books on Architecture and the City, 2019).

[2] Bernard Pullman, *The Atom in the History of Human Thought* (New York: Oxford University Press, 1998).

[3] Margaret Samu, "Impressionism: Art and Modernity," The Metropolitan Museum of Art, October 2004, http://www.metmuseum.org/toah/hd/imml/hd_imml.htm.

[4] Linda Null, *The Essentials of Computer Organization and Architecture,* Fifth edition (Burlington, Massachusetts: Jones & Bartlett Learning, 2019).

[5] Leon Battista Alberti, *De re Aedificatoria,* (1452).

[6] Malcolm McCullough, *Abstracting Craft: The Practiced Digital Hand,* (Cambridge, Mass.: MIT Press, 1998).

While architects do think and act on the scale of the atom or pixel, this thinking only aims to develop particular qualities of a larger comprising component. Various chemical admixtures physically manipulate concrete components giving them discrete finishes, densities, pouring techniques, and formal requirements. Similarly, central to many digital rendering techniques are bump and displacement mapping, in which collections of pixels are extracted from scanned images and (often manually) reinterpreted in grayscale to overlay depth onto their flat parent texture.[7] This technique is a low-data solution for encoding detailed formal qualities to otherwise crude polygonal surfaces, initially developed for the video gaming industry. In both chemical admixing and displacement mapping, architectural form is mediated through processes that are both literally and metaphorically akin to atomic interactions. The important distinction between both cases is the intention behind these granular operations to construct and describe, respectively, qualities of the architectural component rather than to act as the component itself.

Architectural translation occurring at the level of the component/signifier is due to the human incapacity to directly comprehend formations of material components on the granularity of the atom. The primary motivation for the development of computational technology was to provide humans with the indirect ability to exceed these limits of numerical cognition, expand the possible index size, and sort the humanly unsortable. The aforementioned scalar discrepancy between the sedimentary assembly of the architectural image and the amalgams that convey its meaning lies along the gap between human cognition and digital cognition.

In an attempt to alleviate this cognitive gap, recent digital cognition developments identify and reconstruct visual signifiers from pixels. Generative diffusion—more commonly and incorrectly referred to as "AI imaging"—contributes to the alleviation, but it has been unsuccessful in mimicking architectural thought (so far). Built upon a synthesis of several image-based machine learning models developed over the past few years, generative diffusion technology has most prominently emerged as a commercial service that creates an image to match a user-given text prompt. Subsequently receiving massive popularity, usage, and speculation, these controversial services have threatened image-based laborers due to their remarkable efficiency in simulating qualitative cognition through incalculably rigorous quantitative operations.

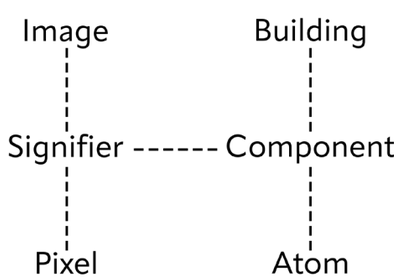


Figure 01. A diagram of the Image-Building Relationship. Courtesy of the Author.

[7] Kimmo Karhu, "Displacement Mapping," Tik-111.500 Seminar on Computer Graphics (Helsinki University of Technology, April 2002), https://www.academia.edu/2696670/Displacement_Mapping.

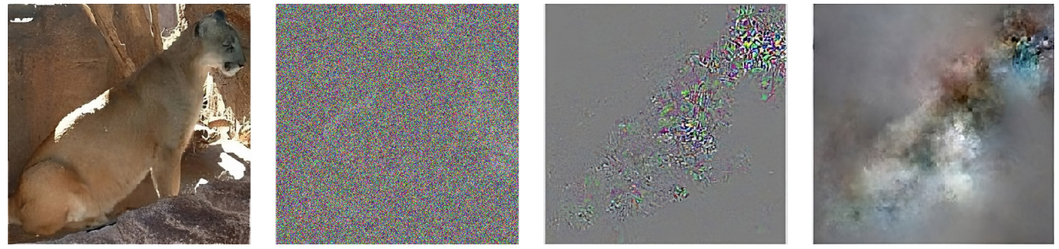


Figure 02. An image and three methods of analysis: L^{∞} metric, L^{2} metric, and decorrelated space. (Olah, Mordvintsev, and Schubert, "Feature Visualization.")[8]

Under the hoods of generative diffusion software are several stages of analytical algorithms that search, study, and index sets of billions of images (and non-consensually include the work of many digitally-native artists). Each image in the set is algorithmically attributed a value on a scale of its relation to the signifiers present in that image; an image can be paired with thousands of scalar values that abstractly describe its graphic contents. These sets of images and paired scalar values are called *latent* spaces. When prompted with text on the user-end, generative diffusion sifts through latent space to find images with related scalar values, assesses their pixels for relevant patterns, and iteratively hones a field of randomly generated pixels to mimic the patterns they identified.[8] When prompted for a specific boulder, it will study billions and systematically construct the requested representation grains at a time. It is a near-atomic metamorphosis.

For all its indexical might, commercial generative diffusion is currently incapable of architectural thought. If prompted to mimic an architectural representational medium—like a plan or detail, for example—the returned array of electric signals is just that, a meaningless index of values assembled from the unitary deconstruction of another index of values. Because of the level of detail at which the technology operates, and the mediation of signifiers through language, *it is unable to recognize the architectural meaning of the components it has graphically assessed, decomposed, and reassembled.* Several related machine learning technologies currently under development specifically attempt to study and create spatial conditions; notably, the most successful have transcended the pixel as their fundamental operative unit.[9,10,11] The current failure of generative diffusion to replace the architect can be attributed to the continued discrepancy in the scale upon which its cognition operates. In order to serve as a replacement for architectural thought, it must understand the meanings of architectural components.

[8] Chris Olah, Alexander Mordvintsev, and Ludwig Schubert, "Feature Visualization," Distill 2, no. 11 (November 7, 2017): 10.23915/distill.00007, https://doi.org/10.23915/distill.00007.

[9] Stanislas Chaillou, "AI + Architecture | Towards a New Approach" (Harvard University, 2019).

[10] Ben Mildenhall et al., "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis," 2020, https://doi.org/10.48550/ARXIV.2003.08934.

[11] Kyle Steinfeld, "Drawn Together," in Distributed Proximities: Proceedings of the 40th Annual Conference of the Association for Computer Aided Design in Architecture, ed. Brian Slocum, vol. 1 (Association for Computer Aided Design in Architecture, Philadelphia: Association for Computer Aided Design in Architecture (ACADIA), 2021).

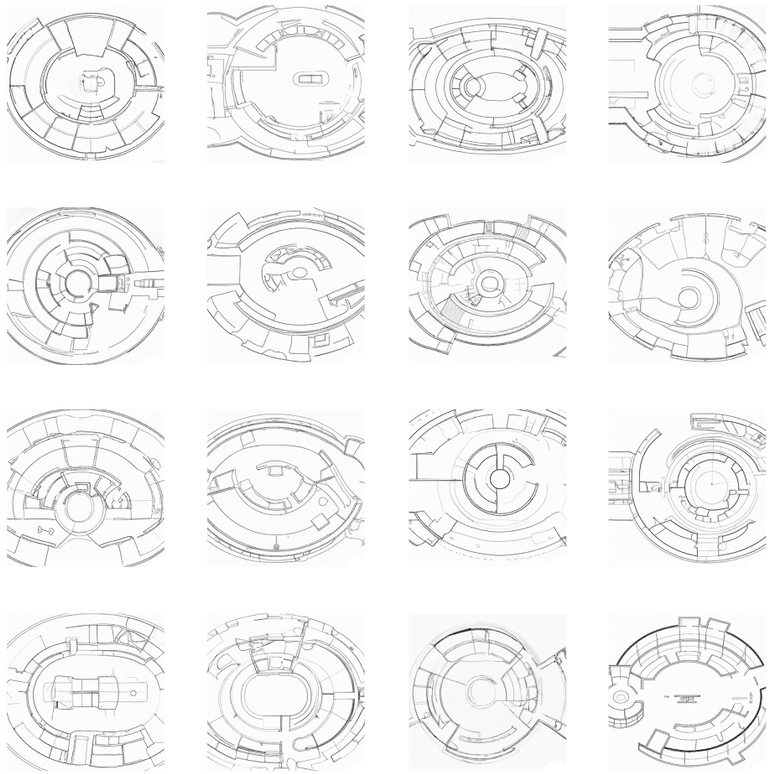


Figure 03: The output of "an architectural plan of an ancient Roman basilica with a circular apse" from DALL-E. Clearly failing to understand the concept of a room, door, window, or any architectural component at all, these images exemplify the limit of DALL-E's cognition: the composition of pixels. Courtesy of the author.

Classifying architectural thought as it relates to the limits of digital cognition opens new lines of inquiry into the contemporary architectural process. Taking a clear position on the level of detail on which a project operates, both digitally and physically, allows engagement with history, context, labor, and materiality through direct, fundamental operations. The remainder of this essay describes three projects within the context of these inquiries: 15 Clerkenwell Close by Groupwork, Ensamble Studio's Musical Studies Center, and 35 Green Corner Building by Studio Anne Holtrop. Each exhibits a unique approach to the architectural unit and its relationship to digitality, framing themselves as possibilities for transcendence from the digital image as the site of architectural action.

Groupwork's 15 Clerkenwell Close directly engages with the material history of its North London context. On the site of a former 11th-century Norman Abbey, the present-day neighborhood of Clerkenwell contains only fragments of the street lines and limestone facades that formed its urban boundary nearly a thousand years ago. Material research led Groupwork to an active French quarry mining the same limestone vein, which likely supplied the

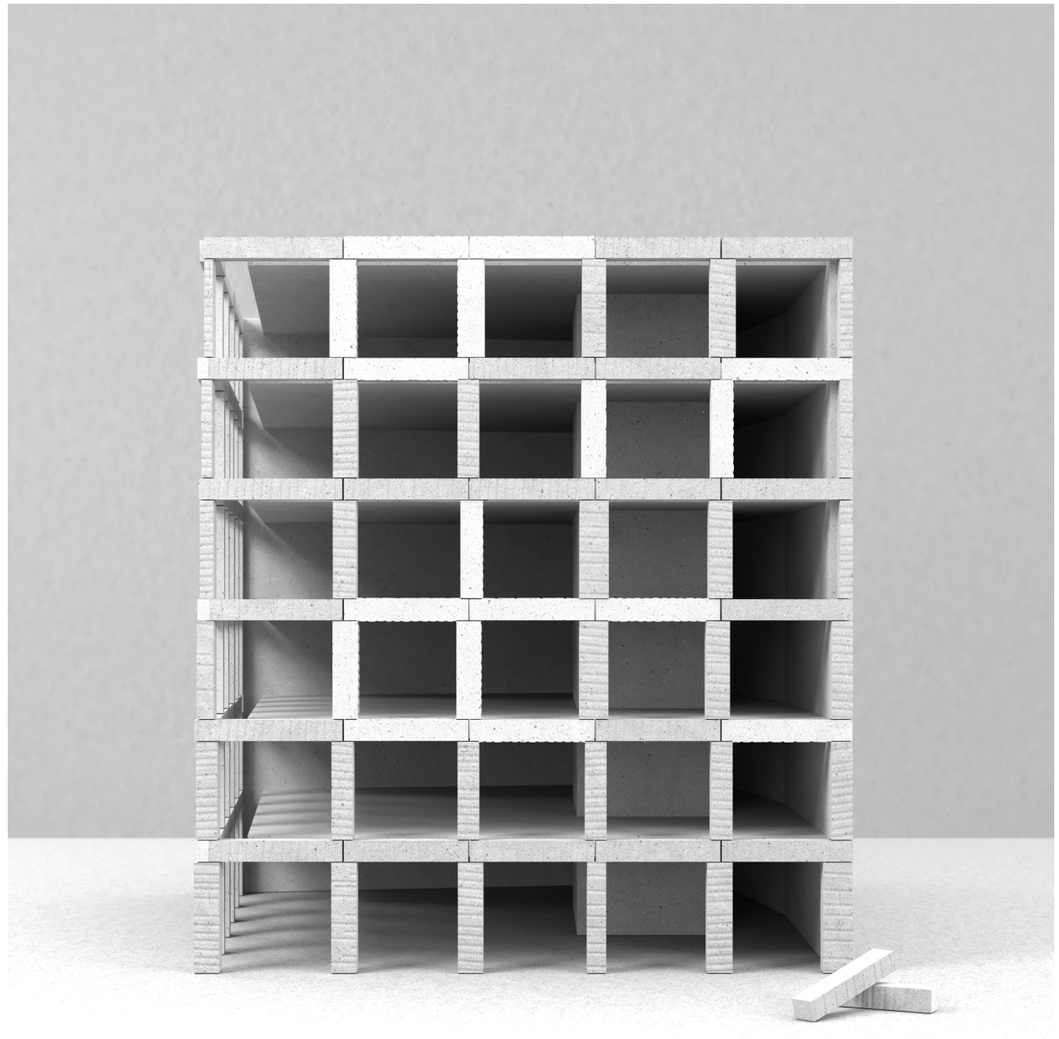
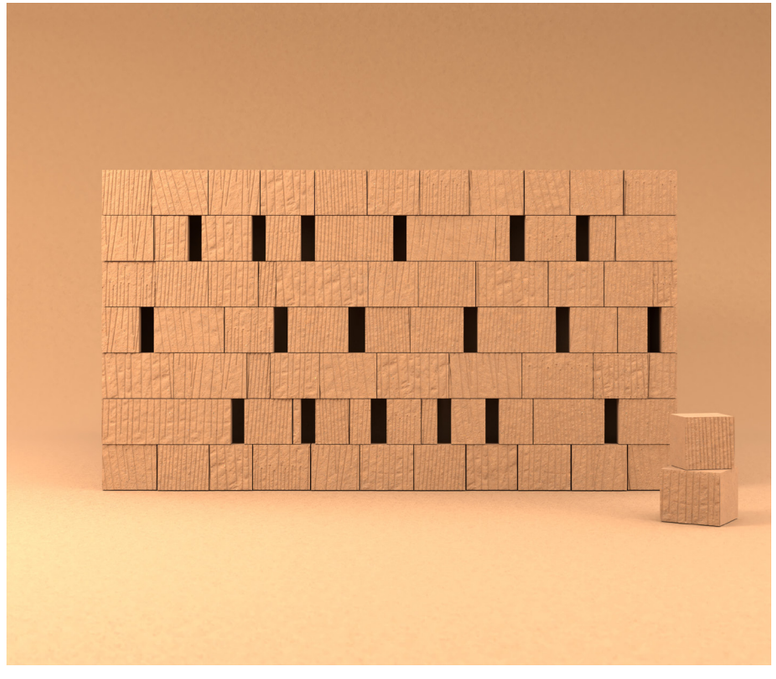


Figure 04. A render-made image of an architectural model of Groupwork's 15 Clerkenwell Close. Courtesy of the author.

[12] Douglas Murphy, "A Song in Stone: 15 Clerkenwell Close, London, UK, by Amin Taha + Groupwork," Architectural Review, July 18, 2018, https://www.architectural-review.com/buildings/a-song-in-stone-15-clerkenwell-close-london-uk-by-amin-taha-groupwork.

[13] Jason Sayer, "Amin Taha Wins Fight to Stop 15 Clerkenwell Close Demolition," *The Architect's Newspaper*, August 15, 2019, https://www.archpaper.com/2019/08/amin-taha-wins-fight-stop-15-clerkenwell-demolition/.

Figure 05. A render-made image of an architectural model of Ensamble Studio's Musical Studies Center. Courtesy of the author.



original abbey. The facade of 15 Clerkenwell Close retraces the thousand-year-old boundary with stacks of off-cuts from the quarry, achieving historical continuity while remaining contemporary.[12]

By identifying the off-cut as a unitary spatial device, the architecture escapes digitality by embracing chance; any attempt at digitally representing this building would result in an abstraction. In many ways, the form and history of its facade came *before* its digital representations, not after.

The inability to communicate its exact form through a digital medium has put the building under threat. Islington Borough Council has repeatedly launched demolition notices on the basis of a discontinuity between the constructed facade and its planning-approved representations. However, the Council would fail in its demolition efforts. After a "pyrrhic victory," Amin Taha, founding principal of Groupwork, has fought for and won the right for 15 Clerkenwell Close to stand.[13]

The inability to communicate its exact form through a digital medium has put the building under threat.

There are apparent graphic similarities between 15 Clerkenwell Close and Ensamble Studio's Musical Studies Center. However, the oscillation between rough and smooth in the former is less prominent in the latter. Instead of limestone off-cuts, thick panels of un-surfaced local Galician granite densely line the facade, abandoning the beam-column binary for a more monolithic reading. The graphic quality of the exterior is entirely dominated by the natural striations derived from granite's required extractive labor. *Stereotomy,* the practice of shaping wood or stone, is deliberately approached with an embrace of spontaneity and the pentimenti of industrial form-giving. Although it appears as a simple unitary stacking, this building could not be constructed the same way twice.[14]

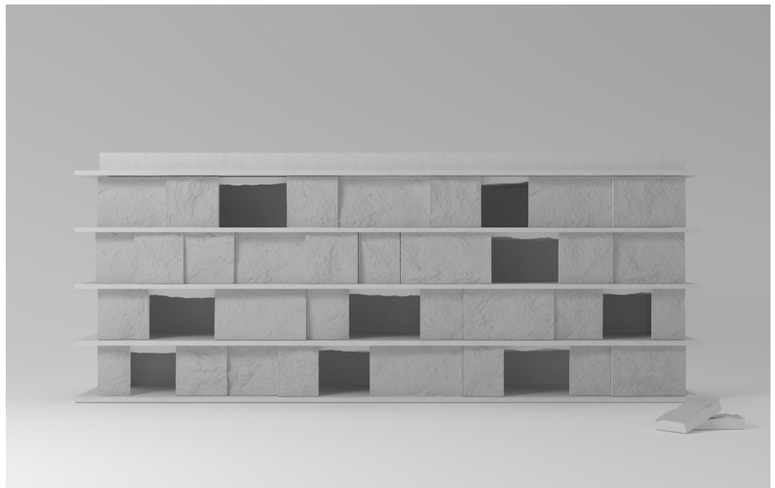


Figure 06. A render-made image of an architectural model of Studio Anne Holtrop's 35 Green Corner Building. Courtesy of the author.

Similarly, Studio Anne Holtrop's 35 Green Corner Building in Muharraq, Bahrain, imbues contextual resonance. Instead of utilizing an extractive method of production, each unitary element is sand-casted on-site, a few feet from its destination. Akin to displacement mapping, each component has a discrete length and width, with their depths variably determined by arbitrary sedimentary values.[15] Holtrop is dedicated to randomness, citing the Rorschach test as a primary inspiration for the speculatory intrigue provided by abstract form.[16] The human mind is naturally tuned to the qualitative reading of an object; its quantitative values are only made relevant by their inclusion in contemporary techniques.

These projects are testaments to the possibility of thoughtful engagement with contemporary issues outside the strict confines of digital discretion and exactitude. Identifying a built component and graphically communicating its assembly is distinctly architectural; this process has been the foundation of architectural practice since its inception and will remain so past the use of digital tools. Boulders are made of countless distinct sediments, each with its own form, origin, and significance. However, these minute distinctions fall second to the forms, origins, and significance of the boulders themselves. How they were processed metaphysically through our computer systems becomes pertinent to how we interpret architectural meaning.

[14] *Architectural Practices*, vol. 142, El Croquis (Madrid: El Croquis Editorial, 2008).

[15] *Studio Anne Holtrop, 2009-2020*, vol. 206, El Croquis (Madrid: El Croquis Editorial, 2020).

[16] Marina Montresor and Stephan Lando, *Defining Criteria* (Luzern: Quart Verlag, 2018).