Envisage:
Investigating Design Intentions, Visual Perception through Eye Tracking of Architectural Sketches

By

Xiaoyun Zhang

Bachelor of Architecture
University of Notre Dame, 2017

Master of Architectural Design and Urbanism
University of Notre Dame, 2019

SUBMITTED TO THE DEPARTMENT OF ARCHITECTURE IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN ARCHITECTURE STUDIES
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2021

Signature of Author: ______________________________
Department of Architecture
May 14, 2021

Certified by: ______________________________
Takehiko Nagakura
Associate Professor of Design and Computation
Thesis Advisor

Certified by: ______________________________
Axel Kilian
Visiting Assistant Professor in the Department of Architecture
Thesis Reader

Accepted by: ______________________________
Leslie K. Norford
Professor of Building Technology
Chair of the Department Committee on Graduate Students

Thesis Committee:

Thesis Advisor: Takehiko Nagakura
Title: Associate Professor of Design and Computation
Thesis Reader: Axel Kilian
Title: Visiting Assistant Professor in the Department of Architecture

Envisage:
Investigating Design Intentions, Visual Perception through Eye Tracking of Architectural Sketches

By

Xiaoyun Zhang

Submitted to The Department of Architecture
on May 14, 2021 in Partial Fulfillment of the
Requirements for the Degree of
Master of Science in Architecture Studies

ABSTRACT

Are we able to perceive an architect's intention through observation of his or her sketches? Yes, but it requires a probing process of observation. Across time and continents, master architects have developed a collection of the processes for expressing powerful design intentions through succinct and dynamic representation, or design sketches. Different types of sketches describe, express, or gesture about the architecture they represent. They deliver active ideas that are not limited to objects but provide a raw sense for both the perception and creation enabled through visual thinking.

I propose a method to utilize eye-tracking as a translator between the graphics and the architects' perception of three types of intention: shape, composition, and circulation. My hypothesis is that we can perceive how architects represent these intentions -- through the means of graphics, which allows a more ambiguous and dynamic translation between intention and sketches, we can probe the underlying process by observing a viewer's eye movements. Furthermore, heat maps, obtained from eye movements, can be adapted to a machine learning algorithm -- Image-conditioned Generative Adversarial Networks (GANs). I use this algorithm to translate the raw sense of space and visual gesture to capture human-level information acquisition of these intentions.

To demonstrate the work, I first discuss the history of visual power in design and a shift towards units and segmentation, covering the development from the emergence of design drawings to the innovation in parametric design. I then proceed with an eye-tracking study where I asked graduate architecture students to observe sketches by Louis Kahn. I study how the graphics of heat maps from eye-tracking decode the participants' perception of intentions in sketches based on a shared educational background in architecture. Then, I propose a framework of utilizing such a representation system to train machines to predict human-level view patterns. Finally, I examine how effective this system will function with an image-to-image machine learning algorithm known as the image-conditional GANs.

From the study it can be implied that mechanical eye-movements reveal a shared visual-thinking procedure that has been unconsciously practiced by human designers. Such a procedure, if learned by machines, will facilitate a creative process that utilize such informal dynamics derived from eye movement in visual representation in design.

Thesis Advisor: Takehiko Nagakura
Title: Associate Professor of Design and Computation
Thesis Reader: Axel Kilian
Title: Visiting Assistant Professor in the Department of Architecture

“The marble not yet carved can hold the form of every thought the greatest artist has.”

---- Michelangelo

“The value of uncompleted things is very strong…If the spirit is there and can be recorded, what is lost? The drawing is important, the incomplete scheme is important, if it has a central gravitational force which makes the arrangement not just an arrangement but something which gives a richness to the associations which are lost. Recording of that which has not been done must be made much of.”

---- Louis Kahn

Acknowledgements

Thank you, Takehiko, for your tireless encouragement and trust throughout my time at MIT. Your dedication and deeds in pursuing the best solutions possible through personal challenges have greatly influenced me in many aspects of both study and life. Thank you, Axel, for your brilliant support and inspirational work that have broadened my perspective from an architectural design student to a learner in the realm of machines and humans. Thank you, Larry, for your constant challenge, sincere advice, and outstanding attitudes that guide my path to a greater world of scientific research. Thank you, George, for your charismatic lectures and steadfast inspirations in seeing the world in a more spectacular way. Thank you, Terry, for your ingeniously guidance and serving as a role model to me for my life. Thank you, Skylar, for your wonderful studio lectures and introducing me to the fresh world of computational design.

Thank you everyone in the computation program and Department of Architecture for your presence and invisible encouragement whether inside or outside building 7. Although the sudden pandemic has brought an obstacle for me to see and talk with most of you, just thinking of all the shining stars who compose the sky of architecture at MIT has made me to be grateful and always thinking of all your talents and ambitions.

Thank you Rania, Elina, Alex, Bowen, Jeremy, and Ki for being one of the best groups of peers I could have ever imagined, and for constantly encouraging me to see more and move forward. Thank you Tonya for your heart-warming conversation and a bright attitude of life. Thank you Inala for your great help and tasty cookie that remind me of joy and appreciation. Thank you Tomoko for the ever-warming smile and encouragement.

Thank you Krupali, John, Semir, Ettore, Giovanna, Alan, JC, Duncan, Lisa, Selena, David, Brian, Thomas, Michael, Donna, and Father Tom for always supporting me from our home of the Welsh Family Hall at Notre Dame. Knowing and learning from you will always be one of my most beautiful experiences.

Thank you Department of Architecture at MIT for all the trust and challenge you granted me to navigate and search for my goals. It has been a great privilege to be accepted as one of your disciples.

Finally, thank you mom and dad, and Corgi Mini for living with me as a family bounded by blood, joy, and spirits. This thesis, and my experience so far would not become possible without any of your selfless love, trust, and sacrifice. I am truly lucky to be sheltered under your angelical wings.

--Xiaoyun (Margaret)

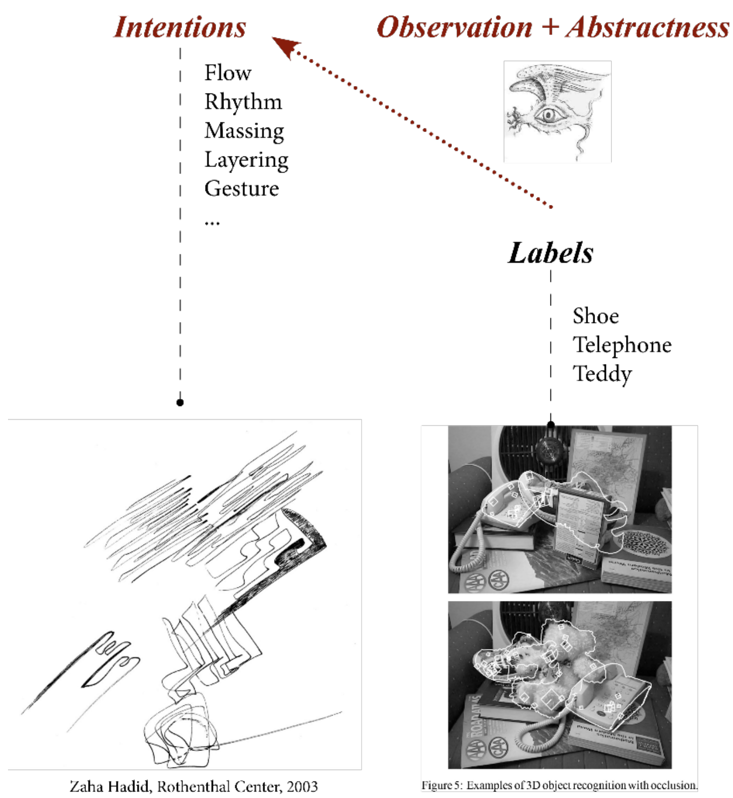


Figure 1. Perception from labels to intention is assisted by active observation and deliberate abstractness.

Table of Content

Introduction

“The marble not yet carved can hold the form of every thought the greatest artist has” (Quotes of Michelangelo, 2021). This quote of Michelangelo emphasizes on visual potentials of artists’ eyes to “carve out” design intentions from a piece of stone by finding their own interpretation of the images within. Such images are undefined, dynamic, and enriched with visual possibilities, in contrast with a stone’s natural properties which are defined and measurable with labels and numbers.

How designers, especially architects, apprehend these ambiguous and hidden intentions within the stones is what I have been interested in since the first day I learned about how to observe in order to design. To retrieve any innate design power of an object in addition to physical properties and generate my own image before eyes. Like the statue of Niki, a thoughtful recess from complete definition of the form brings out a power for imaginations to fly (figure 2.).

Figure 2. Michelangelo’s the “Atlas”, 1525 – 1530 (Wikipedia, n.d.); a block of marble (Marbleicons, 2020); Winged Victory of Samothrace, 200 – 190 BC (Lyokoï88, 2015).

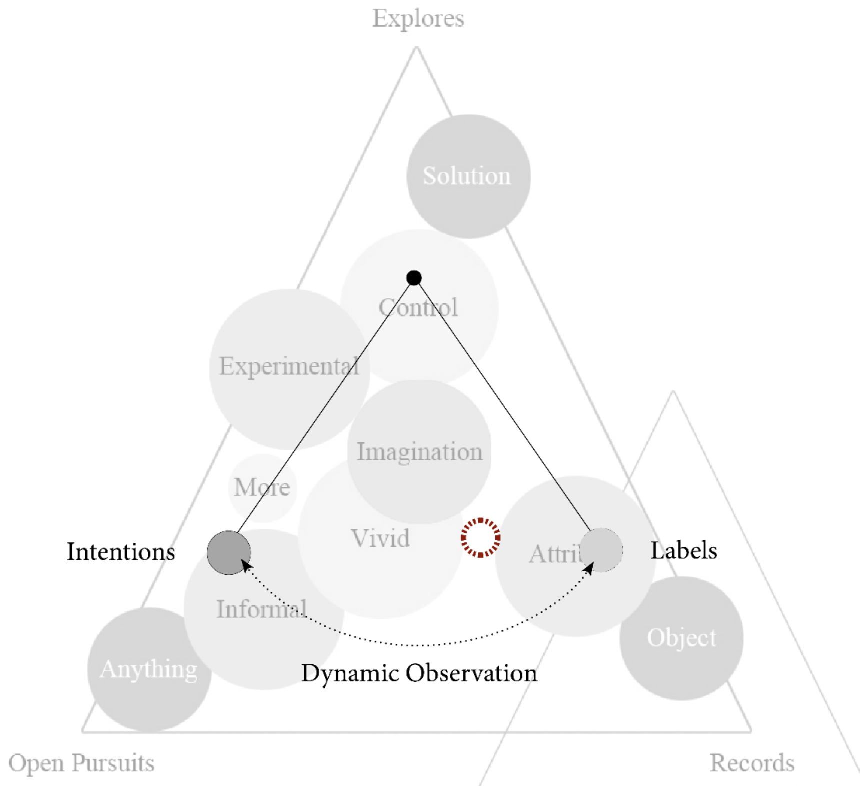


Figure 3. Three major goals of design representation. After James S. Ackerman, 2016, and a pendulum between intentions and labels.

Visions "is a key to man's creative power, even on the most rudimentary level" (Kepes, 1965), and many aspects compose this wing of imagination. How such willful power of the eyes is related to design representations, unsurprisingly, has been a long-established topic among practitioners and theorists, producing connections as well as gaps among how and why to use designers' eyes. For example, James S. Ackerman (2016) discusses visual goals of design and design representation in his book "Origins, invention and revision: Studying the history of art and architecture". On the topic "goals in design representation", Ackerman proposes three major goals: explores, records and open pursuits, forming a triangular relationship (figure 3.). To fill this triangle, I have placed words of small goals which Ackerman uses when explaining the three major goals. Shortly after I put the circles and shades of greys in, I noticed that a gap exists between "record" and the other two if this diagram represents how people and machines see and interpret design images. Then, I noticed that what could make this gap is a disconnection between objective attributes with informal representations. Where I place a red, dashed circle. Unlike other connections between circles, for example (figure 4.): in 1969 Charles Eastman proposed a system to strengthen the link between solution and control in his novel system which was later developed into BIM; Google's image segmentation automated searching of objects by low-level patterns that represent as unique attributes of a label in reading of 2D images (Li, F., Johnson, J., & Yeung, S. 2017). Zaha Hadid's sketch connects imagination visually with experiments and other projects on image generation and matching added a form of control after experiments of algorithmic translation for visual perception. This gap here, between vivid imagination and objective attribute, is where I like to explore. That means attempting to fill the gap requires a combination of human observation behaviors and understanding of the images, or "design intentions."

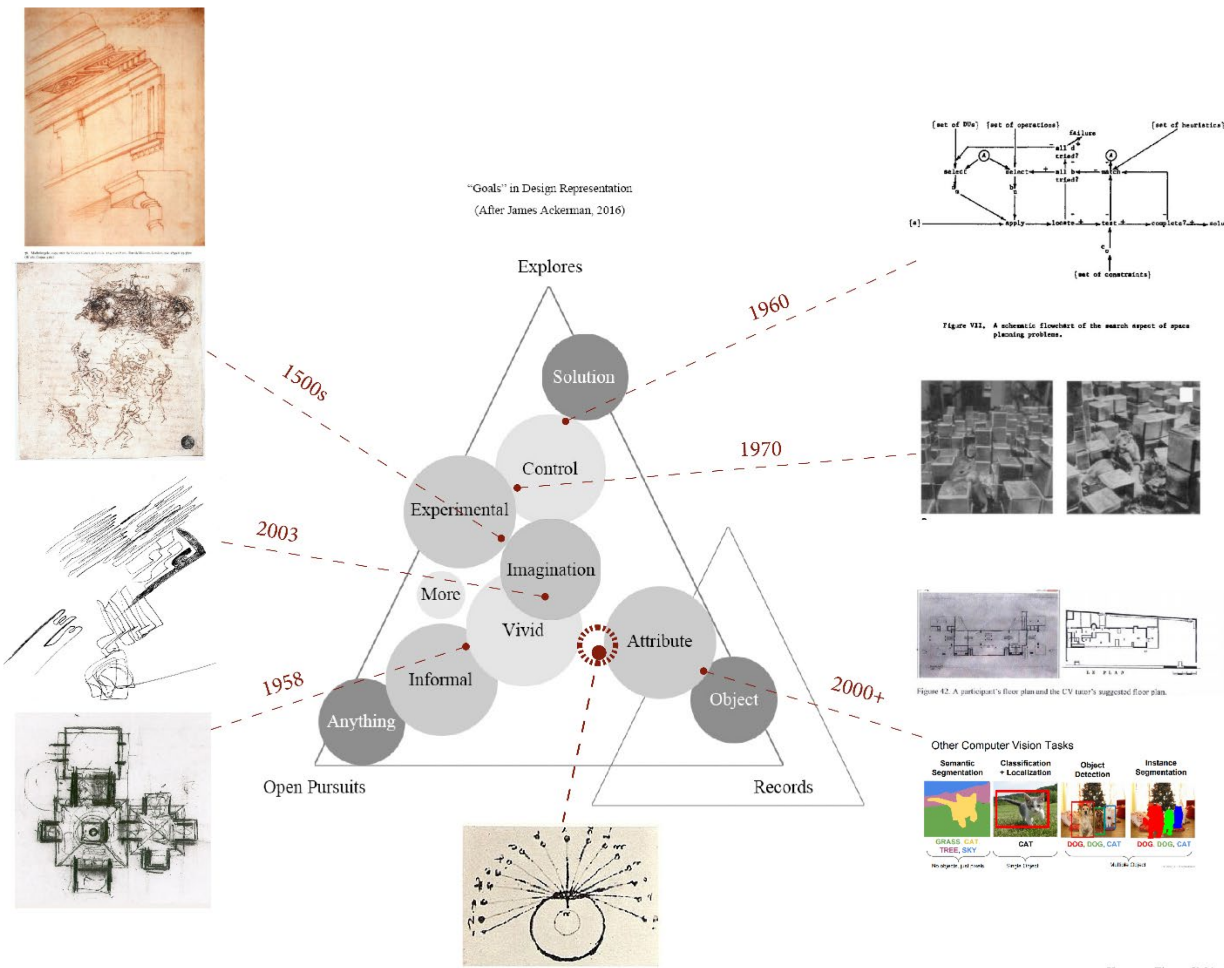


Figure 4. Examples of visual tasks the bridge pairs of goals in design representation. From top left, counterclockwise: Michelangelo, study of the entablature of the theater of Marcellus (Brothers, 2008); Da Vinci, study of the Battle of Cascina (Brothers, 2008); Zaha Hadid, sketch of design for Rothenthal Center, 2003; Louis Kahn, plan study of the Morris residence. (Architectural Archives, University of Pennsylvania, 2021); Da Vinci, diagram of lines of sight (1955); Li, Fei-Fei, Detection and Segmentation (2017); Park, Synthetic Tutor (2015); Negropoonte, the Architecture Machine (1973); Eastman, prototype of BIM (1969).

Apprehension of intentions --and more specifically, in architectural design -- rests beyond object segmentation in more abstract qualities and robust, graphical representations that emerge and change. It requires a probe of observation and is especially powerful to architects when designing without defining labels while seeking big moves of volume and its visual manifestation. Different types of drawings, some describe, some express, and some gesture about ideas that are not limited to objects, but raw senses of the space in many exemplary buildings. To explore this further, my motivation is to inquire representations of these dynamic visual intentions that are innate to design sketches, and a machine learning algorithm to possibly translate these raw senses of spaces and gestures visually.

The primary idea is that graphical representation of eye movement from gaze-tracking studies can extract and represent "intention" from design sketches, as well as interpret human designers' observational process into an image-based form learnable by machines.

I developed my work into four chapters, the first two chapters discuss the history of visual power in design and a shift towards mechanical description and segmentation. The third chapter includes documentation and analysis of gaze-tracking experiments of architecture students' reaction to design sketches. The fourth chapter proposes a framework of observation named "Envisage" based on the first three chapters' discussion and tests, followed by two prototype tests using sets of trained neural networks on predicting human-like viewing patterns, and translating viewing patterns into conceptual design graphics using visual imagination.

Chapter 1: The Power of Seeing: Record and Imagine.

The first chapter starts by discussing a timeline of evolution in visual media and tools from cave paintings to machine vision algorithms. Following the identification of these representational methods, I discuss how architectural designers utilize them to assist design processes. I look at visual powers in three major areas: designers, tool and machines, and machines' visual intelligence. By linking discoveries parallelly on the timeline, I examine the emergence of graphical representations, their adaptation in the Renaissance, Bauhaus System, and later human-machine collaboration initiated by Sutherland's sketchpad in 1960. I also discuss the development of imaging tools and algorithms such as hardware of photography starting in the 19th century, and statistical pattern recognition that founded the base of machine vision in the mid-20th century. Finally, I conclude this investigation in history by suggesting a potential of utilizing the three realms in a cohesive manner, which relies on the translation of information among all.

Chapter 2: Discovering by the Eyes: Intention from Designers' Observation.

In this chapter, I investigate a linkage between designer's intention and its manifestation in the form of architectural design sketches. I answer questions of "where" and "how" architectural intentions are represented. To explore a fuller spectrum of intention, which has been developed and applied in architectural design, this chapter investigates one of the most straightforward visual expressions of design ideas: sketches, with works from classical, modern, contemporary examples. I analyze design drawings in the early Renaissance of Michelangelo, Da Vinci, Louis Khan, Mies Van de Rohe and Kazuyo Sejima; and "visual thinking" systems discussed by Negroponte, Kepes, and James S. Ackerman. To compare to the approach of architects, I discuss another process of segmentation includes image recognition algorithms in popular search engines and machine learning algorithms for assisting visual design, includes Google's Image search, object detection, origins of BIM and a MIT PhD thesis in 2015 using pattern detection to match plan drawings. From analyzing these examples, I conclude important inspirations for the eye-tracking study in chapter three.

<u>Chapter 3: Interpreting the Eyes: Architectural Design Intention in Eye-Tracking.</u>

In the third chapter, I explained logistics behind eye-tracking to explore architectural intentions in design sketches. Then, I report a set of tests and assessments on a particular representation strategy empowered by gaze-based interaction for analyzing forty-five plan sketches of Louis Kahn. I asked three questions to each participant regarding observing architectural intentions: shapes, composition, and circulation. Following test documentation, I conduct analysis of heat maps and graphical relatedness, and proposed three major findings from the heat maps and sketches.

<u>Chapter 4: Creating with the Eyes.</u>

The fourth chapter will propose a framework of observation named "Envisage" based on the first three chapters' discussion and tests, followed by two prototype tests on representational strategy and eye-tracking software for conceptual design using visual imagination. This prototype utilizes Image-conditioned Generative Adversarial Network, a machine learning algorithm which is recently proposed to translate one image into another. The first section consists of descriptions of this application. The second section consists of examples and discussions on the two prototypes that will be an attestation on how effectively this framework can become in delivering high-level representation of design intentions from new sketches: one is "Sketch to Intention", and the second "Observation to Gesture". The third section will be a discussion of all the work so far and explain why this application will be a step towards a more dynamic, observer-based design system that encourages more natural human-computer interaction.

Chapter 1 The Power of Seeing: Record and Imagine

Visual perception and creation are crucial for artists and designers. Among many tools that designers utilize and express to compose their works, vision is one of the most sought after and multi-leveled aspects of human sensory. Others includes tactile, olfactory, auditory, and motor (Negroponte, 1969). Continuous training and development of visual ability are crucial to a designer in both two-dimensional and spatial imagination. It facilitates visual thinking and reference-making when one observes and envisions subsequential decisions without adhering to objects and their physical properties. Either functioning to absorb the “primal sanities of nature”(Walt Whitman, 1900), or “reproduc[ing] them in the world [man] shapes for himself”(Kepes, 1969), how these two visual forms of thinkings could be mastered by designers has been an ever-attractive exploration among hunters and gathers, artists, art theorists, psychologies, neuroscientists, sociologists, and computer scientists alike -- since the very dawn of one of the first artistic visual expressions created by human ancestors in the cave painting in Lascaux, France in 15,000 BC (figure 5.).

Figure 5. A painting of the Giant Deer from Lascaux (HTO, 2009)

Although how and for what visual communication were created can be attributed to an exceptionally long and deep stream which is as complicated as the human neural system, the primary goals are twofold: first, to record events and objects in the real world; and second, to experiment on something imaginarily. The former goal has existed as long as people started observing the natural world around them, and the later were systematically formed into branches when the pursuit of art and design were deemed as intelligent activities. The following discussion will provide some historical examples to explain these two goals.

Visual power in recording has widely appeared across time and places. Recording of a natural object using graphical interpretation is resulted from haptic interaction between the observer and the object. Paleolithic Arts such as the cave paintings in Lascaux, while they are hardly attributed to exact uses (Karstworlds, 2018), are products of recording what the drawers had seen in life: one famous example among all is the painting of "the Giant Deer" (figure 5.), which was thought to depict hunting or ritual scenes composed of animals; hieroglyphs are a more abstract form of figurative recording and interpretation of meaningful objects in order to communicate: many characters in Egyptian hieroglyphs resembles natural object to signal a specific trait related to that object, and so does ancient Chinese characters (small seal script) and prototypes of Phoenician alphabet (figure 6.); Early Roman frescos in houses and temples depicted scenes of human or deity lives, presenting another dimension of reality by "showing" what a dream house or a luxury party will look like from a human perspective; and funeral murals prevailing in both ancient East and West embodied ideal rest places for souls. Records and interpretation of visually meaningful existence made it possible for those who had never seen before to relate and to remember what had happened and what could happen, in the most straightforward way of visual stimulus.

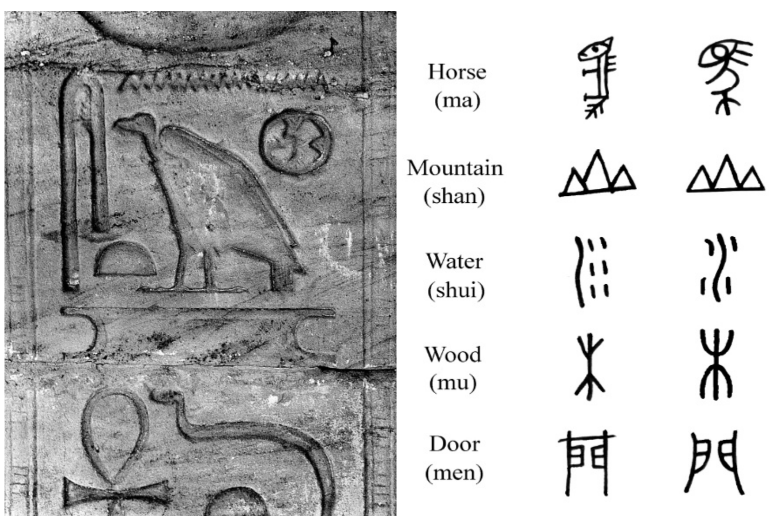


Figure 6. Visual information of languages delivered through abstract characters. From left to right: Hieroglyphs at Amada, at temple founded by Tuthmosis III (UNESCO, 1960); Pictographs examples in Chinese Characters (Buckingham- Hsiao, 2018).

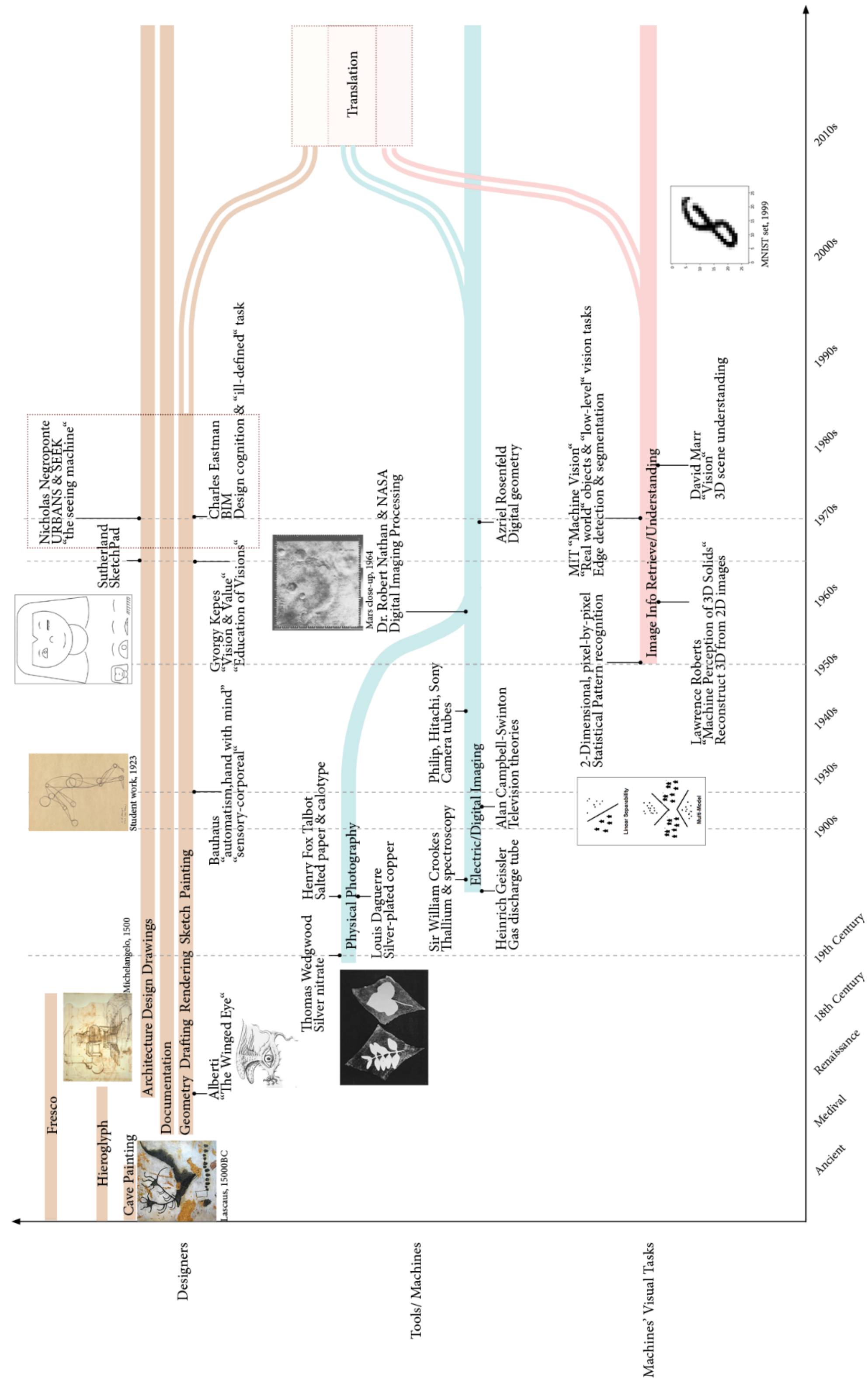


Figure 7. Selective major events in the development of visual communication of Designers, Tools, and Machine Intelligent.

1.1 Eyes of Observers: Perceive and Imagine

Moving forward to the second blooming of Western Arts in the Renaissance, the second goal of visual communication has been more or less acknowledged. The goal is to extract and therefore facilitate intelligent visual imagination, or in other word, design. When the re-discovery and adaptation of ancient Greek and Roman Arts had been a pursuit of high-level design intelligence such as architects, they started seeking a level of understanding towards incorporating ancient treasures and ruins into practical design and celebration of new lifestyles. One of the most significant contributors in the heart of Renaissance is the development of sketch freely, which was initially discovered in prints in the mid-1500s (Ackerman, 2016). Such an innovative method on how drawing became a way of abstraction, reflection, exploration, and most importantly, or testing and materializing design intentions. James Ackerman, an art historian specializing in Renaissance works, points out that in the creative process, sketch "liberated from convention" in comparison to "deliberate preparatory drawings" (p. I), and it is therefore purposed for three major kinds:

"One that explores possible solutions for a particular work; one without a specific goal – an open pursuit of potentialities that may not even represent anything; and one that records an object or occurrence in the external world…with some degree of reliability" (Ackerman, p. 2).

Ackerman's three goals of making a sketch can also be interpreted into a set of goals to create visual representation of a certain existing or to-exist object. Renowned masters of architecture and art in this period demonstrated how such connection can be made: between what is visually presented and what can be visually created. This perception-creation circle is a recurring process enriched by multiple goals of describing and imagining. According to Ackerman (2016), the starting of sketch exploration emerged from the painter Pisanello from Verona (1395 - 1455), succeeded traditions in "pattern books composed of nature studies", documented an inauguration event in 1438 with brief line drawings of horses and textual description for later paintings. Leonardo da Vinci utilized sketch to "study nature, solve problems, develop compositions, and fantasize" (p. 8). Later, Michelangelo's graphical interpretation of architectural motifs from ancient ruins and documentation drawings are not merely representation of existing conditions but "was extended to human analogies" (p. 17), seeing them as a base of abstraction and inspiration to morph and alter. Leon Battista Alberti greatly appraised the imagination and artistic discovery brought forth by "the Winged Eye", as seeing and observing became an inseparable component of holistic beauty when designing architecture. Andrea Palladio, who initialized architecture career as a stone caver, later produced "freehand fantasies" in ink (figure 8.) to "just exercising his imagination" (p.

17). As Ackerman summarized in his book "Origins, Invention, Revision: Studying the History of Art and Architecture", sketch, as a part of visual thinking, "unrestrained by tradition, loose and indeterminate in structure, and issuing straight from the artists' inspiration and vision, and the hand" (p. 17).

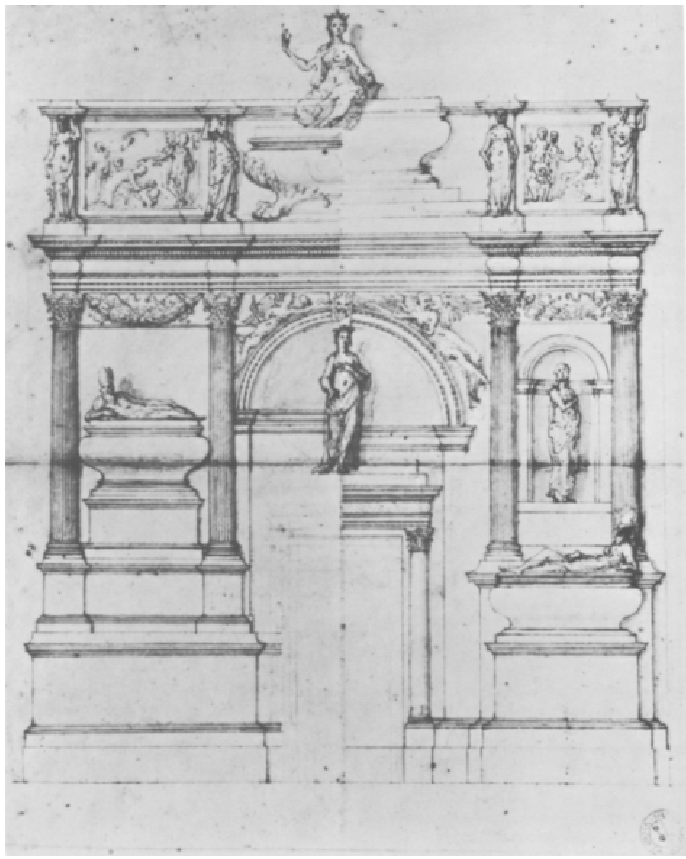


Figure 8. Andrea Palladio, 1564, Proposals for the Grimani Tombs at S. Francesco della Vigna, Venice (D -17). Museo Civico, Vincenza. (Joachim, H. 1981).

Not only from the hands of Renaissance innovators the use of vision has been flourishing, but the discovery of photography techniques also streamed into the long river of visual perception and creation, about one century earlier than the revolution of the Bauhaus. To "capture the world in its types and regularities" on "objective splash" (Daston and Galison, 2017, p. 11), this raising of technology to recreate images of real-world objects on a durable, physical surface, establishes the industry of "imaging" through human-designed "eyes" of machines. By the end of 18^{th} century, Thomas Wedgwood proposed the first process of visually "fixing" silhouette images on stable medium like paper using light-sensitive chemical reactions, physically imitating the process of human retinal receiving light signals, and then forming an image (figure 9.). About half a century after the discovery of physical photography techniques, advances in electricity enabled creation of electrical imaging equipment to generate lights, forming images in addition to capturing them using silver-plated paper. In 1857, German physicist and glassblower Heinrich Geissler invented the Geissler Tube, which is deemed as the primal version of gas discharge tubes that later advanced into television and many digital imaging devices. This technique made the visual interpretation of natural objects no longer a human-only ability but gained a level of objectivity and solidity, for the image could become a direct product from natural interaction – chemical or physical --- between objects and objects, and without human interventions. How to make more true-to-nature[1] images has become a crucial consideration since the emergence of these techniques.

For example, many photo-realistic renderings, either still or animated, are one type of reflections of "reality." These digital images of an unbuilt design are heavily modeled and executed through mechanical simulation, mapping natural phenomena such as light reflection and dynamics of water and winds. More comprehensive simulations of realistic environments facilitate a process of describing and predicting towards complete objectivity. Such process then assures the translation between imagination and the real word as a product of verifiable, scientific interpretation[2].

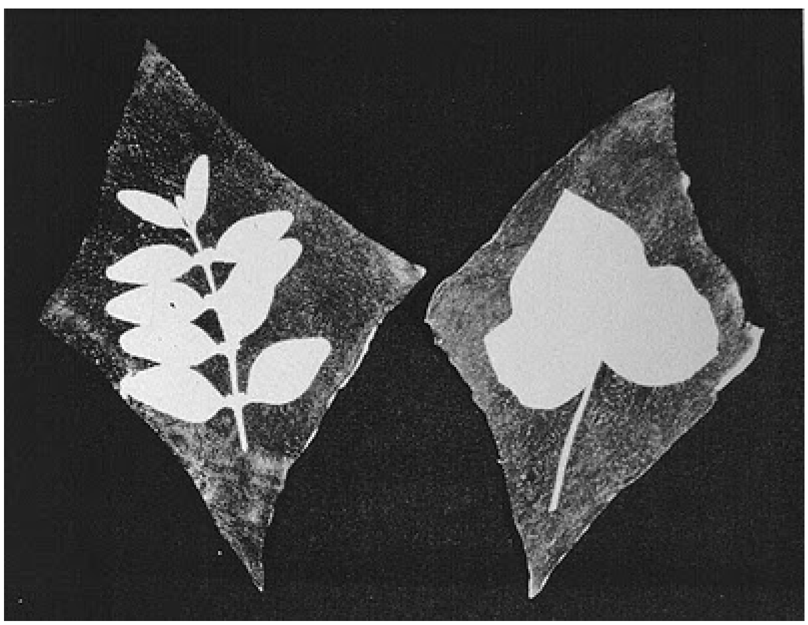


Figure 9. Images of leaves captured using similar methods as Thomas Wedgwood's proposal to "fix" lights using chemical reactions (Photohistoryreview, 2008).

[1] True-to-nature was a "precondition for mechanical objectivity, just as mechanical objectivity was a precondition for trained judgement." - Daston and Galison, p. 18.
[2] Discussion in this paragraph was inspired by Prof. Kilian.

Up to this point, shortly after Alan Campbell-Swinton's proposal of television theories in 1915, the rapid advances of machinery, and the increasingly realistic interpretation of visual signals, the field of design looked forward to another systematic approach of expressing visual information in a meaningful while encoded through machines perception. In addition to its claim of renewing the design industry and pursuing "comprehensive artwork", the Bauhaus incorporated mechanical aesthetics as an important aspect in visual education and creation. Clean, efficient, and industrial visions of design have given visual observation another level of intentions: allocating functions of each part in addition to general composition and appearance. While eyes of designers capture appearances, gestures, and compositional values through observation, they are also responsible for extracting machine-like austerity and dynamics from massing and abstract representation of shapes and relations. The Bauhaus' shift from visual aesthetics to visual analogies has made the roles of designers' eyes to perform another task: to "see" internal and elemental connections and properties of the design as a system (figure 10.): a link of objects to mechanisms, to how the design should be operated.

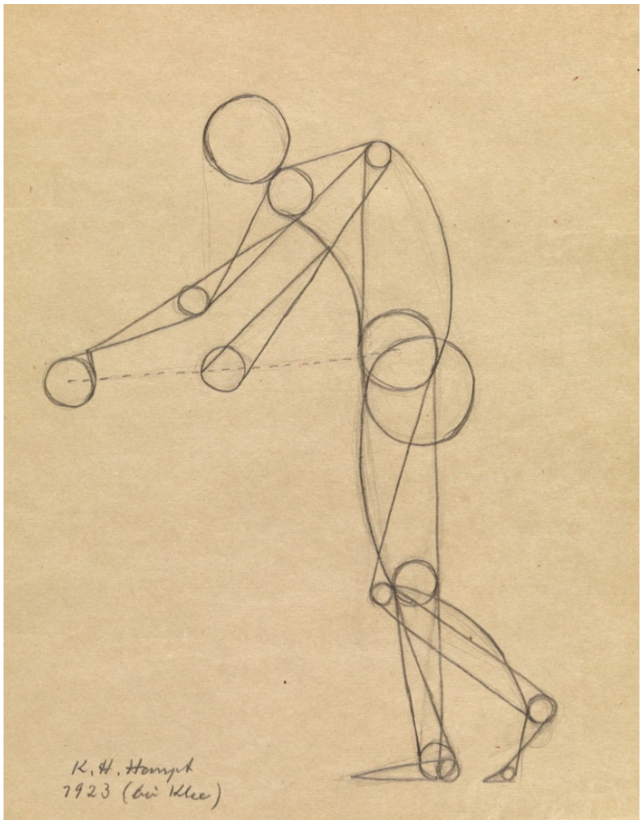


Figure 10. "An edifice of flesh, muscles, bones" (Schlemmer, p.41). Figure drawing for Paul Klee's course, Karl Hermann Haupt, 1923. Bauhaus Student Work, 1919-1933. The Getty Research Institute. More than a diagram of joint-and-bone structure, the resemblance between mechanical gears or conveying belts and human body is obvious in this orthogonal view of a standing figure.

1.2 Eyes of Machines: Efficient Communication

In order to operate design products, more substantial and attribute-oriented “vision systems” have risen given the increasing demand on construction communication and efficiency. Tools and machines have been developed to perform one similar task as that of what designers are able to perform: recording objects and events of the real world but with greater accuracy and efficiency. Utilization of machines’ precision and fast processing of complicated images has been developed fiercely. Starting in the 1960s where the first digital imaging of a close-up of Mars surface was captured by NASA with the work of Dr. Robert Nathan, around the same time pattern recognition and three-dimensional shape perception were explored on pixel and contour-based (Roberts, 1963) Machine Vision tasks. A turning point for designers to enhance their ability in order to create visual representation and link to a computerized processing system is the invention of SketchPad by Ivan Sutherland at MIT in 1963 (Sutherland, 1963).

SketchPad made it possible to link designer’s hand input, which was originally applied to unalterable, physical surfaces only, to be translated into a digitized form through another visual medium know as the digital screen using mathematical representations. Therefore, machines started to “see” what has been drawn by referring to zero-dimensional Boolean operations: lines are represented by two points, and points are represented by a pair of numbers of their relative location on a screen coordinate system. While machines’ efficiency made input more definitive and convenient to be transferred, they also reversely made visual perception of lines, for example, to be viewed more likely as a mathematical expression than an abstraction of any possible visual interpretation.

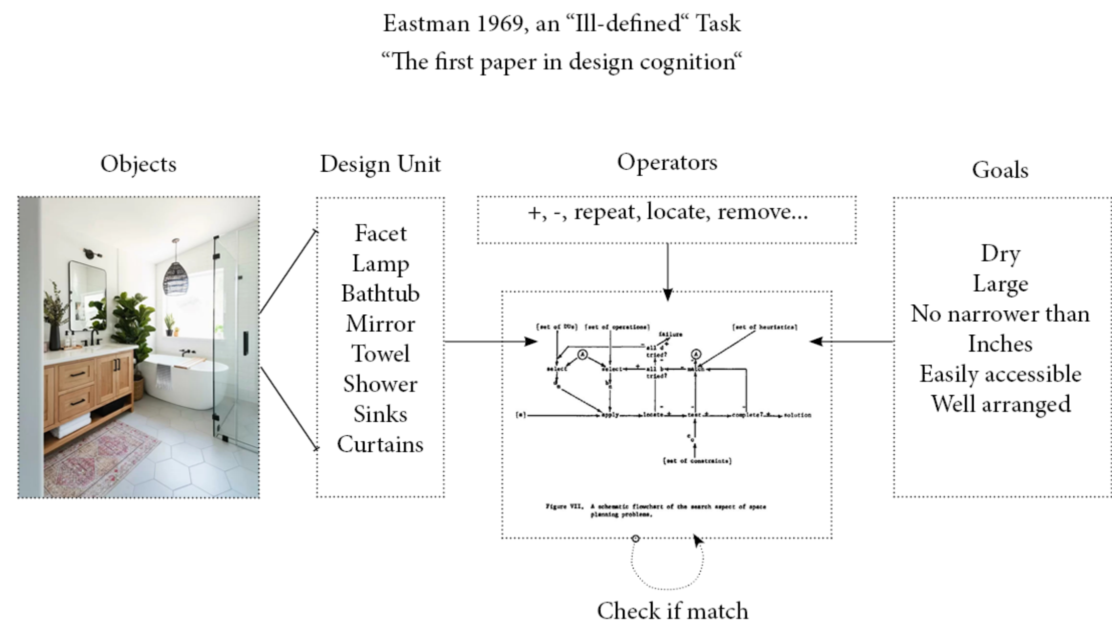


Figure 11. Conceptual pipeline of Eastman's system. "A schematic flowchart of the search aspect of space planning problems", Eastman, 1970, p. 690.

Shortly after the invention of SketchPad, Charles Eastman summarized and criticized existing vagueness and uncertainty in describing design problems using conventional process (Eastman, 1970). His influential response[3] to the "ill-defined" problem in architectural design is to specify "formal languages" to translate abstract design goals such as "'more luxurious' and spacious", and to establish a flow of operation (figure 11.) that identifies "design units" and modifies their properties for the sake of achieving "design goals". These goals, or systemized intentions, are mathematically evaluable by finding numeric thresholds to size, feeling of space, accessibility and so on. Identification, composition, and decomposition of "design units" also "widen the solution space" (p. 672) that augmented solution-finding process by human designers. His work changed the vision of what or who can express design intentions: from trained individual observational reflection to a systematic pipeline of discerning, allocating, and compilation.

However, designers' self-aware powers of vision have not lost their deserved attention. Emergence of the Computer Aided Design (CAD) programs, on the other hand, promoted further exploration of human

[3] Eastman's paper was regarded as "The first paper in design cognition".

designers' pursuit and their reflection of visual communication in design: on meanings and on dynamics in addition to "recording" and "enhancing". Almost at the same time when Eastman proposed the prototype of BIM, artists, and architecture design theorist such as Georgy Kepes and Nicholas Negroponte of MIT, in the 1970s, emphasized on high-level abstraction and compositions, which human designers are rather highly competent in their performance, proceeding far beyond just reasoning about mathematical and material representations of design forms[4]. Along with theoretical approach of Rudolf Arnheim, they assigned tasks of vision to another level of dynamics, to thinking and create:

> *"In all these instances the elements of a problem situation are changed, rearranged, and transformed; the emphasis is shifted, new functions are assigned, new connections are discovered. Such operations, undertaken with a view to attaining solutions, constitute what is known as thinking."*

---- Rudolf Arnheim, "*Visual Thinking*". In Gyorgy Kepes (1965), "*Education of Vision.*"

1.3 Eyes from Tools: Visual Tasks Enabled by Digital Tools

"All problems can be said to consist of translating some entity (A), into some other entity (B), which is specified in terms of goals to be achieved."

—-- Eastman,1969. "*Cognitive Process and Ill-defined Problems: A Case Study from Design.*"

Time-accumulated training of designers' use of visions and progressive advances in digital imaging systems have signaled a merging of the two into a cooperative entity. A shared goal has been to augment and reflect design processes. Systems in delivering design solutions such as Eastman's proposal have given another definition for what to be seen in the eyes of digital tools. Starting approximately in the mid-1960s, discussion has risen to seek for a merging or balance between the context-based information acquisition from designer's human vision and computerized design units and shapes: as an alternative

[4] Shape Grammars by George Stiny.

interpretation of Arnheim's (1965) emphasis on "the next steps" performed by vision and visual thinking, that is to "shift emphasis, assign new functions, and discover new connections".

Architects have been adept at abstracting design intentions and materializing them in tangible forms and designing by haptic, audio, and visual interactions. Especially in vision, these designers reflect upon meanings of their work and see various potentials even from a brief acquisition of visual signals. According to Brothers (2008), Ackerman (2016), Negroponte (1970) and Kepes (1965), it was more likely to base on the abstraction process of embedded ambiguity and morphing potentials in images. Such attitude towards visual experience is also acknowledged by late 20th century design theorists such as Ronald Finke (1995) and Alexander Koutamanis (1995) on creative cognitive power and visual memories.

Around the same time, an "architect-machine partnership" was envisioned, which would be achievable through intelligent sensory exchange between the two sides (Negroponte, 1969). Negroponte not only suggested a foreseeable future when humans and machines collaborate to solve architectural problems, but also proposed frameworks on which this relationship would be built: mapping and interpretating human sensory in computer interfaces, including "visual, tactile, olfactory, auditory, and extra sensory or motor command" (Negroponte, 1969). Among these sensory aspects, the architect emphasized the importance of vision, and the potential for machines to "challenge and question" a design problem by extracting high-level information such as "probabilities, commonalities, intents and patterns." Researchers in the computational design realm have explored many opportunities of this mentality: Multimodal sensory applications are implemented into personal experiences (figure 12.) for a comprehensive description of the spatial quality (Papadopoulou, 2014; Jensen, Foged, and Andersen, 2020); computer vision algorithms are explored to establish computational models that represent human-level experience of architectural spaces (Koile, 2000; Peng, 2018); and interactive visual experiences are incorporated into learning and creative process in exploring design opportunities (Sung, 2013; Park, 2015).

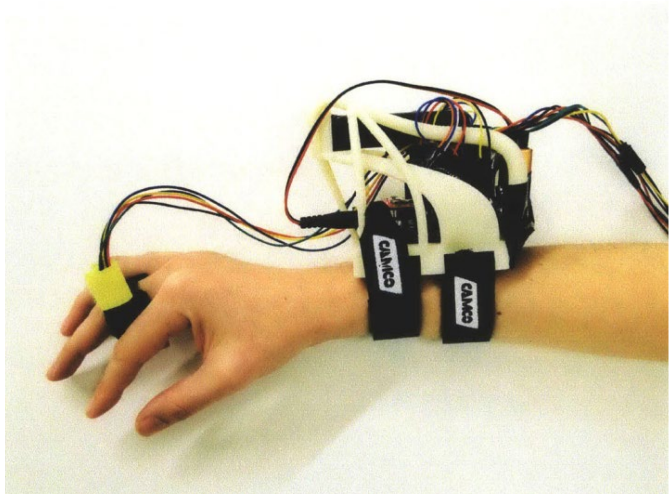


Figure 12. Papadopoulou, 2014. Perceptual Prototypes.

However, before making decisions on the next systems, how to acquire visual information from a design "situation" is even more challenging for machines to accomplish. One persistent problem regarding machines' reading and "understanding" of meaningful information in a graphical stimulus, which is embedded with multiple "intentions" that are aligned with human designer's individual interpretation and wills. Especially in computer vision, a distinctive difference between how people and machines perceive an image is that people are able to pick up contextual and high-level semantic meanings from visual stimulus, while machines are more adept at picking up sub-level, pixelated pattern groups that are deemed to be "distinctive" to a specific image. Machines, especially computer vision (CV) algorithms, deems uncertainty and abstractness very differently from that of human designers: in most situations' ambiguity leads to a drop in optimization score, brittleness of the recognition model (Ilyas et al., 2019), and thus becomes targets of "reduction"; while in the visual design process, abstractness leads to further discovery and exploitation, becoming a crucial generator of opportunities.

Many pioneers have ventured in the realm of constructing artificial intelligence algorithms and its relevance to the process and teleological aspect of design, some recent examples include small scale graphic design tasks, primal spatial arrangement of furniture, and scenic graphics generated from gesturing strokes. The backbones of these AI designers fall into two major categories: one establishing on the exhaustive mathematical modeling of effective properties, and the other is supported via the versatile pattern recognition and matching. The AI is responsible for generating the outcomes that have a particular meaning graphically to human beholders, yet very little to the generating algorithms, as one of the well-known inadequacies in computer vision: what is meaningful to a recognition algorithm is not always meaningful to human perception (Ilyas, A. et al., 2019). Researchers discovered that some minimal disturbance to a test image could completely trick a high-performance, well-trained recognition system to produce errors.

Comparing the two approach of building up a new design, the former is structured from the very elements of formal qualities (Koile, 2000; Takehiko, 1995; Peng, 2018), and the latter is achieved by acknowledging an expressive nature of the imagery itself and therefore designate the imagery as the basic operation unit (Knight, 2015). Computer vision algorithms that distill unique imagery features of a given input to pairing up with a predetermined graphic data (Isola et al., 2018), optimization algorithms that take in numeric interpretation of spatial qualities, and label-matching procedure that links a certain graphic outcome with words and sentences. They are sophisticated, well-described procedures of matching and re-combining. In order for computerized design tasks to give meaningful responses, specific descriptions are required to be presented as a "translation" of the designed context.

In the next section, I discuss two examples of utilizing encoded machine vision in assisting perception and design work.

1.3.1 Google Image Search: Feature Matching

Google's image search, or "reverse image search" was implemented in the summer of 2001. According to Alex Wild (2021), the search enables back-tracking of similar images without needs of keywords, and thus makes tasks like locating credits and seeking image popularity possible. In the same year of its release Google (2021) posted a short video explaining briefly how the algorithm works: the engine "finds" out "distinctive colors, points, lines, and textures" from an input image; then each of these features will be sent to the "back ends for matching against" Google's database. Then the engine returns "visually similar" results by correlating recognizable features as "labels" of an identifiable object.

Figure 13. “Examples of 3D object recognition with occlusion.” Loew, 1999, p. 6, figure 5.

More specifically, the algorithm is a feature detection algorithm called Scale-invariant feature transform (SIFT), which was published by David Lowe in 1999. As an alternative to earlier template matching algorithms (Robert, 1963) which are constrained by slight variation of the image, it utilizes a training database where important “features” of an image would be extracted and labeled as a distinct “description” of the input. For an object to be recognized in a 2D image, many distinctive “image keys” (figure 13.), which are sub-areas within an object that are visually persistent regardless of slight scaling or rotation of the object.

For example, in Lowe's (1999) paper, the author represented “image keys” in white rectangles within outlines of each segmented object. The higher contrast or more stable relative to “image translation, scaling and rotation”, the better the “keys” are. These keys are obtained based only on appearance visible.

Noticeably, Lowe discussed the connections to “biological vision”, and along with other neuroscience scholars they might agree on specific regions of a visual stimulus will contribute to the recognition of particular objects, such as “shapes” (p. 8). “The feature responses have been shown to depend on previous visual learning from exposure to specific objects containing the features” (Logothetis et al. 1995). However, how these “local features” actually make sense or have higher-dimensional meanings rather than a “found” stimulus has been consistently ambiguous towards “recognition” of meanings of the images themselves.

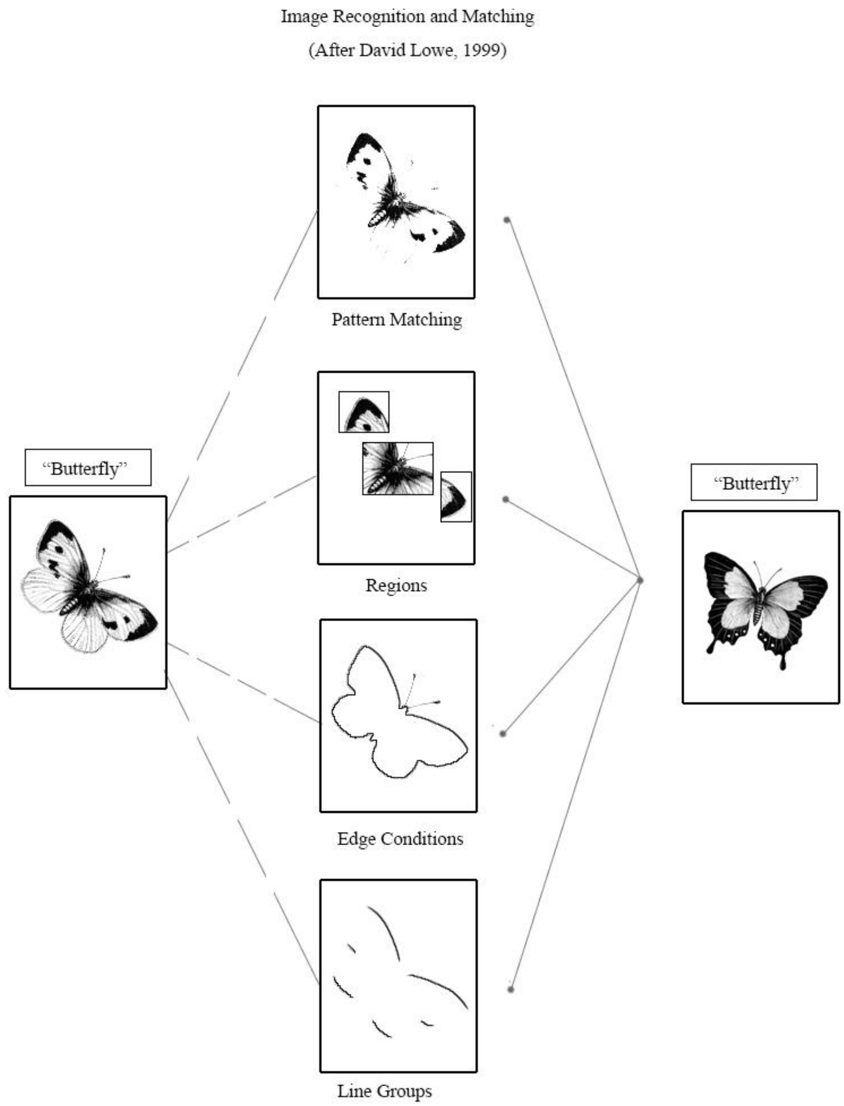


Figure 14. Searching algorithm of Feature matching recognition, Lowe 1999

1.3.2 "Synthetic Tutor": Feature-based Pattern Matching Algorithms on Two-dimensional Architectural Diagrams

In 2015 MIT PhD student Ju Hong Park explored in his dissertation on feature-based pattern-matching algorithms (p. 19) in assistance for finding similar-appearance plan drawings in conceptual design phases. An "Intelligent tutor system" named as "the Synthetic Tutor" was developed to provide "customized teaching materials", including graphical suggestions of a drawn plan, serving as "precedents" to facilitate students' design learning experience (p. 11). This section will focus on a summary of how Park implemented the algorithm to capture "semantic" information from plan drawings.

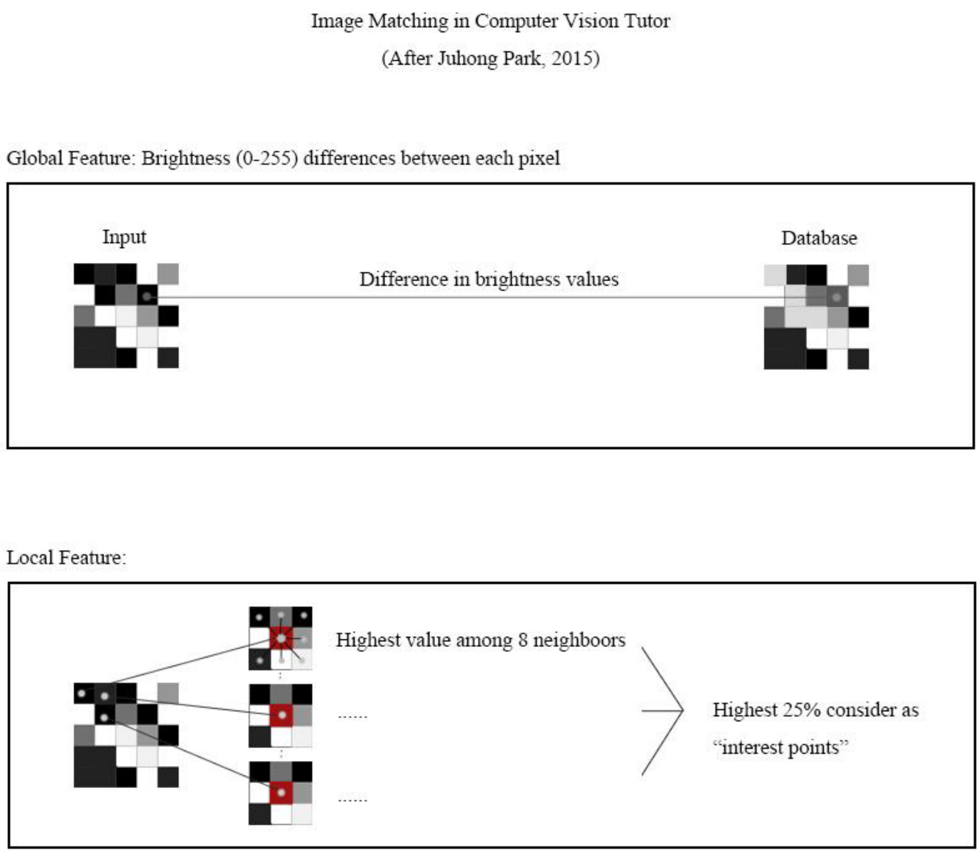


Figure 15. Pattern Matching in Computer Vision Tutor, after Juhong Park, 2015.

Park discussed about the positive potentiality of utilizing image recognition to provide "instructor-like feedback" (p.54) in the "computer vision tutor", in a form of architectural plan drawings. One goal of Park's system is to extract "meaningful" information from input drawings through various processing of the input image. Another goal is to perform "understanding" by the machine on "meaningful" information, by a "global and local feature-based" (p.56) algorithm similar to the Google Image "reverse search" in the previous section (figure 15.). A straightforward global feature evaluates average of the sum of Euclidean distance between every pair of pixels from input and database images; a local feature detection by using interest points, which are recognized from high contrast value from pixel information.

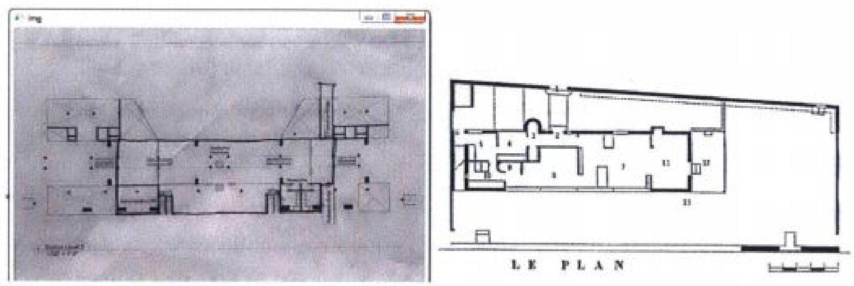


Figure 42. A participant's floor plan and the CV tutor's suggested floor plan.
These two images show the result of suggested design feedback. The left image is a participant' floor plan and the right is a floor plan that the CV tutor suggests (Le Corbusier's Villa Le Lac floor plan).

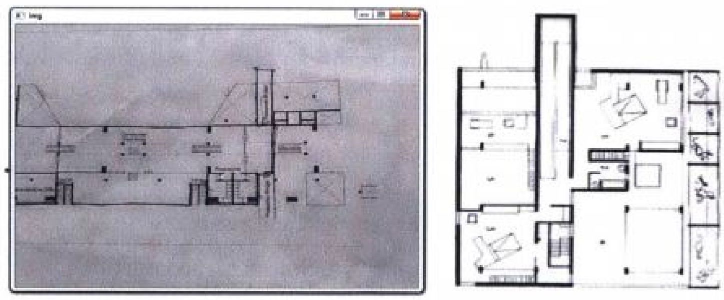


Figure 43. A participant's second attempt and CV tutor's suggested floor plan.
A participant takes a picture of the right part of his floor plan and the CV tutor suggested Le Corbusier's Villa Shodhan floor plan.

Figure 16. Examples of human input and the "Tutor's" suggestion of recognized plan drawings (Park, 2015).

Park commented positively on the pedagogical feedback provided by the CV tutor (figure 16.). From the example result provided in the paper, it is clear that a good match for a clean, definitive plan drawing can be effectively "translated" and thus related to another clean plan drawing in the database (p. 107). However, the author had yet demonstrated drawings in a more ambiguous manner, or in the "sketch" stage for developing schematic design. One possibility of a lack of "sketch" situation is due to a more uncertain graphical representation of information extractable from the inputs. Since the matching algorithm directly retrieves pixel-based, vectorial information and pattern groups from the image itself, it has yet to overcome the disadvantage of capturing "intention" from a design image, which includes recognition of shapes as contextual entities, and their interrelations that structure the whole composition on the canvas.

From the two examples, it is obvious that current computerized visual thinking is mostly benefited by low-level features that make little sense to humans, or distinctly defined image group that respond to search algorithms. They are either too atomic, or too complete. One major guiding principle in machines' training of matching images is to minimize a predetermined loss function. Such goal is meaningful mathematically, however, it is lack of high-level intentions, or contextual understanding of the images themselves. These visual perception algorithms, therefore, reflect only partial of the human design processes that are often emergent, reflective and ephemeral. Greater opportunities reside in the magical "abstractness" of vision that translate a defined object into dynamic representations of visual thinking.

Such thinking process requires a high-level, willful execution of design intentions. "Talking freely about the things I see, and that they alter at will, as I go on[5]" is supported by the reasoning of individual designers. When a novice architect is asked a simple question: "Why did you draw that?", the person should have a decent explanation, whether graphical, mystical or verbal, about the "intention" within. Each of these transient moments in the design process is dynamic, and the next step is executed based on acquiring meaningful information from the previous state, which is to be captured visually.

If machines are capable of performing a similar process of understanding visual inputs, it will be possible to further bridge the making with thinking in a visual manner, which will allow intentional extraction of abstract representation towards a given form.

However, before thinking about how machines will possibly perform a design-like visual task, it is crucial to think first about how designers use visual power and abstractness. Therefore, a history of the designer's visual power is needed to be examined: to find the discovery moments that have shaped and translated design imagination onto visible forms. How designs were progressed when acknowledging the "intentions" that have facilitated those transformation.

[5] Stiny, 2021. "Calculating in imagination's magical realm", script.

Chapter 2 Discovering by the Eyes: Intentions from Designers' Observation

In this chapter, I focus on investigating the link between designer's intentions and their manifestation in the form of architectural design sketches. To explore a fuller spectrum of this relationship between intentions and observations which has been developed and applied in architectural design, I investigate one of the most straightforward visual expressions of design ideas: sketch, with works from classical, modern, and contemporary times. The classic examples will be focusing on works by Michelangelo and Da Vinci, whose use of lines and abstractions are considered to be the first in architectural design graphics. The modern examples will be focusing on works by Le Corbusier, Louis Kahn, Mies van der Rohe and Kazuyo Sejima; and later on, design framing assisted by parametric and interactive programs. Then, I examined three digital examples that consist of machine vision and machine learning programs that consider abstraction as one goals in addition to object segmentation. Sequentially this chapter answers to "how" and "why" intentions have been perceived and created by human architects. To conclude this chapter, I identified three inspirational aspects from the examples of machine algorithms and how they would fit in to the broader narrative of dynamic observation in perceiving design intentions: how to express "verifiability", a time-sensitive description of enclosed shapes, and an image-to-image translation enabled through machine learning algorithms.

2.1 Origins of Intention in Architectural Design Drawings: Find Abstractness through Observation

Subtracting pure descriptive drawings and obtaining abstracted contours that are efficiently expressive is one strategy of invention through drawing practiced by Michelangelo. From the extensive study of the architect's progression in finding design intentions through graphical clues, negating the needs of "binding category or definition to forms" by taking profiles in descriptive drawings away enables visual information in shape and composition to be clearly presented. In Brothers' (2008) book "Michelangelo: The Invention of Architecture", the author describes that by "omitting measurements and origins of the details," Michelangelo transformed a "drawing of an ancient monument" and into a "design drawing" (p. 66). Deduction of every singular detail makes abstracted forms to have enough representative power to indicate what form should be completed: such as the horizontal running lines of dentals and curved profile in the upper partial triglyph. More specifically, his copy of Codex Coner presented a shift from descriptive line drawings towards a form of abstractive qualities that encapsulate both the forms and gestures of an architectural motif, such as the trabeation of the Theatre of Marcellus (figure 17.). The original graphics in Codex Coner are descriptive and definite, with orthographic projections of ancient Roman monuments with measurement attached to notes of properties of size and location. However, Brother argues that Michelangelo's interpretation of less-defining lines is a continuation of the architect's "graphic tendency" from his figure studies, where embedded motions and gestures are seen inherently with the graphic itself. "A profile", as the author points out, "also has an inherent ambiguity of positive and negative qualities that Michelangelo fully exploited" (p. 67). What is useful and linkable to his own design is selected and therefore extracted. These useful and linkable graphics are "intentions" to be explored in the following section.

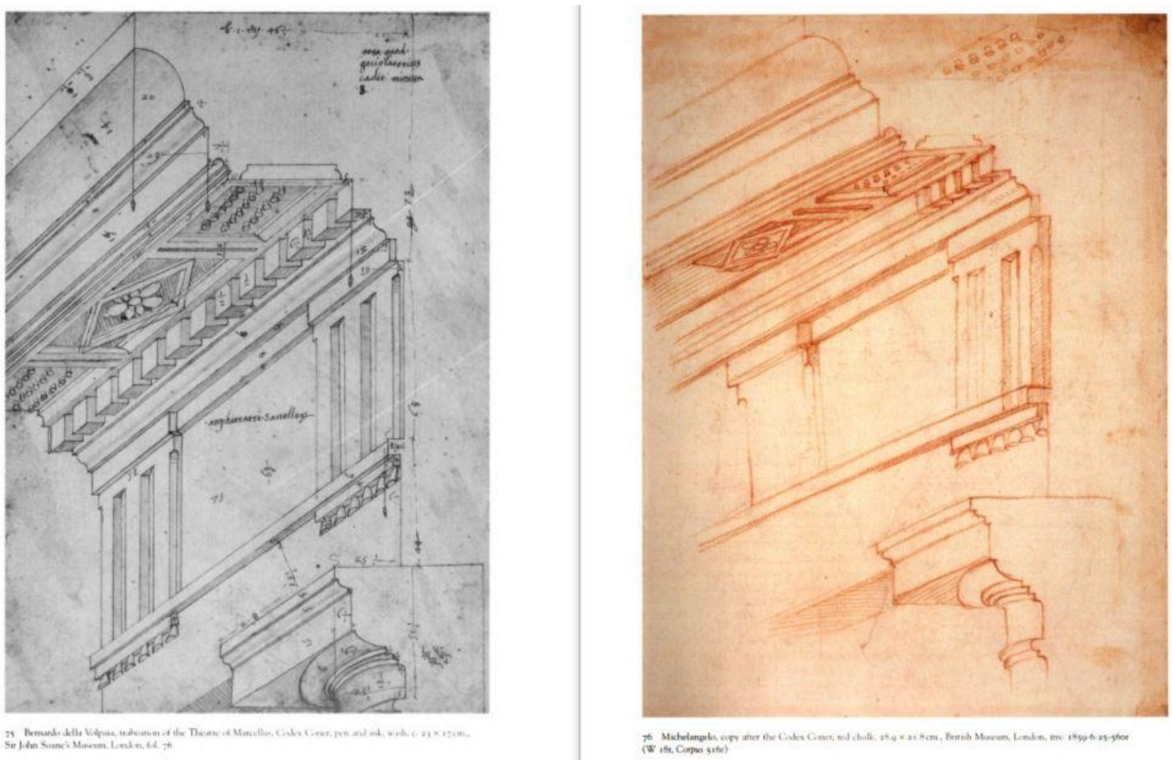


Figure 17. Bernardo della Volpaia Codex Coner's documentation (left) and Michelangelo's study of the same motif (Brothers, 2008).

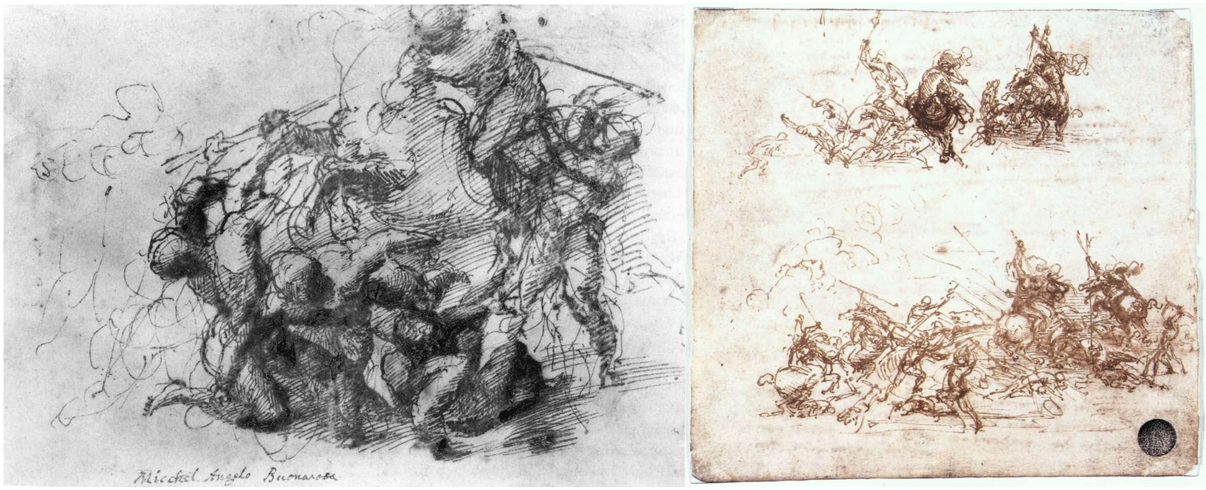


Figure 18. One sketch from Michelangelo's study of the Battle of Cascina (left) and Da Vinci's sketch study of the Battle of Anghiari (right) (Brothers, 2008).

Leonardo Da Vinci's influence in visual experiment using sketches also shifted the use of visual imagination from recording of object properties to free explorations based on perception and calibration. Most importantly, a visual process of extracting useful representations for foreseeable choices of "next line to be drawn" based on observations. In Da Vinci's sketch study of the Battle of Anghiari and Michelangelo' study of the Battle of Cascina (figure 18.), movements and collisions among figures and gestures of the whole composition are loosely applied. Multiple possibilities of posture and combination of these figures thus become visible according to how the artist will see them as the next step when interpreting the curves and shades. Michelangelo's habits that he developed for drawing to search for figure poses affected his later interpretation of architectural drawings (Brothers, p. 62). Selective abstraction of the descriptive graphics serves as indications of "design." Retraction from the descriptive drawing makes the visual focus to work more efficiently and therefore facilitates visual imagination, seeing "more" than objects but to gestures, movements, layers, and jointly, to intentions.

As a polymath, Da Vinci's inquiries on the human eyes are as fascinating as the use of them in artistic pursuit, suggesting that anything that "reaches the eye through this central line can be seen distinctly" (Hunziker, 2006). This "central line" (figure 19.) was latterly studied extensively in modern science as "fovea vision" in different from "peripheral vision." The distinction between perceptive level at these two visual areas is also relatable with the recognition of labels and perception of "intentions" in a visual clue.

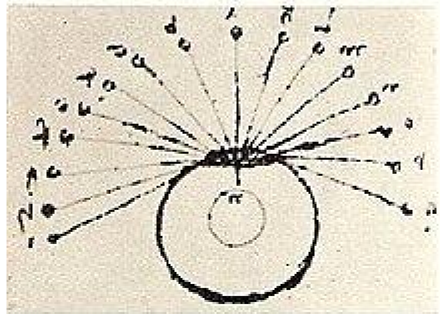


Figure 19. A sketch by Leonardo Da Vinci on the "eye line of sight" (Da Vinci, 1955).

To the discussion of "intention", what is more important than the sketch technique and the sketch themselves are the designer's observational ability to acquire and imagine the dynamics of possibilities, while appropriately recessing from identifying what exactly every part in the drawings is. A process or method of observation facilitates graphical clues in the retrieval of "intentions": the observer's personal language obtained through observation.

Using the example of Michelangelo's interpretation of the entablature of the Theatre of Marcellus, how the architect's abstraction of the original drawing helps to explain a designer's observational method. After the studies practiced by Medieval craftsmen and the progression of more durable rag paper, the architect was able to perform slight transformation of the original profiles and contours into abstract forms that suit his ways of observing design intentions----"Incorporating forms into a personal language". Such is how Brothers (2008) describes Michelangelo's attitude towards his copying of the measured drawing by Bernardo della Volpaia, "Codex Coner" in 1514.

As a method to "seek the best examples and improve upon them", Michelangelo only took models and parts he deemed worthy visually and filled the rest with imagination. Brothers also notes that imagination came from multiplicity of readings latent in an abstract profile, or expression of edges but without measurements and origins of the details, which, more specifically explained by the author as "expressing a quality apart from an object." From the original drawing, the architect saw volumes, connection among parts, repetition of elements, and contours of the whole composition to inform depth and alignments.

These abstract qualities of the profile add another layer of a figurative description of the drawn object, rather than deduce the description to be an abstracted product. As Ackerman (2016) comments: "As in the fortification of drawings, there are lines that have nothing to with structure, representing in this case visual and conceptual axes" (Ackerman, p. 15). Brothers also states that "Michelangelo's tendency to favor the profile to the exclusion of other aspects may have depended on the abstract quality of the profile...A profile also has an inherent ambiguity of positive and negative qualities" (Brothers, p. 23). Thus, his drawings made another level of visual interpretation for architectural drawings, and furtherly enables "Selection of subjects that suggests a link to his designs" (p. 24).

Michelangelo's example enables retraction from pure objective description and indications of compositional visual dynamics, in order to serve as a starting point to search for imagination (Brothers, 2008). Vision connects what is objectified to representations of design, and towards a form that internally retains gestures and movements. The architect's work suggested a more distinctive approach in the works of modern pioneers. Design intentions have become not only a manifestation of architectural ideology, but a justification of the power of visual perception beyond objectivity towards movements and dynamics.

2.2 Activating Forms in Abstraction in Architectural Design Drawings

> *"Each day of my life has been dedicated in part to drawing. I have never stopped drawing and painting, seeking, where I could find them, the secrets of form."* —Le Corbusier (Pauly, 2018).
>
> *"Le Corbusier used drawing to imprint images on the mind. [His] method of drawing on the Acropolis in particular, fast, excited, exploratory sketches served to actively propel his reasoning forward more than merely recording and memorizing fixed conclusions."* -- Geoffrey Baker, 2001, p. 182.

From the masters of the Renaissance, the sketches, as one of the first forms in abstract visual communication, have been succeeded by generations of architects in education and practice alike. "Training of the eyes" is an important realm of all designers of both the classists and the pioneers. The trust and practice of the power of vision and envision have been silently passed down to the era of machines and industries. In this second section of chapter two, I discuss exemplary architects' sketches and how visual intentions are generated and retrieved from these "rough-and-delicate, loose-and-disciplined beauty" (Merrill, 2010, p. 13), in other words, the abstract representation of intentions in design images.

2.2.1 Le Corbusier: Elements of Evocation

Being directly personal and self-reflective (Harris, 2015, p. 1), the self-claimed crow-like architecture pioneer used sketches to facilitate, improve, and shape his ways of thinking on observation, turning records into innovation. His sketches were "often minimal, fast and emotive with which he perfected both his analytical and design ideas" (p. 1). These powers of "schematic evocation" were usually done with a few quick lines. As early as his "Grand Tour" through northern Italy, the young art students developed a "repertoire" of "added analytical sketches that captured the core of spatial forms and became a means of shorthand visual note taking" (Brillhart, 2018). His drawings are deemed as a way of thinking and inquiry beyond objectivity to the composition and dynamic of forms, as Giuliano Gresleri (2000) commented: "...his awareness of 'being able to begin again.'...The notes, the sketches, and the measurements were never ended in themselves, nor were they a part of the culture of the journey. They ceased being a diary and became design."

Discussed thoroughly in Harris' investigation of Le Corbusier's sketches, these lively, abstract lines and gestures serve more than tools of memory but as "actively investing[ation]" (Harris, p. 2), to "transform his thoughts into original interpretations of the essence of classical architecture and discover its new relevance to twentieth century architecture." Unsurprisingly, this approach of translation "memory" into "thinking" has been a key point for Michelangelo's personal interpretation of Codex Corner and thus catalyzed the first architecture design drawings. More specifically on the quality of these evocative sketches, the architect's focus gradually formed to architectural qualities such as "the relationship of architecture and ground, and the representation of movement and spatial sequence" (Harris, p. 4). The two following sketches and analysis diagrams (figure 20.) by Harris show how one sketch, although seemingly as a "recording of a scene", can be seen and understood by same-minded designers as three distinctive representations of movement, depth of field, and "balanced asymmetry of contrasting graphic techniques" (Harris, p. 14).

Figure 20. Two examples of Le Corbusier's sketches and potential diagrams drawn after observation (Harris, 2015).

2.2.2 Louis Kahn: Form and Design

"Form is 'what', Design is 'how'. Form is impersonal. Design belongs to the designer."

---- Louis Kahn, 1960.

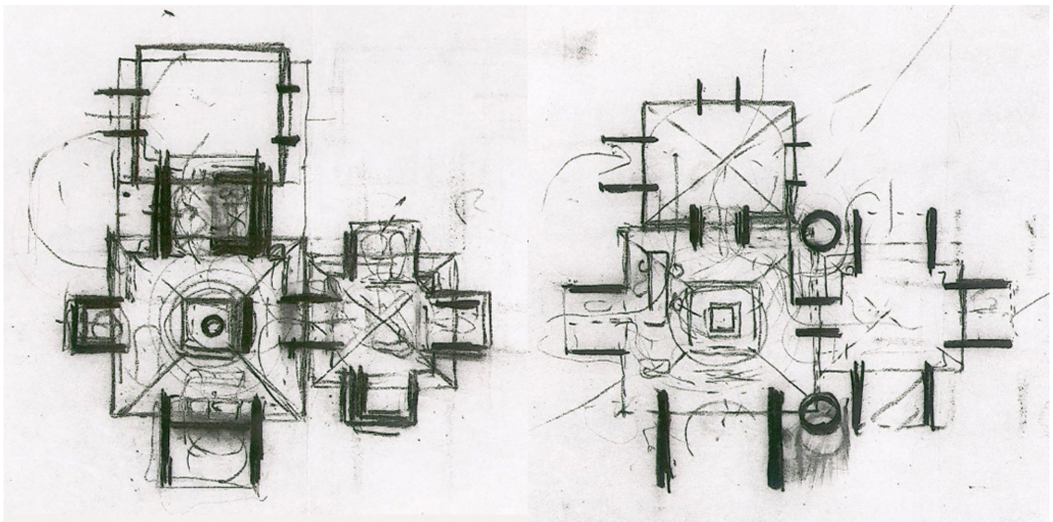


Figure 21. Two sketches by Louis Kahn: Plan studies of Morris residence. (Architectural Archives, University of Pennsylvania, 2021).

Producing hundreds, if not thousands, of plan sketches through his design career, Architect Kahn drew in an expressive, yet diagrammatic way, linking definite shapes with emergent forces on paper. "Kahn's insight into the architectural process as one uniting intuition and ratio, Platonic idealism and realism…representation and its transcendence…" (Merrill, 2010, p. 30). His sketches were about experimenting with the "Form", or in Merrill's perspective (p. 30), a defining process about the "hierarchical and reciprocal relationships between its activities and was thus the architect's insight into the unchanging essence of the institution in question" (p. 30). The expression of the "Form" is not restrained by objective design contingencies such as budgets and materials, but to demonstrate and to flow. Essences of space making, in Kahn's eyes, travel beyond the frames of programming and even scientific reasoning. They instead were given to and initiated by "intuitive powers", which is "probably our most accurate sense" (p. 29). Investing greatly in his sense of "appropriateness", Kahn advocated for trusting and exercising a process of architectural design that took roots from human and historical inspiration.

Norberg-Schulz stated in his discussion of "Aesthetics" that form and expression of an artistic object has no conflicting cause, and thus "expression belongs to the form" (p. 73). Instead of obtaining expression from elsewhere, the form itself facilitates the expression. From the two sketches above, although the compositional elements are almost the same (squares and rectangles in a series of cross-shaped arrangements), by selectively darkening certain edges Kahn presented two possible distinct intentions: one on the enclosure of spaces (left), and another on the progression of connective spaces (right). By giving graphical clues on what to see when seeing, the architect suggested his intentions of space-making in those loose lines and their compositions.

2.2.3 Ludwig Mies van der Rohe: Composed Dynamics

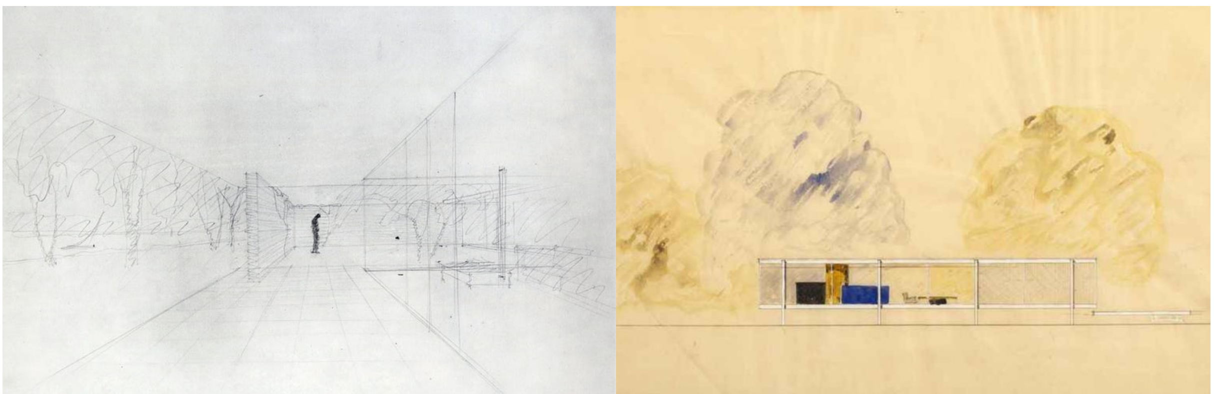
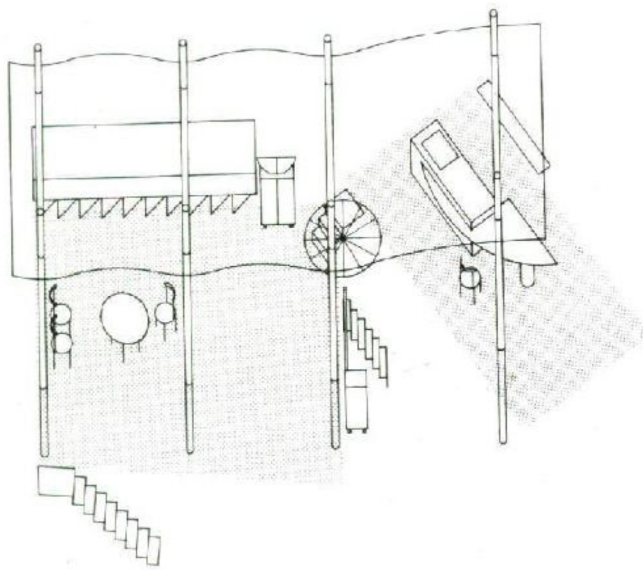


Figure 22. Barcelona Pavilion sketch (left) (Mies van der Rohe); Farnsworth House, Plano, Illinois, Elevation, 1945 (right) (MoMA Mies van der Rohe Archive, 2013).

Appraised as “the Master Composer” (Artemel, 2021), the architect earned this grand adoration by his “restrained-yet-powerful” composition emerging from collages and drawings. Formerly worked at his father’s stone-carving shop, his use of lines and planes in drawings inherent to a disciplined expression from a builder, and a space-maker. Although the objectivity of planes and spaces are much more recognizable and formalized in these drawings, a sense of a composing moment varies depending on where and when an observer chooses to focus on. For instance, if one focuses on the vertical planes in the first sketch of the Barcelona Pavilion, it will imply about a progression of transparency and a distinction in texture and massing; if one stead focuses on the highest contrast, human figure, then the drawing is about the convergence of spatial movement to a defined, one-point perspective setup. Similarly, the two horizontal stripes in the second drawing of an elevation of the Farnsworth House hint about the extent of longitudinal mass; and the tri-colored rectangles suggests interior and subdivision of spaces restrained by the previous white stripes.

2.2.4 Contemporary (after postmodernism): Diagrammatic and Parametric

Kazuyo Sejima's diagram for project Platform I is assembled out of geometric forms. Toyo Ito wrote in his commentary on the work of architect Sejima, "You see a building as essentially the equivalent of the kind of spatial diagram used to describe the daily activities for which the building is intended in abstract form. At least it seems as if your objective is to get as close as possible to this condition" (Vidler, 2000).



2

Figure 23. Kazuyo Sejima's diagram, for project Platform I, 1987 (Rodriguez, p. 382).

As a new star shining in the once male-dominated realm of architecture and space-making, Kazuyo Sejima is undoubtedly one of the most skillful seekers of new lifestyles with architectural ingenuity. Her use of visual medium not only justifies spatial arrangement, but also indicates other "sensual potential" (Rodriguez, 385) and textures. The drawings themselves are rather objectified and synchronized with actual, physical imaginary of the space to-be-made in an inspirational manner to provoke multiple sensers at once. In this case the "intentions" are less ambiguous as probing a thinking process, but leaning towards connectivity to real-world, tangible objects that actually compose the story composed by curves and enclosed shapes.

Undeniably popularized by its affinity to the computational approach of generating and manufacturing buildings, parametric design has been a force to absorb internal attributes of objectivity into the dynamics and accessibility of digitized design and construction interfaces. Extending further into contemporary parametric design that has been widely accepted in the current architectural industry, "internal logic" (Jabi, 2013) of objects is ascribed and manipulated as attributes of a recognizable object. Design representation is therefore more of a reflection of internal forces of identifiable objects and mathematized logics to connect these components. Visual powers of intentions that was heavily relied on architects' personal abilities has been set to the background, seen as a reminiscence of "instinct" and "slow" compared to the computational power of machines.

While architects have partially shared their visual power with the machines for reasoning and efficiency, how machines actually perceive or utilize such visual power is also an interesting topic to be explored.

2.3 In the Eyes of Machines: Shifting Perspective

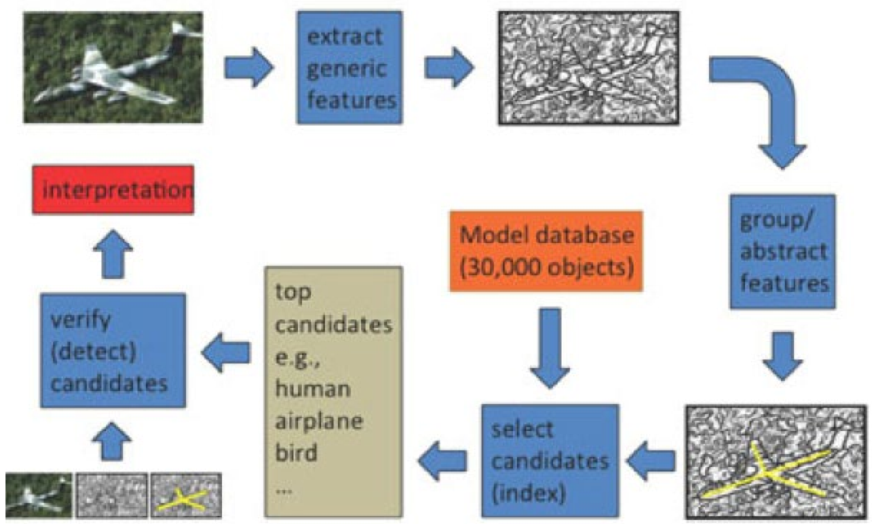


Figure 24. A classical image recognition model of extracting shape features. Proceedings, 4th Mexican Conference on Pattern Recognition (MCPR), Perceptual Grouping using Superpixels, 2012, S. Dickinson, A. Levinshtein, and C. Sminchisescu, p. 14, Fig. 1

As discussed in chapter one, machine vision has gained a systematic approach to how and what image signals could be translated into digitized expression that synchronizes with computational frameworks. Usually, these frameworks are established on detecting "shape features" (Dickinson et al., 2013, Chapter 1) as a right-or-wrong response to recognize a specific category from a database of candidates, as the targets to be seen as verifiable, solid, and definitive entities. To summarize, from the 1950s to 1970s just in the span of two decades machine vision was developed from a test of algorithmic analysis to a system of retrieving visual information of the real world: to distinct and describe as realistic as possible. The first "pattern recognition" algorithm separated points of different properties into groups by linear borders, and MIT's curriculum established for "Machine Vision" with the aim to understand real world objects empowered by "low-level" vision tasks such as edge detection and segmentation. The answer is usually "right" or "wrong", as similar to the judging point of "matching" or "not matching" a design criterion in Eastman's design cognition system[6]. "Segmentation" has therefore been one important core of Machine Vision. Mathematical criteria are implemented to examine whether a trained machine will perform the tasks of identifying objects in a least erroneous manner. Its fall back, however, has been a "weaker, domain-independent shape prior" (Dickinson et al., 2013, Chapter 1) as contrary to the grouping effects in human visual perception.

However, as designers, we face another question if adding machines' visual power to our bags of tricks: can machines give answers on an abstract level? The answer can be partially found in recent explorations in the form of "extracted" abstractions of shapes, from photos of objects, as a partial response to the abstraction enabled by human eyes when observing images (Dickinson et al., 2013). The following discussion will focus on three examples with abstractness and machines.

More importantly, before discussing any disparities between human and machine vision it is crucial to acknowledge that from conflict comes mutual understanding. "To understand machines, we need to become machines first[7]". To inform the limitation means exploring unseen opportunities. The following examples provides not only some insight of machines' processing power on "abstractness", how they are similar or dissimilar to that mastered by human architects, but also three inspirations that inspired the eye-tracking study I conducted: how to express "verifiability", a time-sensitive description of enclosed shapes, and an image-to-image translation machine learning algorithm known as the image-conditional Generative Adversarial Networks (GANs).

[6] Prototype of BIM, discussed in chapter 1.
[7] One quote from Prof. Randall Davis from his MIT 6.835 Interactive Multimodal Interface lectures.

2.3.1 Abstract Perception of Shapes in Machines: is It Verifiable? Is It Dynamic?

Figure 25. “Black, amorphous blobs”, when grouped and perceived together, forms a scene of a horse and a rider. (Street, 1931, p. 55, Fig. 8)

Machines perceive, or have been encoded to perceive, abstractness based on individual frameworks of finding and matching shapes from predetermined search spaces. This design of searching is partial similar to human perception in the sense of finding meaningful graphics. When looking at a particular visual composition, like the famous example of “a horse and a rider” (Street, 1931, p. 55, Fig. 8), the human visual system is able to “group the fragments to form a set of abstract parts, then group those parts into larger configurations, then ‘queried’ your visual memory for similar configurations” (Dickinson et al., 2013. Chapter 1). This grouping effect from quick, multiple visual perception to form a composed meaning from a given stimulus is called “perceptual grouping.” The related research works had been thriving in the latter half of the 1990s, however it has been declining due to the conflicting operation required by object detection and segmentation from a large database in machine learning (Dickinson et al., 2013). Resulted uncertainty from the composing process to reformulate image information from components has not justified to be effective in detection problems for a particular target object.

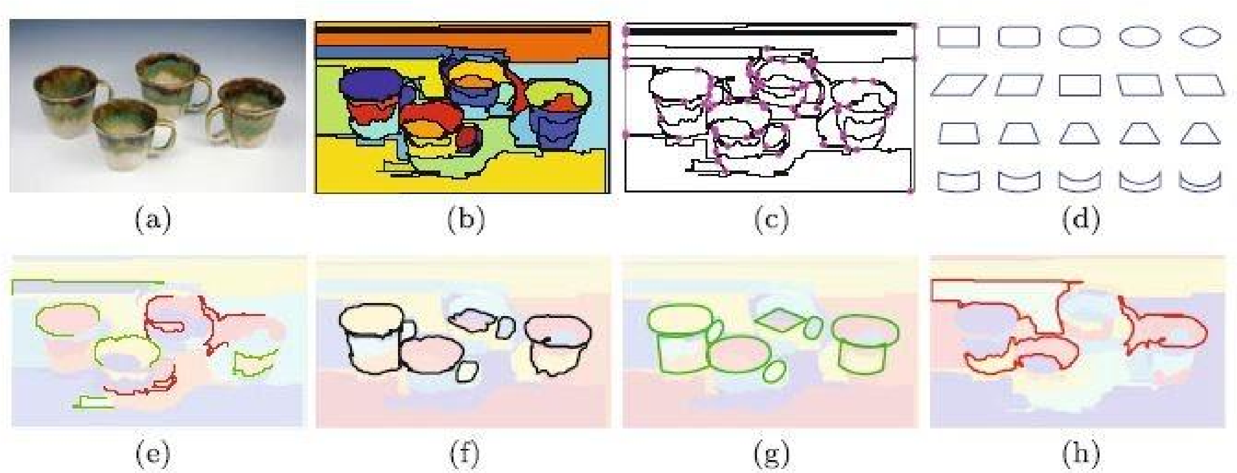


Figure 26. Detection of abstract forms from photographs. Springer Science+Business Media: Proceedings, 11th European Conference on Computer Vision (ECCVC), Contour Grouping and Abstraction using Simple Part Models, 2010, P. Sala and S. Dickinson, p. 606, Fig. 2)

However, such grouping and search are also different from human perception due to its emphasis on individuality and distinctive feature when composing a reliable answer. To follow the aim of fidelity and completeness of machine vision, one of the strongest disadvantages of composing abstract patterns done by machines comparing human designers is the "verifiability" from recombination of observed patterns. To achieve a certain degree of any verifiability is to regulate solutions search spaces into a collection of describable features or units. These search spaces excluded any furthering inquiry and extension of the abstract shapes. The above image showing retrieving primary shapes from a photo heavily focuses on the "objects" that provide the shapes, which are fairly decisive in the photo provided. The collection of abstract shapes in the "vocabulary" (Sala & Dickinson, p. 606) are enclosed, well-defined shapes with distinct features and parametric maneuverability, such as symmetrical rectangles with different levels of smoothed corners, and a series of trapezoids of different ratios between bases. To this extent the abstractness in the segmented result is still considered as labels that can be searched through data space and possess little contextual meanings with the objects from which the shapes are conducted. Identification of objects in a clean background, real-life photos, the extracted shapes are only verifiable by the predetermined database as individual "most similar" to a pattern template. An area in the photo usually is identified with one label and then such label becomes a verifiable segmentation that exclude an area from other parts in the photo. What's more, labeling areas with one system limits the reading of other graphical opportunities: for instance, the facing of the cups in photo (a).

This example of segmenting image regions and composing informative parts into potential matched shapes suggests an approach to allow a variety of different levels of graphic component to be grouped and perceived together. Next example will provide another insight to the machines' eyes.

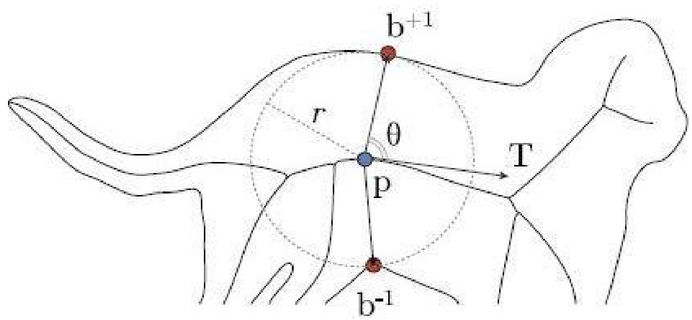


Figure 27. Example of medial representations (Siddiqi & Pizer, 2008)

Another system of describing abstractness in machine algorithms is the medial Representations introduced by Blum in 1967. Unlike the previous edge-detection approach to identify and segment shape from a given scene, the medial representations establish an imaginative skeletal network and a system of inscribed disks centered on p and their corresponding object angles (θ) and radii (r) (Rezanejad & Siddiqi, 2013). This representational system considers both the interior and boundaries of an enclosed line drawing, and therefore describing the contours from within, by the assistance of a sets of potentially moving points (the centers of disks). The circles which determine every pair of boundary points (b^{+1} and b^{-1}) can be furtherly expressed as a moving object following the skeletal structure to "fill" the interior of the shape.

This example in the "rolling" movement of a point on a set of frames provided another inspiration to the eye-tracking study.

2.3.2 Image-Conditioned Generative Adversarial Network

Image-to-image translation has become more and more frequent with recent development in the machine learning realm, especially of designing and training convolutional neural nets. In architecture design, generative networks that translate one graphical representation of design into another have been adapted to floor plan prediction (Wu, 2020) and footprint (figure 28.) and furniture layout (Chaillou, 2019; Huang

& Zheng, 2018). These applications have been focusing on clean plan drawings where each sublevel design unit such as beds and tables have distinct graphical representations. Their commonality is to use color-coded blocks to indicate matching areas with the original architecture drawing[8].

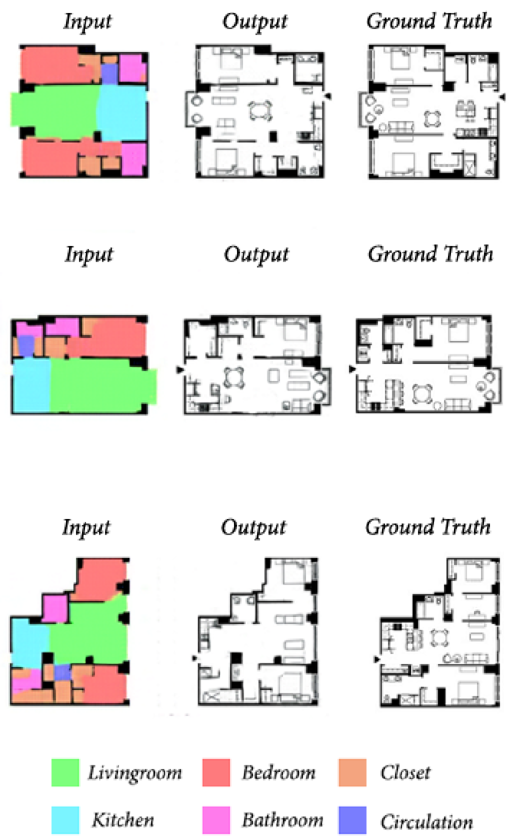


Figure 28. Chaillou, S. (2020). Selected generated floor plan with furniture.

The major distinctive feature of conditional Generative Adversarial Network (cGANs) is to use a discriminator to classify between real and fake image pairs. Conditional GANs are “generative models that learn a mapping from observed image x and random noise vector z to output image y, G: {x, z} → y” (Isola et al., p. 3). Recently developed in 2015, this algorithm was designed to allow a more general solution to image translation problems in computer vision and image processing. As the authors states, this algorithm provides a “common framework” (p.1) to predict pixels from pixels by minimizing loss

[8] Architectural drawings in these examples are computer drafted plan drawings compared to architects’ design sketches.

through a high-level goal, such as "making the output indistinguishable from reality" (p. 2), in contrast to other task specific convolutional neural nets (CNN).

2.4 Expressing Intentions Through a Mechanical Process

Gradually shifting from the trained designers' willful perception to the algorithmic search of clean forms, a gradient of human and machines' perception of given graphics stimulus can be formed. Human designers, especially of greater skills and innovative visual powers, are more likely to use dynamic, yet undefined graphics to assist further iteration of design forms. Their personal experiences, whether of Michelangelo's sculpture practices and Corbusier's grand tour in Northern Italy, produced a large collection of visual data and thus is able to be searched or associated with new abstract stimulus such as the gestures of figure study and quick lines that indicates spatial dynamics. To the machines, however, it is hardly possible to code the processing power of a trained human design completely into definite pipelines. What is possible, is to explore what kind of forces might help to bridge the two sides: one of the experiences and one of the experimentations.

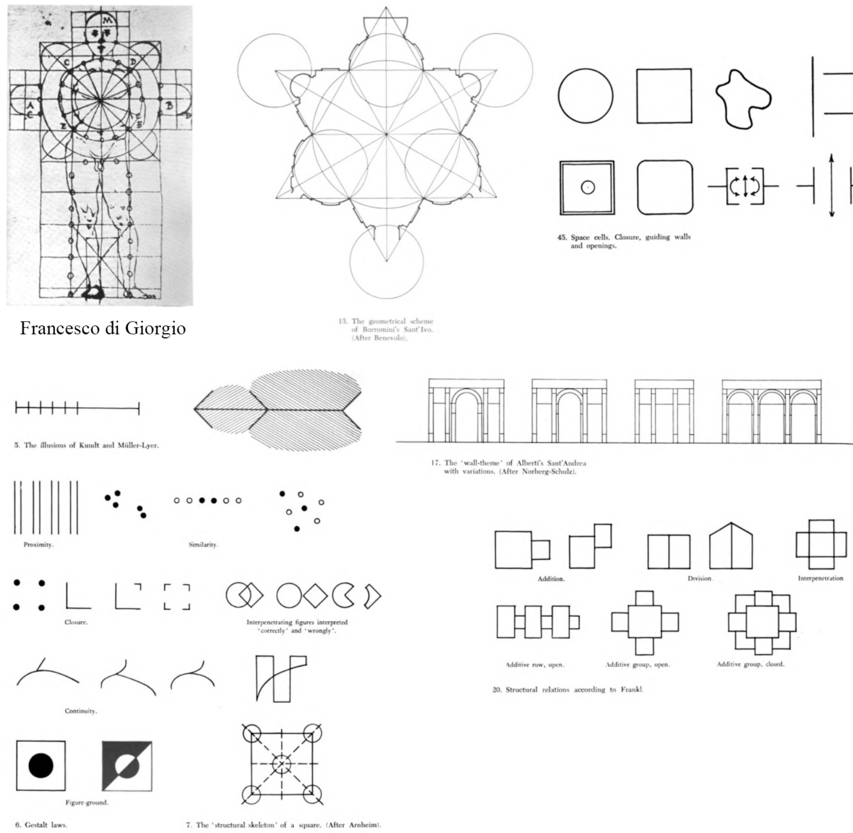


Figure 29. Diagrams for delivering of architectural intentions (Norberg-Schulz, illustrations).

One potential bridging force, therefore, can be found through examinations in the design diagrams, for they are decisive graphical expressions made by trained designers for the goals of guiding or describing design ---- or design intentions. More importantly, they are results from active observation or visual imagination. In these design diagrams by Norberg-Schulz (figure 29), labels and intentions may be fairly distinctive not only because they are represented visually, but how to think based on each in the design process are determined by two mindsets: perception and creation. Previous discussion on the parametric design logic, which categorizes objects into atomic "design units" (Eastman, 1965), synchronizes with conventional computational approach to "perceive" and process visual stimuli as variables which can be identified and evaluated. However, objective attributes are only partial of an entity under the scrutinization of designers. The perception of design intention based on abstractness is another part of the creative design procedure that requests a different level of perception and imagination. "Dynamic observation", as something not easily described by identifying properties or choosing languages, is attached to human consciousness to extract and compose rough or ambiguous perception into analysis and prediction of the potential within an image, sometimes in an even rule-less manner. To capture it while it is active, we must capture the dynamics of its perception.

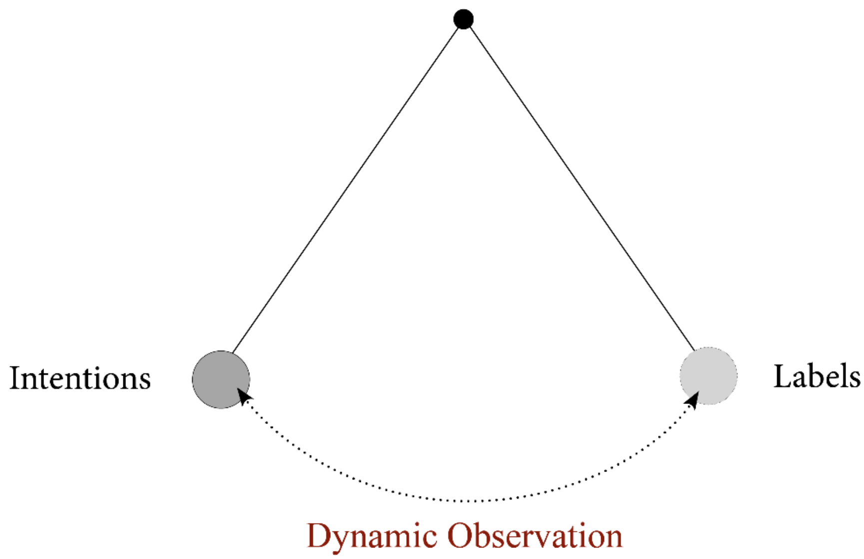


Figure 30. A pendulum between the perception of intentions and labels, swings by the force of observation.

Despite the pendulum of design mindsets that swings between these two ends, the former (label) is usually seen as more rational and mechanical, while the latter one (intention) being sensory and intuitive. However, the artificial line drawn between logic and instinct would be much blurred if taking the answers of "how and why designers see" into consideration when describing and composing design representations or structures. Dynamic Observation, drawn as a two-directional arrow that indicates a dynamic translation between intentions and labels, is tested and discussed in the next chapter for its two manifestations: one from the observational process recorded via eye-tracking, another from the design representation obtained by analysis of both eye-tracking result and visual stimuli.

Using eye-tracking to decode design intentions in architectural sketches is an attempt made here to reverse the conventional roles of "instinct" and "logic": to utilize machine-captured human observation patterns as mechanical logic behind visual interpretation of design intentions; and design sketches themselves as evocative visual representations seen be trained professions. In the next chapter I describe human behavior as a series of captured Euclidean coordinates and scan paths to seek a formal graphical representation; and give uncertainty and interpretation to design sketches that are drawn to be seen not as attributes of objects, but as “observation” that facilitate design thinking on the level of “intentions”. In such an exploration and examination of human behavior and design thinking I aim to contribute to a more sympathetic understanding between designers’ visual power and machines’ power in representation. More importantly, to inquire about the collaborative future between humans and machines when their vision’s power becomes both perceptive and creative.

Chapter 3 Interpreting the Eyes: Architectural Design Intention in Eye Tracking

This chapter consists of three major sections: Design and logistic behind the eye-tracking studies; documentation of the eye-tracking studies on observing three expressions of architectural intentions from Louis Kahn's plan sketches by MIT architecture graduate students; and analysis of the resulted heat maps from eye-tracking studies.

In this chapter I demonstrates three major implications respectively: affinity between mechanical eye movements and the observational goals of architects; goal-oriented observation results in different interpretation of a single design sketch; and view patterns reflect meaningful design information visually.

3. 1 Logistics Behind Eye-tracking for Design Cognition: Why Use Eye-tracking to Reveal Architectural Intentions?

Eye-tracking, or more specifically gaze-tracking for fixation time and viewing sequence, bonds mechanical movement of visual observation to another visual medium such as sketches graphically and takes account of individuality in observation processes. As discussed in the previous two chapters, how designers observe and understand visual stimuli from a given sketch has been a consisting exploration of practitioners of many realms of visual thinking. While many design studies focus on the creative methods and theories of observing and thus the tangible products in forms of diagrams[9], human behavior research projects emphasize on the physical process and relations between mechanical movement patterns and information retrieved according to spatial and temporal sequences (S.Djamasbi et al., p. 308). However, what has not become a topic is that by what process or medium might bridge, or facilitate a transition between, the creative and mechanical process of designer's visual power. If machines, in their nature of Boolean operation, can be furtherly elevated into a level of human-like creative perception, an examination on how these two processes may converge and be translated into each other is meaningful.

Human creative perception is multi-layered, made possible from various goals when observing, and difference in individual experiences. As Christian Norberg-Schulz (1963) described in his book "Intentions in Architecture", how the forms in visual representation will be interpreted differently according to personal inclination and training. "Intentions" in general are perceived from an active

[9] "Aesthetic preference" in works Arnheim, 1988.

process, which goes beyond "simple classifications" (Norberg-Schulz, p. 31) to a "greater 'intentional depth'" to "grasp the situation". For extracting different expressions of intention in a design sketch, therefore, I propose to also give thought to the physical properties of eyes and seeing, or at least to how designers like Leonardo Da Vinci used to interpret them.

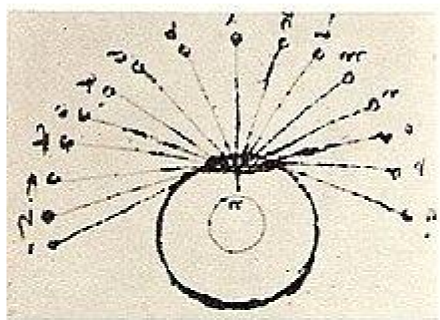


Figure 31. A sketch by Leonardo Da Vinci on the "eye line of sight".

How mechanical movements relate or facilitate graphical recognition of visual design stimuli can be interpreted through a link between the brain's control of eyes and areas of observation that reflect the brain's choice of interest. Gaze location reflects gestures of a designer's observation. Back to the masters of Renaissance, Da Vinci is speculated to be the first in Europe to acknowledge "certain special optical qualities" of eyes, most likely from his practices as an artist and a researcher in human anatomy. He ascribed two qualities of vision, and only what wa at the "line-of-sight" is clearly perceived (Eye Movement in Reading [Wikipedia], 2020). Building up on this artist's interpretation of nature of seeing, contemporary scientists define eye tracking as "an experimental method of recording eye motion and gaze location across time and task.": as Carter and Luke (2020) briefly summarized in the beginning of their review on the eye tracking practices:

> *"The origins of eye tracking can be traced to Charles Bell, who first ascribed eye movement control to the brain, classified eye movement, and described the effect of eye movement on visual orientation. This defined a physiological connection between the eyes and the nervous system, connecting their motion to neurological and cognitive processes and thereby opening a potential window into the inner workings of the mind" (p. 50).*

Echoing the power of eye movement to serve as a reflection of how a visual stimulus is understood by minds, an approach of extracting "design intentions" from graphical representation of mechanical eye movement is made reliable by how the eyes work when processing visual stimulus: "mental processing" and "decision making" reflected by what and how long the eyes are looking at. Because of the anatomy of eyes, only a small portion of images projected into the retina receives strong intention, where the fovea is located in the human visual field. Carte and Luke (2020) point out that "there is a strong motivation to move the eyes so that the fovea is pointed at whatever stimulus we are currently thinking about or processing." More specifically, it is called "the eye-mind link" by Rayner (2009) in his long-time work for understanding eye movements and cognition. According to Rayner, it is even "necessary" to move eyes in order to understand better visual stimulus due to eye structure limitation in anatomy (p. 1459). Experienced visual task performers, such as designers and architects, "have a consistently higher average fixation duration than do novices" (p. 1459). Therefore, not only the locations and durations of gaze are important for understanding what has been "perceived" from visual stimulus, but the time-stamped sequences and relationship between critical visual areas are critical to link mechanical behavior with "understanding intentions".

Therefore, in order to understand how a design image is perceived by human designers, it is crucial to measure "where, how and in what order gaze is being directed during a specific task" (Carte & Luke, p. 50). My studies therefore utilize eye-tracking technique and graphical representation of the result – heat maps – to examine the relationship between extracted visual information according to fixation time and viewing sequences (scanpaths) interpreted in architectural languages. This heatmap based graphical representation is tested in a machine learning algorithm to generate heatmap as "representation of intentions" in chapter four.

Table 1. *The range of mean fixation durations and the mean saccade length in silent reading, oral reading, scene perception, and visual search*

		SL	
	FD (ms)	*Deg*	*Letters*
Silent reading	225–250	2	7–9
Oral reading	275–325	1.5	6–7
Scene perception	260–330	4–5	
Visual search	180–275	3	

Note: FD = fixation duration; SL saccade length. The average fixation duration in scene perception and visual search can very much be influenced by the exact nature of the task that participants are given.

Figure 32. The range of mean fixation durations and the mean saccade length in silent reading, oral reading, scene perception, and visual search. (Rayner, 2009, p. 1460).

3.1.1 Fixation-based Eye-tracking and Visual behavior

Before reporting the design of gaze tracking studies, what visual behavior is particularly focused is explained in this section. Two types of visual behavior are commonly accepted by scientists: "Visual behavior in general can be characterized by fixations, and the rapid, ballistic movements that connect those fixations, which are known as saccades." Fixation is the process of retrieving information. "Most fixations tend to fall on the informative parts of the scene" (Rayner, 2009). During fixations, the eye is briefly stationary and situated such that objects of interest fall within the foveal region of the eye (the central point of one's field of vision) where visual acuity is greatest, with most intentional visual behavior being comprised of fixations" (King et al., 2019; Duchowski, 2017).

On the other hand, areas of fixation also suggested effective graphical cues to the observer. "Fixations reveal what objects are subject to cognitive processing at a given moment (Duchowski, 2017)." Often the duration of these fixations is considered to reflect one's depth of processing" (Josephson, 2005; Rayner, 1998). In Visual Search: eye movement control in visual search: "where and when to move the eyes was one question asked by Rayner in 1995, suggesting that the trigger to move the eyes in a search task is something like: Is the target present in the decision area of the perceptual span?"

A variety of fixation time furtherly assigned four different visual behaviors as silent reading, oral reading, scene perception, and visual search (figure 32.). More information on exact functions during fixations was reported in Carter and Luke's "Best Practices in Eye Tracking" (2020), various visual tasks result or require distinctive length in fixation duration. "Scene perception" and "visual search" are the two functions the plan comprehension tasks my eye-tracking study is focusing on. From the above chart it is obvious that "scene perception" requires the longest fixation time: 260 - 330ms, while "visual search" requires the shortest as 180 - 275ms. These two fixation duration ranges will be used in deciding which areas shown in the resulted heatmaps in the test result section.

3.1.2 Why Use Plan Drawings: Shapes, Composition, and Circulation

Training in Architecture, and the greater realm of design, how to observe has been one of the paramount aspects in enhancing and gaining fundamental abilities of a novice and an expert. Plan sketches, as a form of orthographic projection, are one of the most easy-to-execute and "richest and reliable source[s] of spatial information" (Koutamanis, 1995, p. 18). These sketches not only facilitate reading of spatial arrangement and formal articulation of shapes and composition, but also give clues and prediction on circulation, which is about imaginative movement within the represented spaces.

Two major characters representing intentions using plan drawings: diffusing and overlapping. As Norberg-Schulz (1966) discussed architectural forms and elements:

> *"Elements which are topologically defined have a diffuse, amorphous character, and their 'expression' merely consists in their concentration or closure...The bounding surfaces may be articulated in such a way that they 'characterize' the mass, for instance as a 'block' or as a 'box'" (p. 144).*
>
> *"An 'interpenetration' is created when two elements overlap. This does not mean that they lose their independence, only that ambiguous zones are formed, which at the same time 'belong to' both elements...that a formal separation becomes meaningless" (p. 141).*

How exactly such “characterization” and “interpenetration” are revealed through Area of Interests (AOIs) and therefore establish a system of narrative to capture[10] “intentions” graphically using diffusing and overlapping is then tested and examined in the next two sections: Gaze-tracking methodology and sample heat map results.

3.2 Design of the Eye-Tracking Tests: Gaze-tracking Methodology, Data Collection, Measures and Procedures

To provide evidence for the visual power in revealing “design intention”, I designed a gaze-tracking experiment to collect gaze-tracking heat maps from architecture students at MIT. For visual stimuli, I collected forty-five Louis Kahn’s design sketches of building plans that possess “evocative and dynamic” design information as discussed in chapter two; and possess graphical characteristics of diffusing and overlapping as mentioned from the previous section. Twenty-nine of the sketches were accessed from an online collection of Louis Kahn’s drawings from MoMA; sixteen of the sketches were accessed from digital archives of Philadelphia Architects and Buildings[11]. The detailed information and sources of these forty-five sketches can be found in the appendix.

The objectives of the gaze-tracking study were to provide quantitative evidence from the discussion the beginning of chapter three: affinity between mechanical eye movements and the observational goals of architects; goal-oriented observation results in different interpretation of a single design sketch; and view patterns reflect meaningful design information visually. At the same time, the resulted heat maps were prepared into training images for the machine learning algorithm in chapter four. Participants’ eye movements were traced and recorded in empirical data, including fixation coordinate and scanpaths. Fixation data and gaze coordinates can show where and how quickly participants examine sketches of Louis Kahn when given observational tasks on architectural design intentions: Shapes, Composition, and Navigation (Circulation).

[10] Such capture of graphical clues in intention through heat map results might be an interested parallel to Thomas Wedgwood’s capturing of silhouette images.
[11] These archives were accessed in April 2021 by paid monthly membership fee as required.

Using the gaze-tracking data, a heat map was generated for each sketch to demonstrate overlapping “areas of interests” (AOIs) on and how the focused areas progress in the 10-second window.

3.2.1 Participants

This study focused on using vision to “express” architectural intentions from plan sketches of Louis Kahn. Fifteen participants were examined, and all of them are architecture students or have had professional architecture and design education of undergraduate or graduate level. They ranged from twenty-two to thirty-three years old, and consist of seven female participants, six male participants, and two participants who preferred not to report gender identity (table 1.).

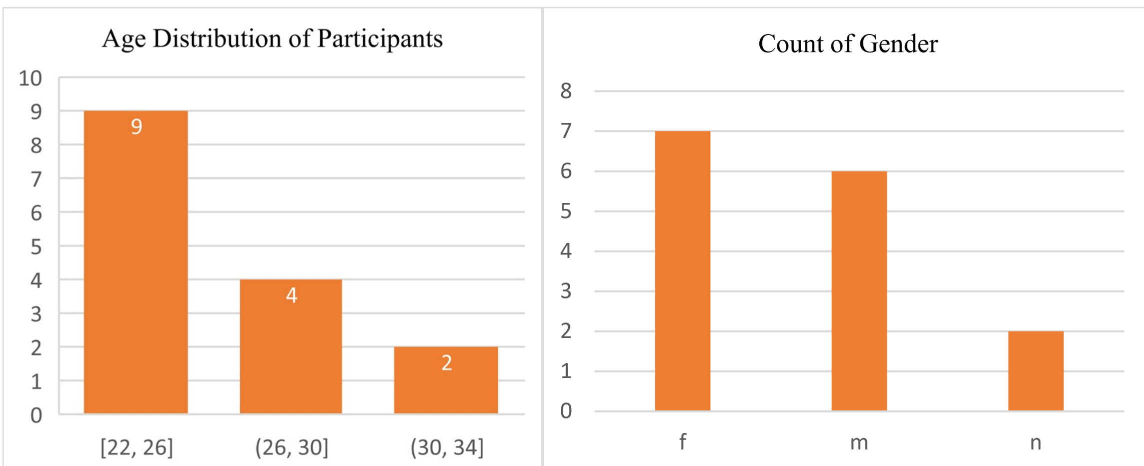


Table 1. Age and gender distribution of the participants.

3.2.2 Devices

Due to the health protocol effective at the time of the thesis, all eye-tracking studies were delivered and collected via an online eye tracking platform: GazeRecorder. The study was composed and distributed using this web-delivered eye-tracking study service, achieved through webcams on laptops screens or desktop monitors to calibrate and track gaze movement. The accuracy of this webcam-based API is ~1.0 degree, with precision ~0.2 degree of visual angle (GazeRecorder, 2021). All participants were asked to keep the head relatively still, and no operation for the auto-played visual stimulus slides.

Below are key settings of this gaze recording system; and two examples of displays used in the tests.

Sampling interval: 35ms +- 3ms, 10s total for each stimulus (512 x 512-pixel x 72 dpi, at full screen mode of laptop screen or desktop monitor.)

Total sampling of each stimulus about 273 – 280 points of eye-gaze focus (fovea vision) for each sketch.

Display 1: Dell Monitor S2715H. 27-inch diagonal display. Image display size 38cm x 38 cm. Optimal eye distance from the monitor: 38 cm.

Display 2: Apple MacBook Mid 2015 screen, 15.4-inch diagonal display. Image display size 20.5 cm x 20.5 cm. Optimal eye distance from the monitor: 21 cm.

Each participant was specifically asked to take the GazeRecorder web test in full screen mode of their browser so that the eye movement can be assigned to more detailed areas in sketches.

3.2.3 Testing Images

The components of each testing set are plan sketches of Louis Kahn in early design iterations, especially reflecting the testing and thinking process architecturally and visually. Why Louis Kahn's sketches are particularly favored are their known quality in expressing visual thinking and "incompleteness" that facilitate recognition and integration of architectural intentions: the "locus of its[architecture] making" (Merrill, p. 13). A collection of all forty-five sketches used as visual stimuli of the gaze-tracking studies are presented as the following (figure. 33):

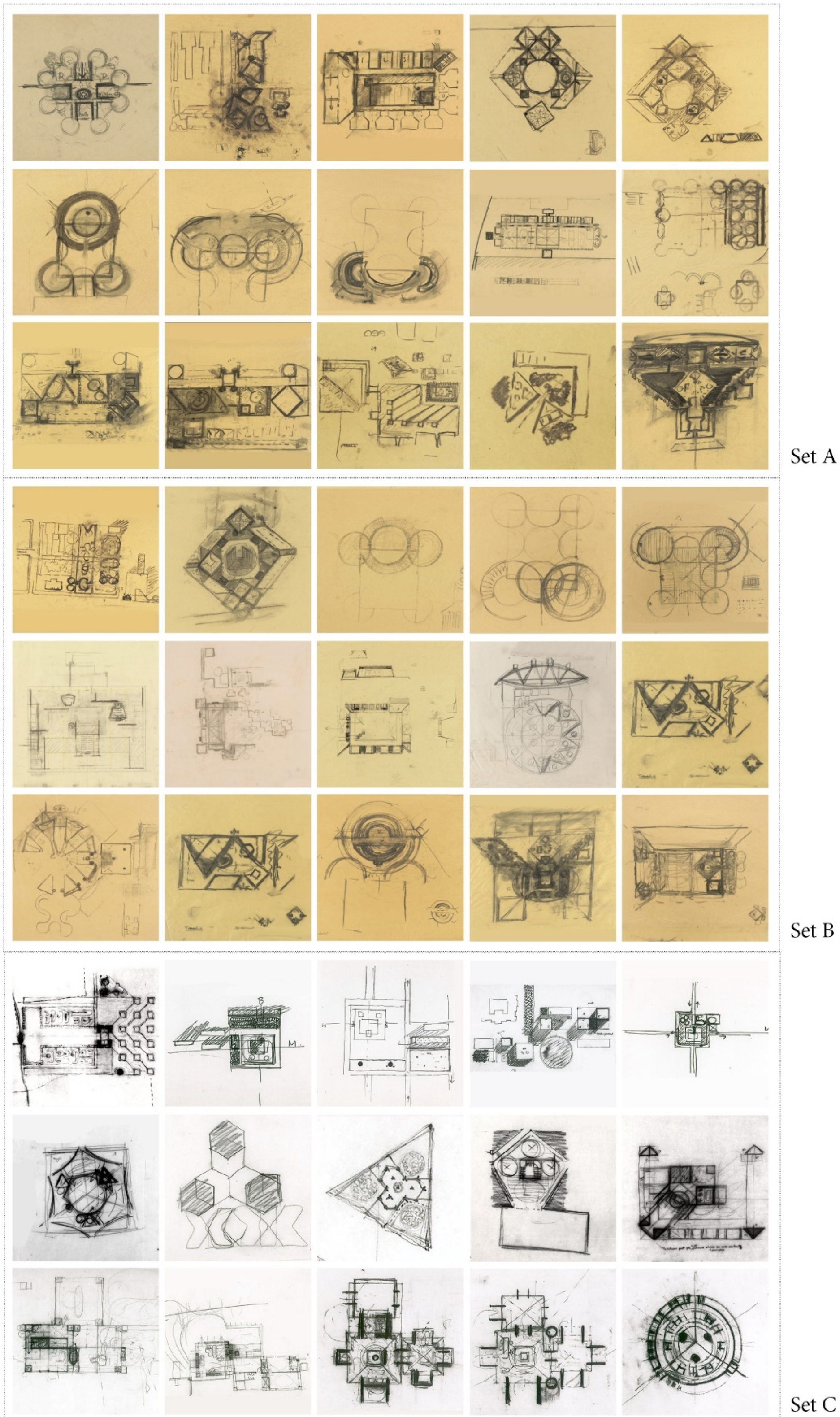


Figure 33. All forty-five sketches in three sets used in the gaze-tracking studies.

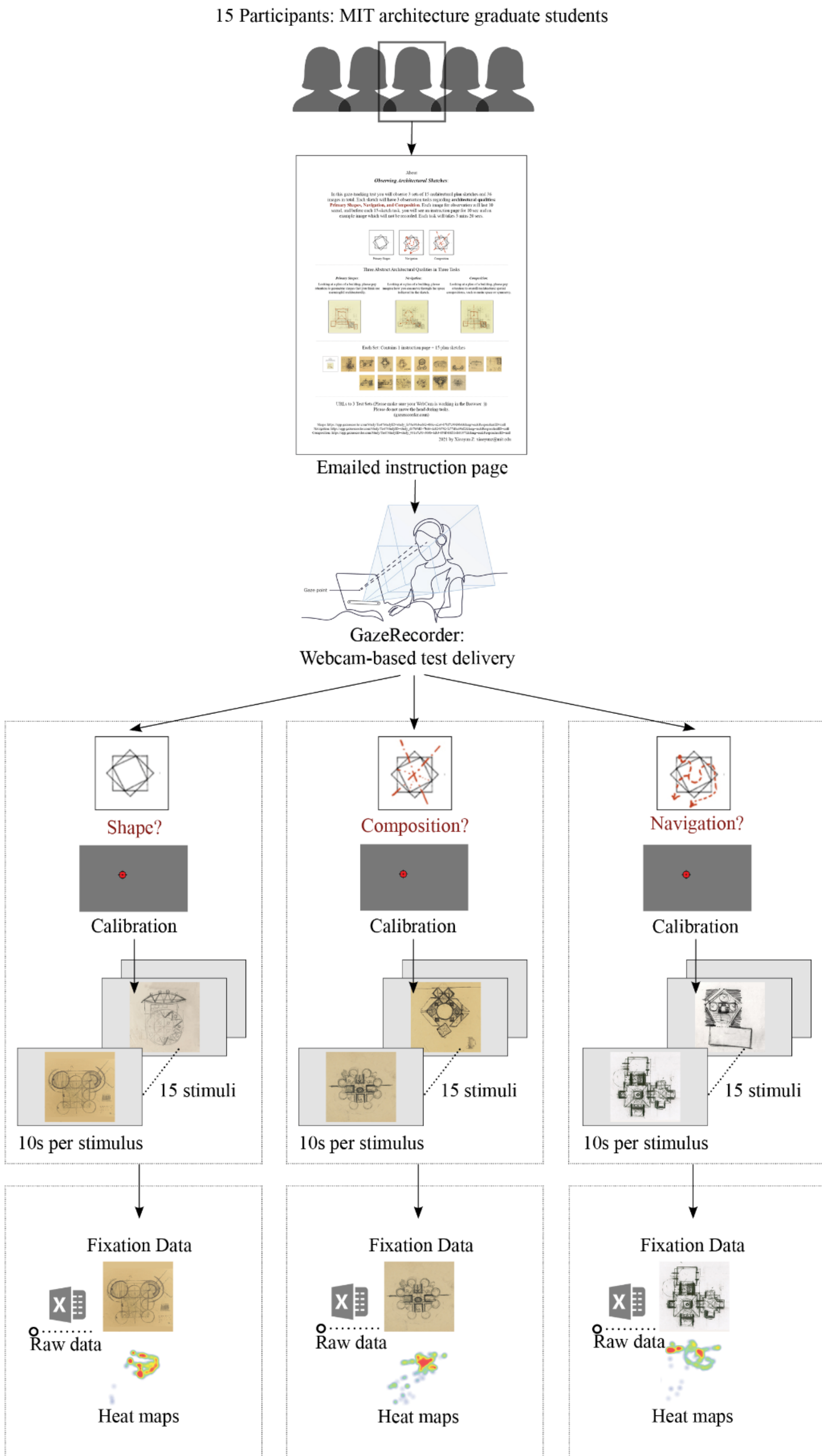


Figure 34. Gaze-tracking procedure.

3.2.4 Procedures

i. Before Gaze-tracking:

Participants were contacted by the author and asked if they would be interested in participating in the study. After agreement, the author sent out emails containing one letter-sized instruction document and three links are contained in the instruction for participants to click and start eye-tracking immediately. Key features of the instruction documents are summarized as below:

Each gaze-tracking is centered on three architectural intentions: Shapes (or Elements), Composition (or Formal Structures), and Circulation[12].

Three sets of fifteen images per set are composed, for each intention the participant was given one of the three sets randomly. Repetition of the same set but for different intention task was allowed.
Each set took 3 mins 20 secs to finish, and before each set a 10-point calibration will take about 2 min to finish. Total time required for 3 sets is 10 + 6 = 16 mins including calibration before each set. All participants had flexibility to choose when to take each of the tests.

ii. During Gaze-tracking:

Calibration

The procedure of calibration consisted of the participants following instructions designed by GazeRecorder at the beginning of each test: to position head to a proper distance in order to fit the given frame and proceed to a 16-dots calibration that took less than 2 minutes to finish. "The calibration procedure measures [participants'] eye position and maps eye movements to targets with a known position. The calibration is done by following a point across the screen" (GazeRecorder, 2021).

Instruction

Due to the health protocol of MIT and Massachusetts, the tests were distributed without any in-person contact. Therefore, instructions before each test session were shown to the participants in a presentation form, same size, and duration as each of latter sketches.

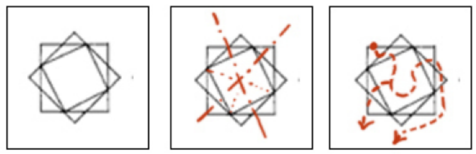
Examples of test instruction

[12] First two taken from Norberg-Schulz's discussion of Form, as "Pride" and "Study of Architecture itself". P131.

About

Observing Architectural Sketches:

In this gaze-tracking test you will observe 3 sets of 15 architectural plan sketches and 45 images in total. Each sketch will have 3 observation tasks regarding **architectural intentions: Shapes, Composition, and Navigation**. Each image for observation will last 10 seconds, and before each 15-sketch task, you will see an instruction page for 10 sec and an example image which will not be recorded. Each task will takes 3 mins 20 secs plus a less-than 2 mins of calibration time.



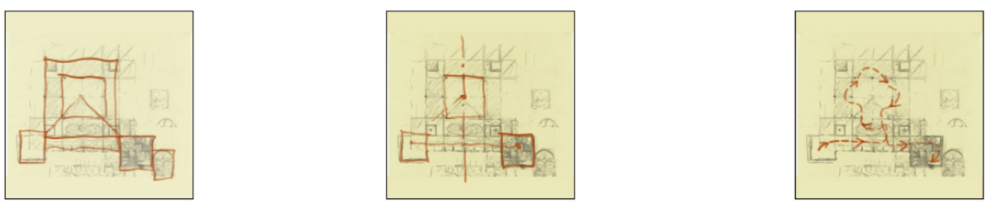
Primary Shapes | Composition | Navigation

Three Architectural Intentions in Three Tasks

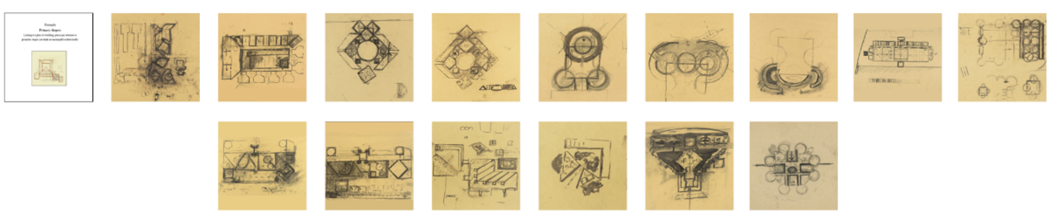
Shapes: Look for shapes that you think are architectural component, such as room, wall, column, stairs, etc.

Composition: Pay attention to spatial compositions such as symmetry and central space.

Navigation: Imagine how you can move through the space in the plan.



Each Set: Contains 1 instruction page + 15 plan sketches



URLs to 3 Test Sets (Please make sure your WebCam is working in the Browser :))
Please do not move the head during tasks.
(gazerecorder.com)

Shape: https://app.gazerecorder.com/Study/Test?StudyID=study_fe76e96d-ad62-486c-a2a4-b7bf7c99d4b6&lang=en&RespondentID=null
Navigation: https://app.gazerecorder.com/Study/Test?StudyID=study_d37b5df3-7b46-4c82-b792-fc77d0c69af2&lang=en&RespondentID=null
Composition: https://app.gazerecorder.com/Study/Test?StudyID=study_931a7c53-306b-4cb3-854f-88f1b4b8197d&lang=en&RespondentID=null

2021 by Xiaoyun Z: xiaoyunz@mit.edu

Figure 35. An example of the instruction page sent out to participants.

Instruction text is the following:

In this gaze-tracking test you will observe 3 sets of 15 architectural plan sketches and 45 images in total. Each sketch will have 3 observation tasks regarding architectural intentions: Shapes, Composition, and Navigation. Each image for observation will last 10 seconds, and before each 15-sketch task, you will see an instruction page for 10 sec and an example image which will not be recorded. Each task will take 3 mins 20 secs plus a less-than 2 mins of calibration time.

Shape: Look for shapes that you think are architectural component, such as room, wall, column, stairs, etc.

Composition: Pay attention to spatial compositions such as symmetry and central space.

Navigation[13]: Imagine how you can move through the space in the plan

Stimuli

Thereafter, the participants received the stimuli.

[13] The use of navigation was replaced by "circulation" when describing this intention category. However, in the eye-tracking tests, all participants were given the instruction in "navigation."

3.3 Analyses of Heat Map Results

This section consists of three sections of heat map analysis and their implications on how architectural intentions are focused and extracted. Specifically, to provide graphic evidence on the following arguments:

1. Affinity between mechanical eye movements and the observational goals of architects.
2. Goal-oriented observation results in different interpretation of a single design sketch.
3. Viewing patterns reflect meaningful design information visually.

The examination for the first argument focuses on the time-sensitive pattern between heat maps and sketches: I compare the mechanical pattern of gaze locations and durations with their corresponding areas in sketches, using one second as a step to find direction and discovery of new focusing area to the architectural intentions[14] the participants were asked to observe.

For the second argument, I examine the representational difference among heat maps and their overlapping areas of same sketches but different intentions. I utilize image-processing software programs such as photoshop to extract informative patterns and compare them accordingly.

For the third argument, I use extracted areas from sketches and discuss on their compositional characteristics in composing or implicating design sketches. I use the concept of medial axes to locate a skeletal system that will interpret gaze pattern movement and architectural meanings as a holistic composition.

Before the analysis, how heat map results from GazeRecorder were interpreted and processed are explained in the following section.

[14] Shape, composition, and circulation.

3.3.1 Graphical Interpretation of Heat Maps

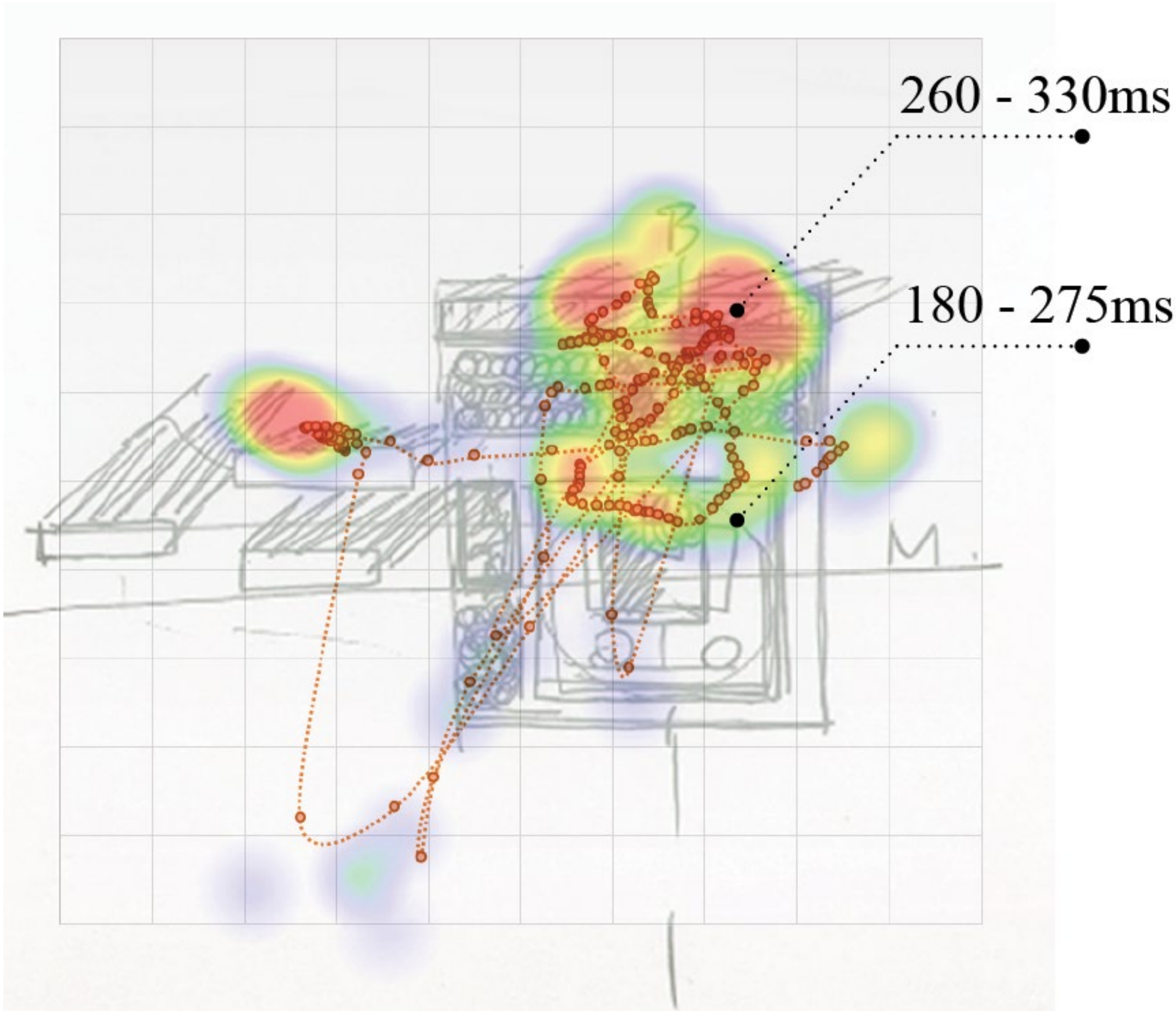


Figure 36. An example of the heat map superimposed on the original sketch.

From the previous chart on the differences in fixation duration between scene perception and visual search (figure 36), it is obvious that “scene perception” requires the longest fixation time: 260 - 330ms, while “visual search” requires the shortest as 180 - 275ms. These two duration ranges will be used in deciding which areas shown in the resulted heatmaps in the test result section. Taking 35ms as an average sampling frequency for each plan sketch, in the total of 10 seconds task time the average sampling points number will be 285. If an area in a sketch overlaps at least with 7 - 9 sampling points, then it is a fixation area. Similarly, if it overlaps at 6-8 sampling points then it suggests a “visual search” in observing the sketch. I selected 6 heat maps from each intention group and calculated the average of scene perception time and visual search time based on the numbers of sampling points that fit into the two categories (table 2.) The resulting values suggest that perception of shapes requires the most of scene perception time, while the perception of composition requires the least time in scene perception. In all three intentions, scene perception times are always longer than that of visual search. The fixation areas are furtherly analyzed for where they would overlap with original sketches.

Architectural Intentions	Avg. Scene Perception time (s)	Avg. visual search time (s)	Perception/Search Ratio
Shapes	7.39	2.61	2.83: 1
Composition	6.29	3.71	1.70: 1
Circulation	6.73	3.27	2.06:1

Table 2. Difference between scene perception and visual search time in each of the three intentions. Average times were calculated through 18 selected heat maps, 6 heat map per intention group, and 3 of each group are attached in Appendix II.

Extracting areas of different levels of interests of perceptive strength is made possible by isolating areas in a sketch which are color-coded differently in the original heat map results according to visual behaviors. As shown in table (table 3.) below, visual behavior such as scene perception are allocated to the red and yellow color range, for their sampling rate corresponds to at least 7 - 9 points within the same-colored areas; green, cyan, and blue areas are assigned to visual search for similar reasons. The distance between each pair of sample points is calculated using the original image resolution: 512 x 512 pixel and count the distance between two sampling circles' centers by the numbers of pixels.

Color(s) (RGB)	Visual Behavior[15]	Distances d between two sampling point (pixels) in 512 x 512 original image
Red	scene perception +	$d <= 4$
Yellow - orange	scene perception	$4 < d <= 9$
Green	visual search +	$9 < d <= 15$
Cyan, Blue	visual search	$15 < d$

Table 3. Meanings of color in resulted heat maps in this study.

[15] Relating to exploration and inspection. (Duchowski A.T., & Krejtz K, 2017)

Many former eye-tracking studies used heat maps stopped at recognizing where AOIs were appearing, rather than asking further about what specific graphical contents within those AOIs were supporting the interpretation of the whole image as a communicative agent for designers (Jahanian, A., Keshvari, S., & Rosenholtz, R. 2018). However, in the eyes of designers, specific graphics of the focused area render additional information on the forms and connections of architectural significance: as “characterization” and “interpenetration” which are proposed by Norberg–Schulz (1965). In the next section I conduct a detailed analysis of overlapping AOIs and sketches to provide examples on how designers’ abstractive visual thinking can be interpreted through AOIs, in addition to a more conventional approach of reading the locations of AOIs.

3.3.2 Affinity between Mechanical Eye Movements and the Observational Goals of Architects

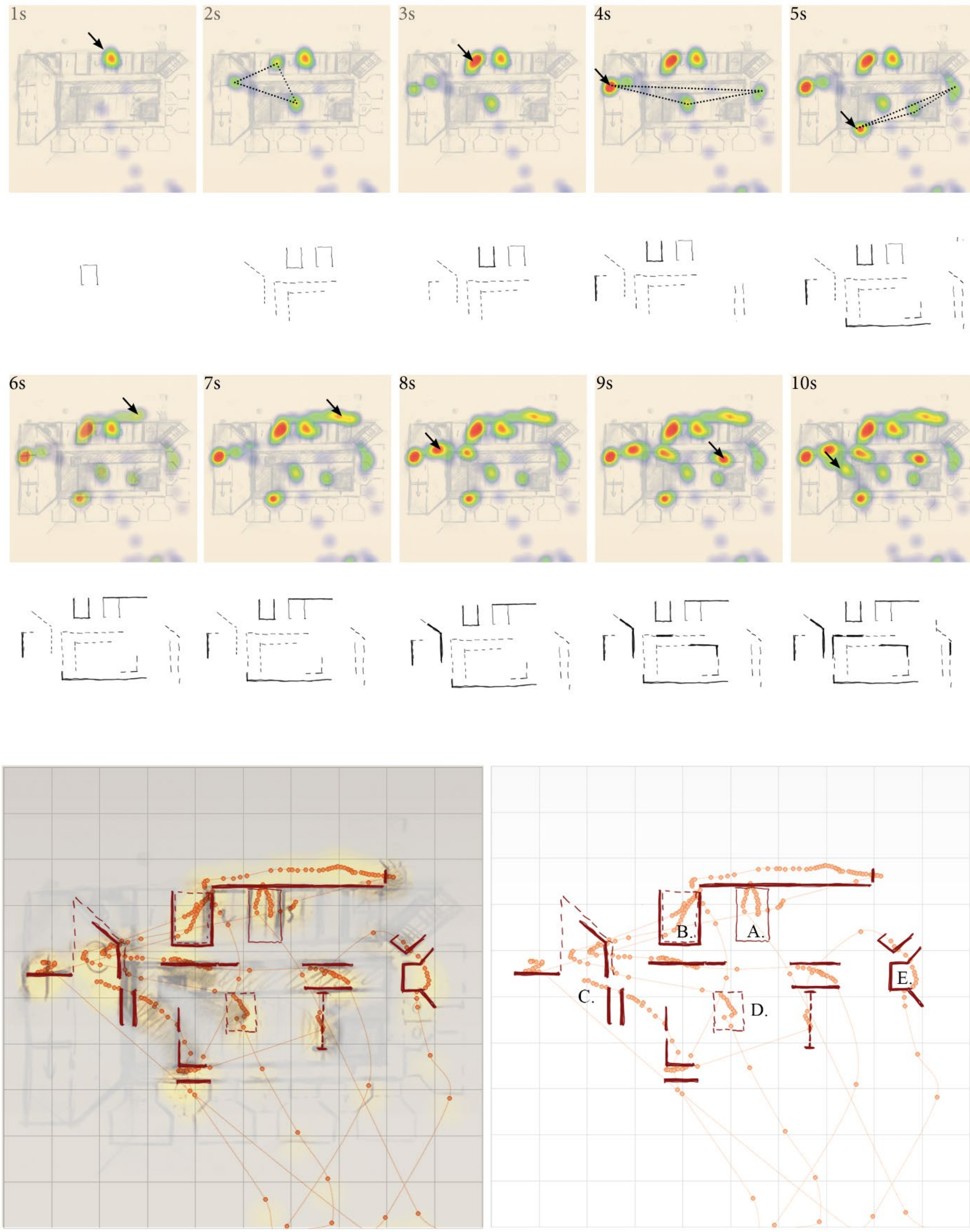


Figure 37. An example of heat map result of “shape” analysis on Louis Kahn’s 1964 sketch.

One of the major concerns in eye-tracking studies in general is how much the heat maps are able to imply given the shapes and areas they are overlapping with original visual stimuli. However, due to the nature of visual thinking of training architects, where and by what sequence the eyes were commanded to move reflect the choice of observation for given goals. One explanation of the deliberate eye movement can be attributed to the requirement of long-time fixation for scene perception: 260 – 330ms. In this section, I take a direct approach to interpret the meaningful observation of shape, composition, and circulation of one participants' heat map results using one-second time stamp to examine whether the mechanical movements captured by the heat maps and scan paths will serve as meaningful graphic clues responsive to the three intentions.

Above is an example of ten heat maps resulted from "shape" analysis (figure. 37): this plan sketch was done by Louis Kahn in 1964 for classroom building for the Indian Institute of Management in Ahmedabad, India. A series of black arrows and dashed triangles indicates how the observer's fixation points shift over the ten-seconds task window. The red areas indicate the longest fixation time, and time declines following the rainbow spectrum from red to blue. The following is a brief description of the intentions captured in this example of heat map:

Time Stamp	Visual Behavior	Focused Areas on the Sketch
1s	Fixing	A rectangle (A)
2s	Scan	A neighboring rectangle (B) and interior of two major shapes (C) and (D)
3s	Fixing	Rectangle (B)
4s	Fixing + Scan	Shape (C), (D) and new area (E)
5s	Fixing + Scan	Contour of (D)
6s	Fixing	Contour of (A) and (B)
7s	Fixing	Contour of (A) and (B)
8s	Fixing	Contour of (C)
9s	Fixing	Contour of (D)
10s	Fixing	Between (C) and (D)

Table 4. Sequential visual behavior observed in a ten-second observation task of the intention: shape in figure xx.

Scanpaths and fixation locations have implied how "shape" was perceived in this particular example. The overall viewing pattern suggests that the participant started at a fairly simple rectangular shape(A), then quickly scanned its neighboring rectangle (B), and searched for adjacent, larger shapes (C) and (D). After the first impression of the relative location (B) was identified, on the 4^{th} second the participant searched another shape (E) to the opposite side of the sketch, continuing building up peripheral shapes by allocating partial contours from the 5^{th} – 7^{th} seconds. Then the participant focused back into the inner areas of the plan, and identified connecting areas between major shapes, such as (C) and (D). A compilation of fixation areas and shape segments can be seen in the bottom two images of figure. 38.

The participant, within the brief ten-seconds window of the observation task, was able to focus on critical graphic clues that were distinct and expressive enough to describe shapes accordingly in the way they were perceived. For example, the obtuse angle of shape (C), two diagonal corners of shape (D), and the extrusion of shape (E). Although these clues of shapes were identified, they were not necessarily dedicated to one shape per clue only. Due to their fragmented appearances, these clues can be merged in another way in order to be imagined into another expression of shapes: such as the edge fragment of (D) and the leftmost side of shape (E) will combine together and produce a long rectangle. Consequently, what the intention of shape in this sketch, through the view pattern of this participant, was the identification and shape clues that best describe the characteristics of potential shapes, and relations among these clues. Scanpaths served as connection that tied these clues together.

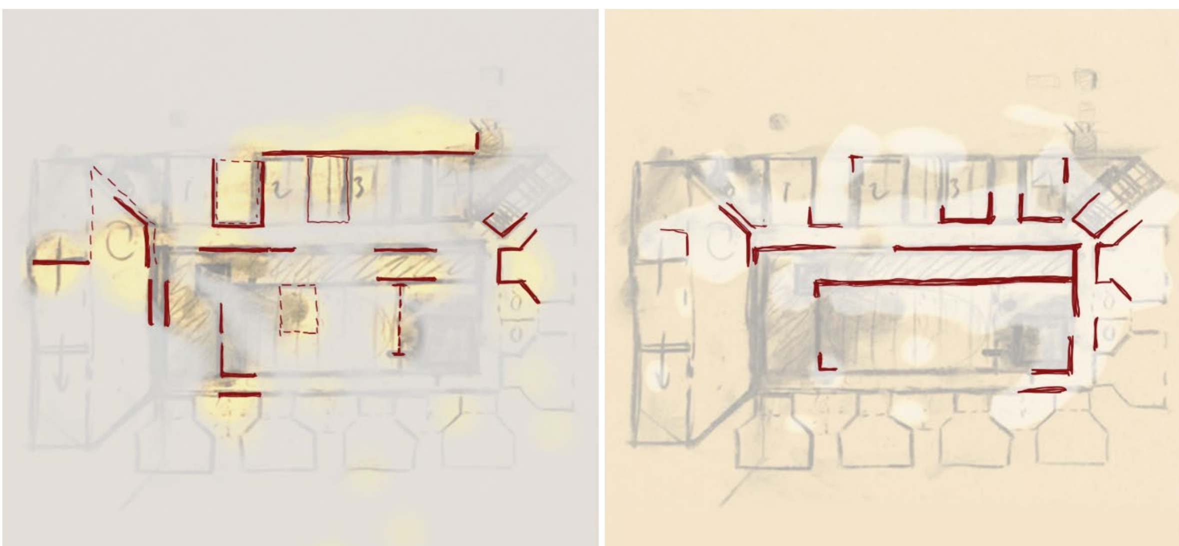


Figure 38. The participant's result comparing to the compilation of all seven participants' result.

Similar approach when observing composition and circulation as architectural intentions can be found using the same analysis method.

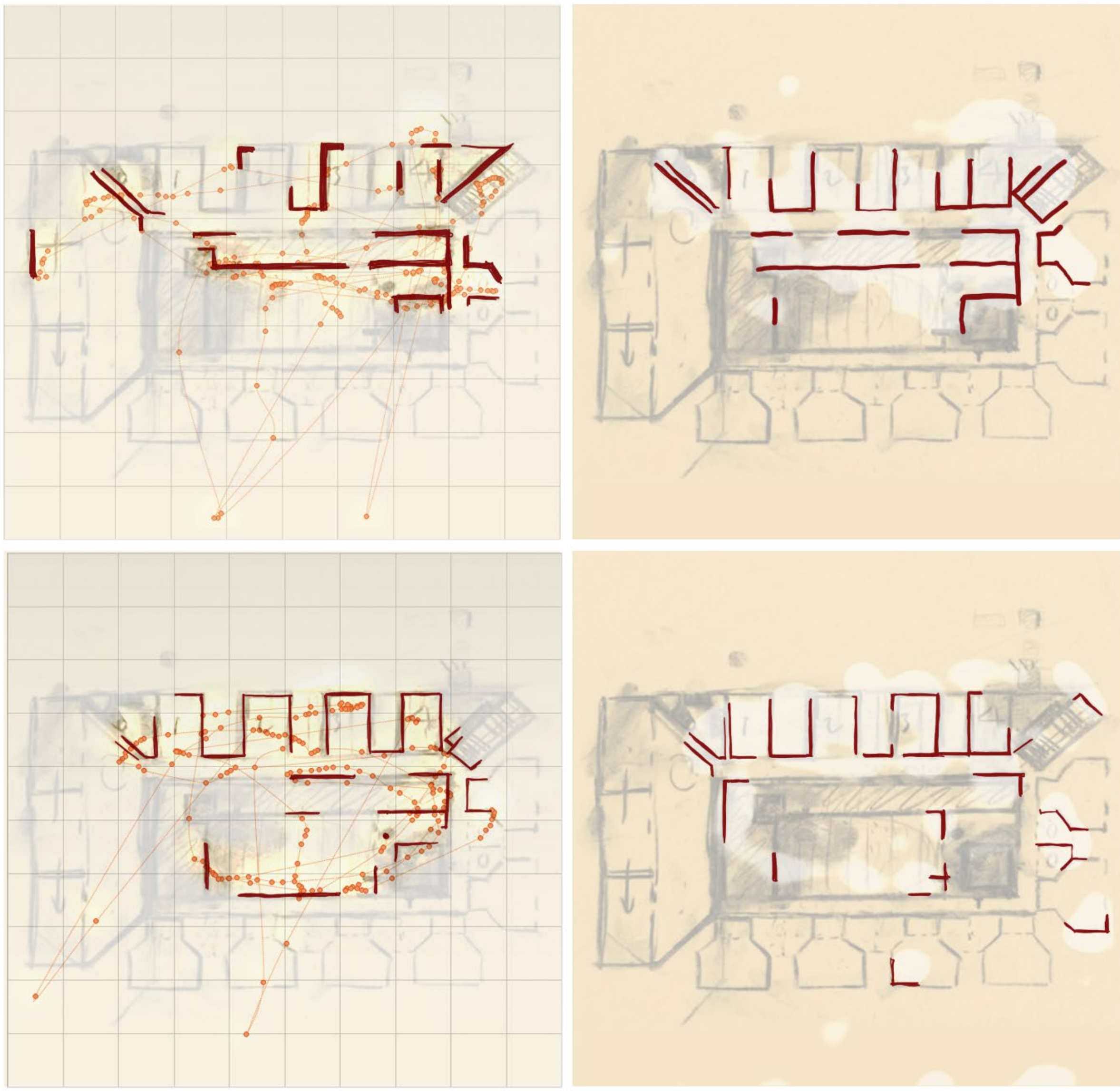


Figure 39. Composition (top) and circulation (bottom) results of the same sketch.

In the top pair of composition diagrams (figure 39), it is clear that scene perception areas (overlapping with at least 7 fixation points) were more spread than in the example of shapes. This is a result from the participant paying more attention to the relationships among shapes as they provided stronger clues on how the composition of entire sketch was presented. We see an increasing amount of fixation on visually similar shapes, such as the row of rectangle at the top of this plan sketch. This is also explainable by what architects consider for arranging inter-shape relationships: by strengthening rhythms, repetition, and connections.

In the bottom pair of circulation diagrams, a greater amount of scene perception areas was given to the inside of shapes in comparison to the former two intentions. Most of these areas running either parallelly or perpendicularly with graphic clues which are interpreted as walls and other plan elements for such sketch. The scene perception areas were even more spread out comparing to the composition, as they indicate imaginary movement in the plan drawing. The scanpaths indicates that the participant was examine through enclosed contours for circulative patterns.

The above three interpretations of intentions and their relation to view pattern and fixation areas imply that the observer's eye moments were impacted by what type of observational goals is effective. How the graphical elements of a sketch will interact differently with view patterns to fit any selected goal, or architecture intention in these examples. Although the sketch of an architecture plan is usually seen as one image that perform one function of representing an idea, the actual perception of such "idea" is rather depending on the one who observes, and the intention to be observed. The resulting variations on graphical clues and eye movement patterns assist the different interpretation of the intentions, facilitating livening a plan sketch to consist multiple innate visual opportunities for further imagination.

3.3.3 Goal-oriented Observation Results in Different Interpretation of a Single Design Sketch.

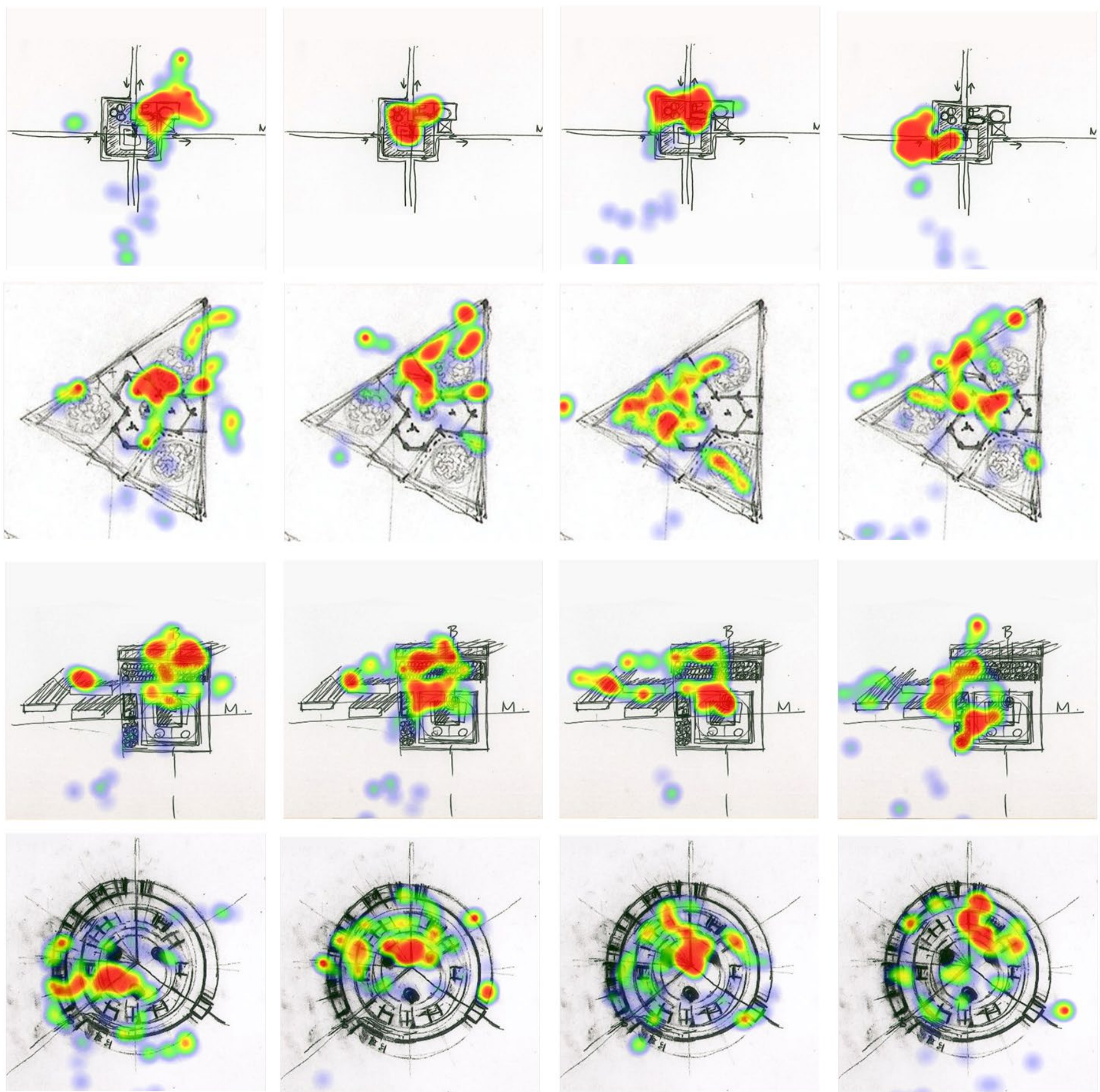


Figure 40. Four Participants' heat map result on the intention of "shape".

Previous discussion on how scanpaths affected graphical clues to be interpreted differently further implies the resulted graphical areas of heat maps overlapping with sketches will likely provide another level of translation between intentions are extracted, abstract visual components. In this section I continue to examine how these resulted heat map areas, especially the areas of interests (AOIs) from eye moments relate to the reading of the three architectural intentions. For a cleaner graphical representation of eyes' fixation areas, in this section the difference in attention levels is divided into two major categories as in the first sections' color code analysis: red, orange, and yellow region as scene perception; green, cyan, and blue as visual search. To start, I compare heat maps from relatively simple plan drawings by Louis Kahn, and progress with more complicated plan sketches.

The first four simple sketches are the following (figure. 40):

First row: Penn center studies A: a centralized square form with almost bilateral symmetrical layout within the square. The scene perception areas are consistence and covering one quarter of the overall rectangular shape.

Second row: City Tower Project: a centralized equilateral triangular layout. Most scene perception areas are grouped and overlapping with one of the three similar portions that compose the triangular plan.

Third row: Penn center studies B: a bilateral rectangular primary shape with three similar additive rectangles on the left. Scene perception areas consist of three major groups: one on the bottom square shape, one on the top, darker rectangle, and the third on the left grouping of three small rectangles.

Fourth row: Civic Center Studies: a centralized form of concentric circles with indication of three similar subdivisions. Scene perception areas are again grouped on one of the three subdivisions, and the division lines suggesting the subdivisions.

By superimposing heat maps of "shape" on four sketches relatively by four selected responses from participants, I summarize three persistent characters that can be found from these sample heat maps: firstly, scene perception areas are covering one characteristic representation of the symmetrical form; in most of the cases on the borders where two shapes of different properties touch; and interior of the shapes attracted less attention than the borders. The symmetry of the composition is indicated very briefly by visually "copying" the overlapping areas under visual perception to their reflected locations.

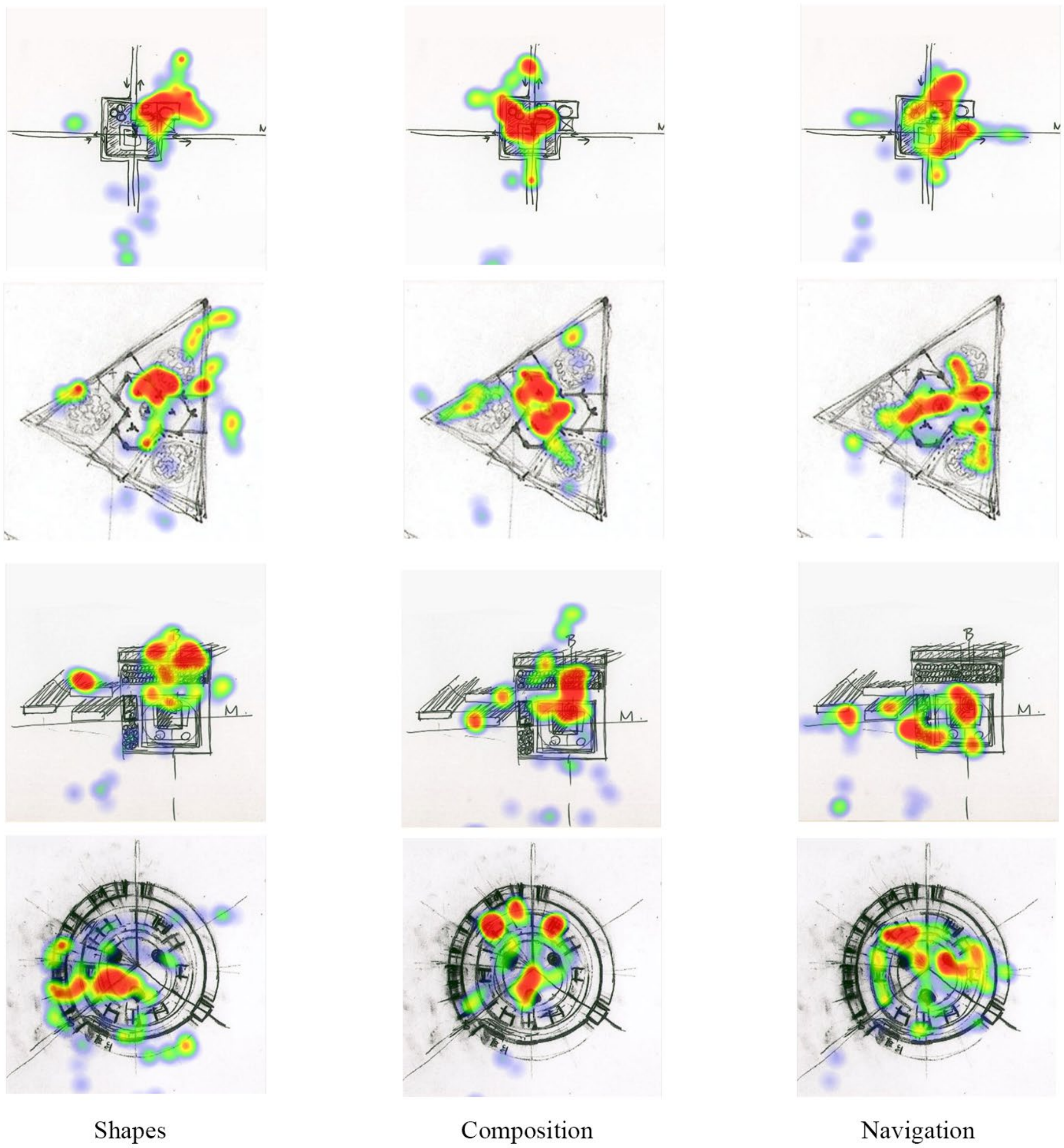


Figure 41. One participant’ on three intentions: shape, composition, and circulation.

Next the variations on heat maps on different architectural intentions are examined. The middle column of the above figure (figure 41.) consists of four heat maps responses on “composition” by the same participant, and similarly the right row is a collection of responses on “circulation.” The first observation of differences between the latter two and “shapes” are their extended scene perception areas in other symmetrical parts, which in “shapes” receive less or little attention. Compared to heat map forms in “shapes”, in “composition” they appear to be more aligned to the compositional characters of each sketch.

For example, in the triangular form, the heat map contains extrusion from the center to all three corners. Similarly in the rectangular and circular forms. As for "circulation", major scene perception areas cover more of the sketches and connect more often than the early two architectural qualities. These characteristics are consistent with the previous analysis of eye movements.

At this point it is proper to link the eye movement by observers for a given architecture intention to the graphical representation in color-coded heat maps. Therefore, a graphic-to-graphic relationship can be established by extracting scene perception areas and find the overlapping segments in the original sketches. These areas begin to demonstrate effective areas varies for different intentions graphically.

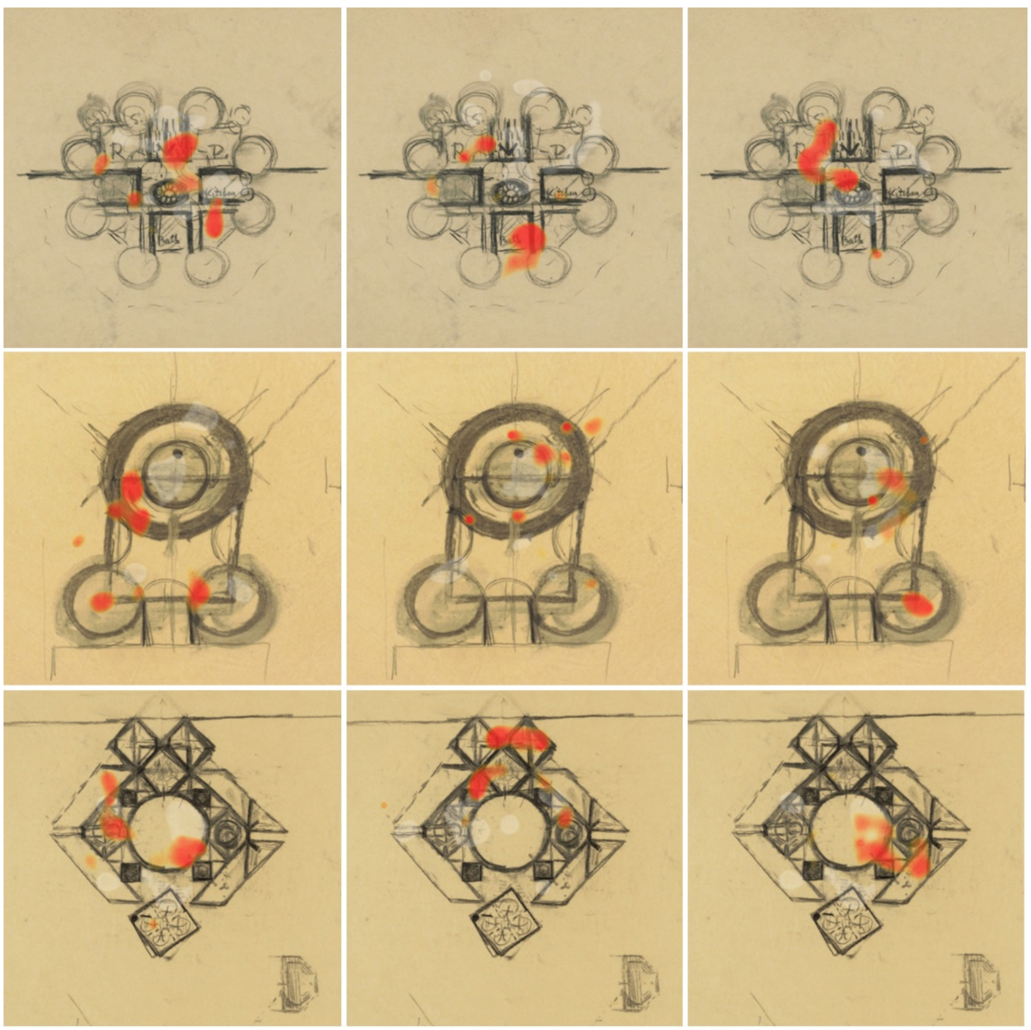


Figure 42. Notation on the effect areas for intention: shape, based on the color-coded heat maps from three different participants' data.

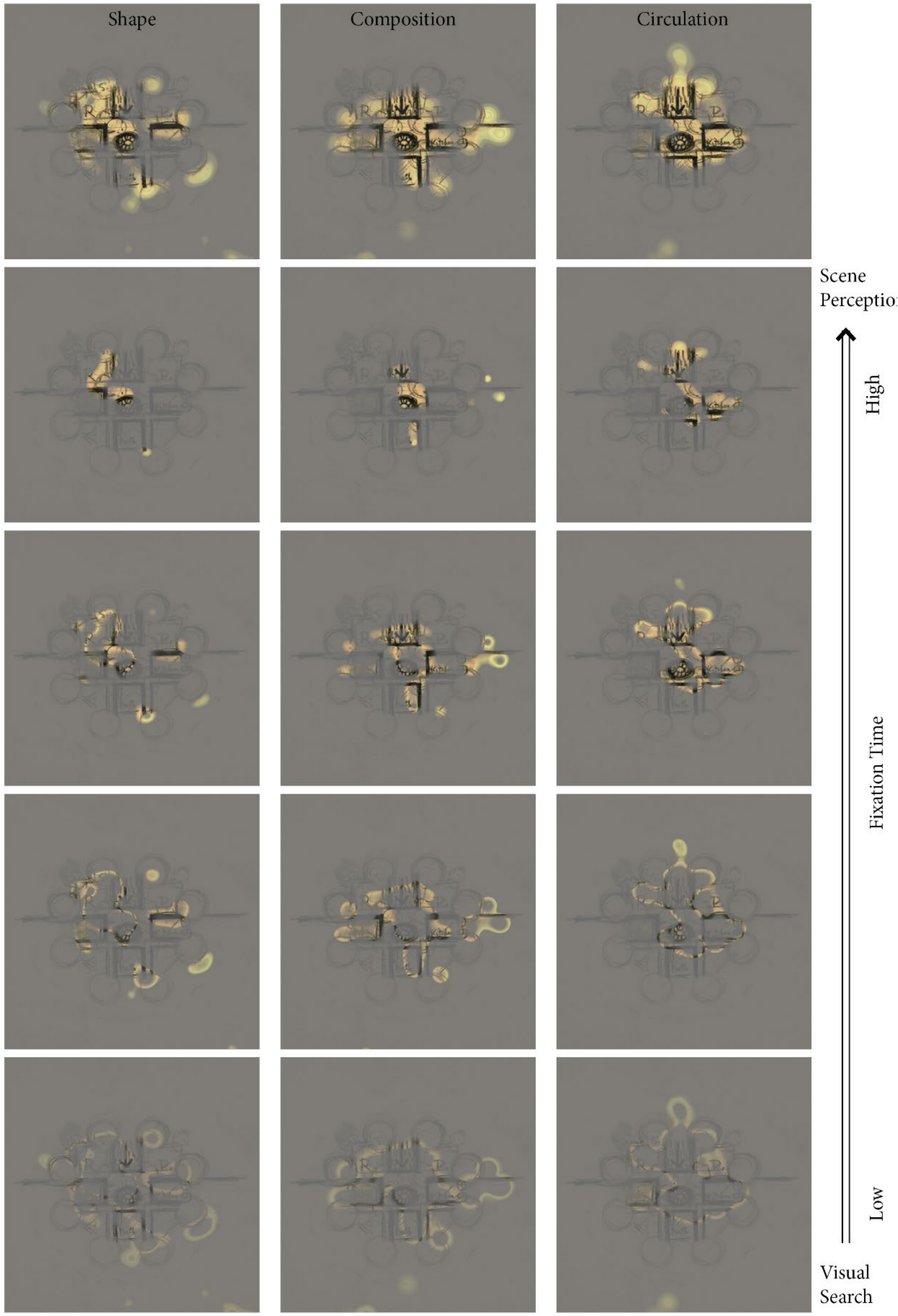


Figure 43. Example of different areas in a sketch receiving different level of attention from the observer.

3.3.4 Viewing Patterns Reflect Meaningful Design Information Visually.

Not only observers' eye movement suggested that the formation of heat maps can be affiliated with deliberate visual process used by goal-oriented design professional, but the resulted scene perception areas possess consistent graphical qualities that aligned with each design intention accordingly. However, before proceeding into the discussion of graphic-to-graphic translation between viewing pattern and architectural intentions, I will reemphasize the implication of graphical translation connecting these intentions with visual representations: abstractness. The more an architect like Michelangelo or Le Corbusier see, the more effective those visual stimuli were translated into a form of abstract, graphical representation of any labeled objects: from a row of columns, they see the interplay of light and shadow; from a repetitive motif they see rhythm and movements; and from blocks of geometries, they see connectivity, negative and positive spaces, and gestures of composition embedded in any explorative sketches. Therefore, this section will be a primitive exploration on these capacities through direct use of scene perception areas and overlapping sketch segments to suggest an approach of utilizing the power of abstractness to facilitate representation of intentions: from perception to creation.

Each intention, as observation goals, facilitate a set of different forces that form the interpretation of the intention through the graphical clues overlapped by heat maps (figure 43.). If these graphical clues are extracted and placed back into a potential representation of the intention respectively, it is possible to obtain a variety of these representations, using only a small segment of extracted visual clue. The following diagrams, taken from the previous four participants' response on "shape" (figure 44. & 45.), present a small selection of what can be reconstructed using the segment in the pink scene perception areas. Figure 45 is an invert-colored figure 44, presenting the scene perception areas as positive solids and graphical clues as negative "carvings" against the solids.

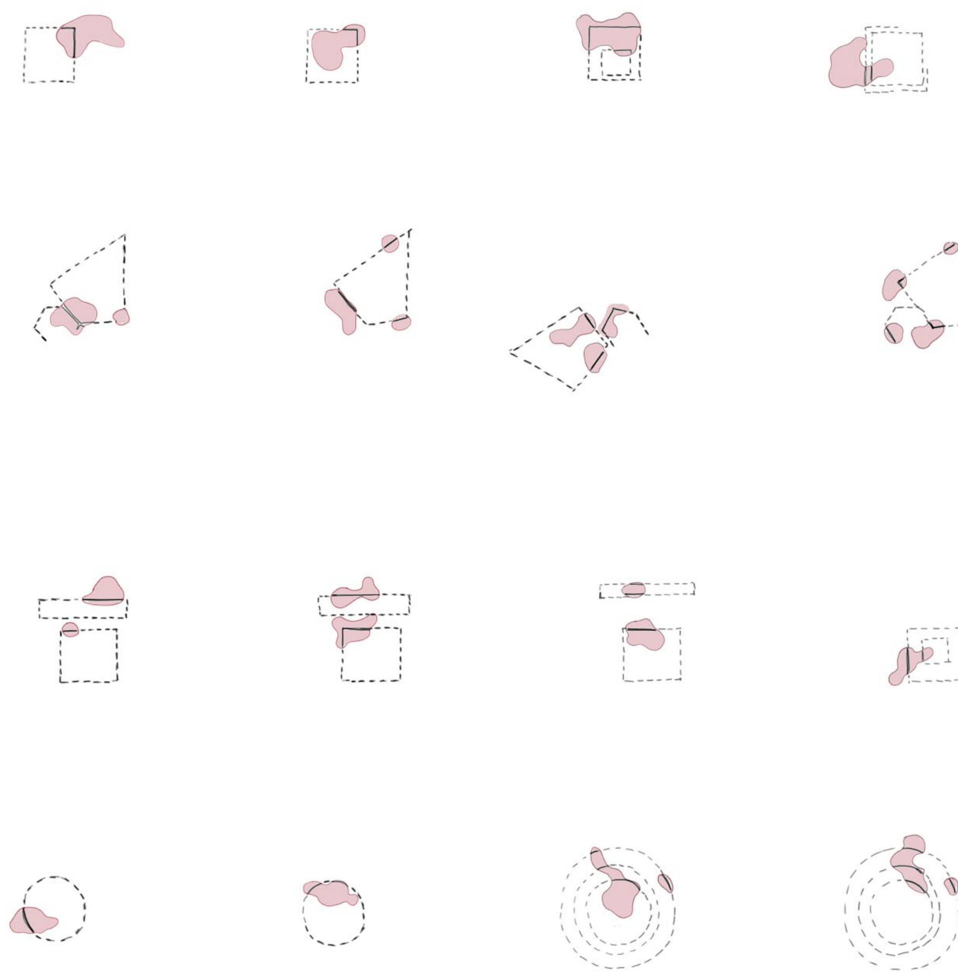


Figure 44. “Reconstructed” potential forms using scene perception areas as indicated in figure 40.

Shapes

Figure 45. Invert-colored figure-ground diagram of scene perception areas in figure 44.

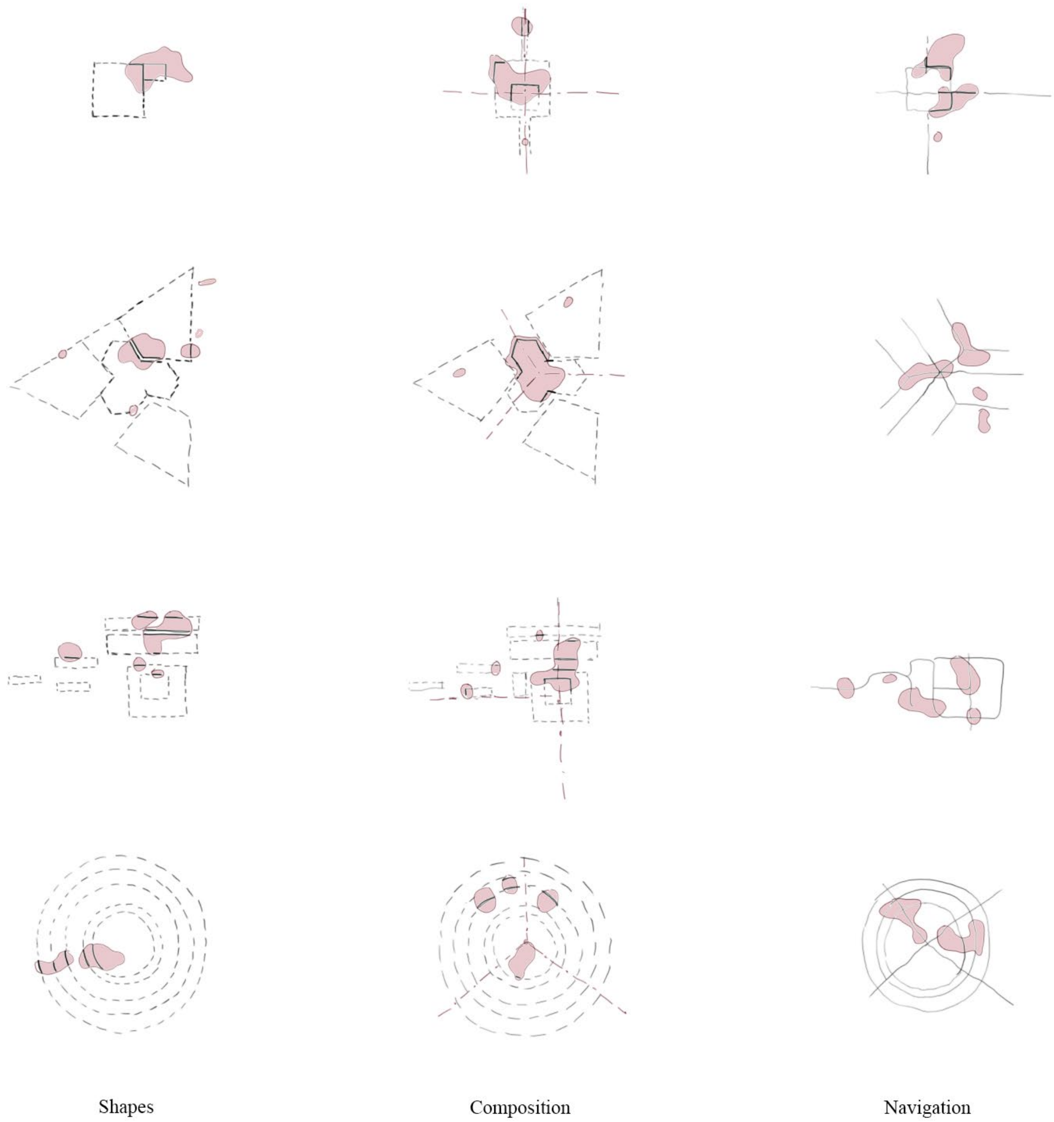


Figure 46. “Reconstructed” potential forms using scene perception areas of more complicated plan sketches in figure 41.

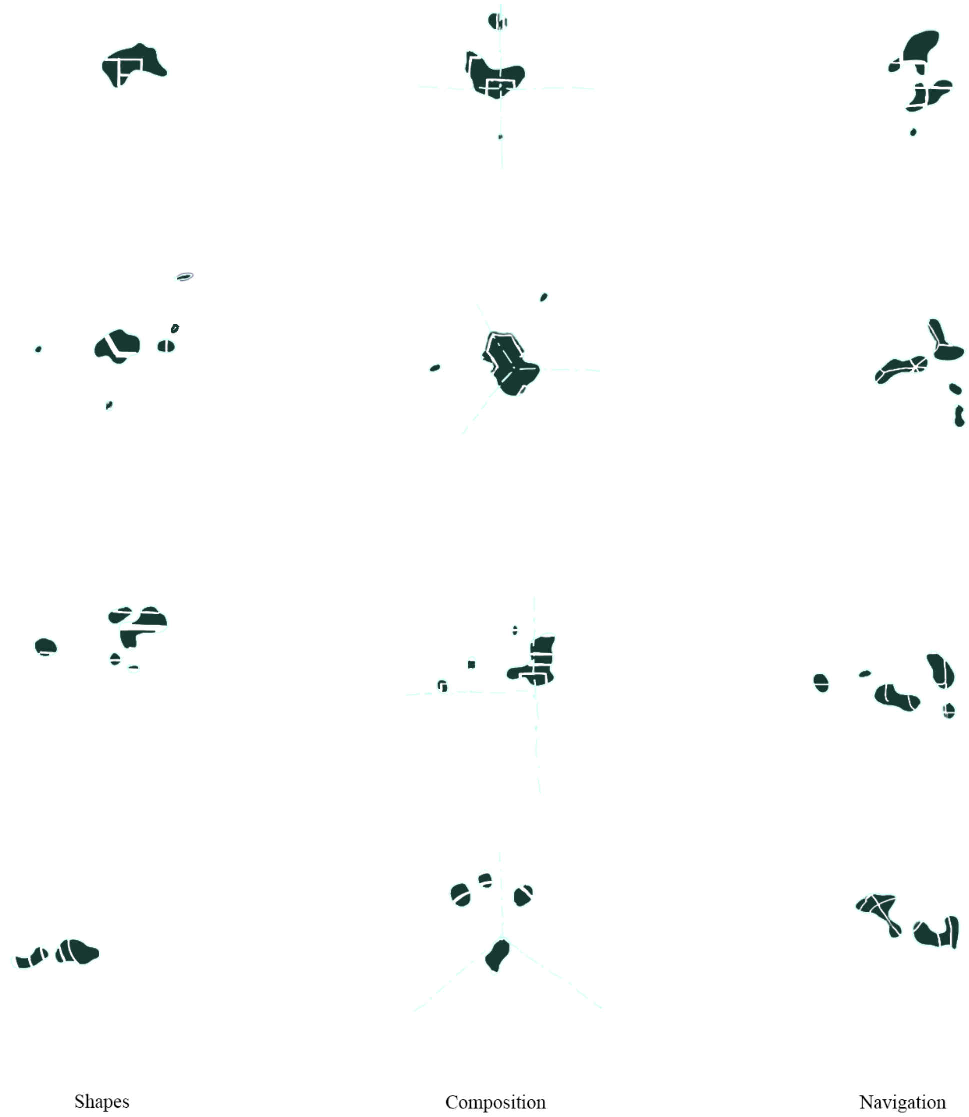


Figure 47. Invert-colored figure-ground diagram of scene perception areas in figure 46.

Findings from the analysis on heat maps indicates that composed intention is itself a dynamic gesture of geometries based on formal characteristics. As eye movement potentially opens a path to form graphic representations of deliberate observation on architectural intentions, it is then possible to reframe how these intentions can be perceived using a more machines-like process. Back to the previous discussion on the three inspirations from machine vision algorithms at the end of chapter two: grouping of graphical clues, a skeletal system based on movements of a point, and an algorithm for image-to-image translation. If the scene perception areas are potential segments for grouping, and the viewing patterns indicates centers of these segments and movement directions, a structure of representing architectural intentions based on eye movement can be proposed for the following three architecture intentions (figure 48.):

Shapes: clustering of corners/segments of distinctive forms and then reflecting using symmetry.

Composition: Successive graphical clues linked by scanpaths.

Circulation: on converging points and "interior" of shapes that mark potential paths in circulation.

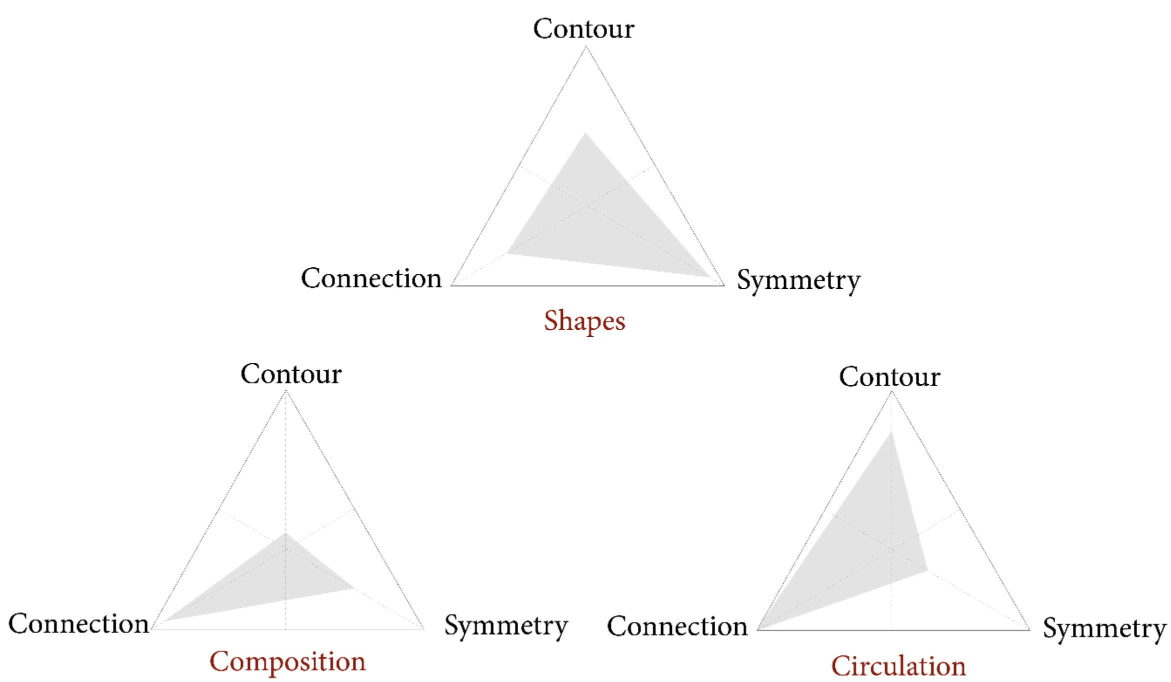


Figure 48. Diagram of graphical characteristic and level of influences in representing each architectural intention.

The viewing pattern and resulted AOIs overlaps suggest a similar system of 2-D vision tasks' richness in perceiving shapes: as two-dimensional shape perception to be considered as a "mid-level vision Gestalt" (Wagemands, 2013), and a mix of "low-level and high-level" (Wagemans, 2013) dynamic "interplay" of visual elements. "Curvature singularity" (Attneave, 1954), which means that where the curvature on a shapes' contours reaches local maximum is the most "informative" component of the shapes. Formal perception is an eminently active occupation (Sánchez & Bragado 2019). Taking in the shapespace

(Wagemans 2013) enabled by uncertainty innate with sketches, and scan path directions from mechanical eye movement, a graph on the representation of potential weighting on visual interpretation of eye movements can be made similar to figure 48. Fixation areas are therefore seen as anchors from where the graphical clues, such as lines and shapes, to be multiplied by their influence in contour, connection, symmetry, and other abstract architectural qualities to be translated into intentions in a graphical manner.

3.4 Discussion on the Eye-Tracking Tests

Through the design and implementation of eye-tracking studies, I concluded three major graphical characteristics that are shared in most of resulting heat maps: a link between eye movement and intended visual perception areas; a correspondence between different intentions and view patterns; and distinct graphical emphasis of fixation areas reflected in each intention. However, many improvements will support a more comprehensive examination of these discoveries.

First of all, the hardware implementation in this study was limited in accuracy. Webcam-based eye-tracking has a lower resolution comparing to other wearable devices (Semmelmann & Weigelt, 2018). Recorded view patterns from online platforms can have "an offset about 211 pixels (18% of screen size, 4.38 ° visual angle)" (p. 460). Additional environmental factors that affected the accuracy of the tests are lighting, resolution, and refreshing rate of the webcams.

Secondly, framing of each intention instruction can affect the participants viewing patterns, both for the term given to describe an intention and the meanings of terms for different intentions. For example, in the question of "Navigation: Imagine how you can move through the space in the plan", the word "navigation" was originally designed for a more active expression of using eyes to trace an imaginative path to navigate in the plan. This word was replaced with "circulation" for a clearer architectural expression when analyzing heat map results. Another example will be the use of "shape" and "composition" as two distinct expressions of architectural intentions. In the study, "shape" implies recognizable geometries that form architectural spaces, such as circles, squares, and triangles; "composition" implies interconnectivity among these recognizable shapes and their formal relations such as symmetry, major and minor spaces, enclosure, and centrality. Although these terms of intention and their implications are familiar and shared among architectural students, for people who have little to no professional training or knowledge, "shape" and "composition" potentially mean the same. In future studies, a comparative study between expert and non-expert groups is required to observe how differently view patterns can be affected by these goals for observation.

Thirdly, the three intentions used in this study are a small selection of possible intentions innate to design sketches in Louis Kahn and other architects (figure. 49). When a participant was given no instruction to observe a sketch, he or she will be more likely to pay additional attention to other possible intentions in presence. Meanwhile, how the three given intentions perception will also be influenced by the graphical appearance of each sketch, if their prominent intention is seen differently from shape, composition, and navigation.

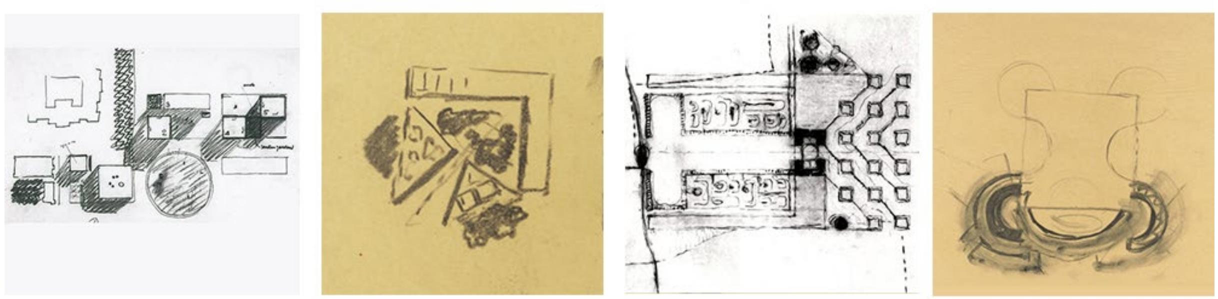


Figure 49. Other implications of architectural intentions can be seen from Louis Kahn's sketches.

Finally, on the individuality of each participant and design sketch, the dynamic and emergent nature of observational process and graphics are highly receptable to personal choice. Furthermore, how exactly Louis Kahn expected for his sketch to express or to be seen would be most likely different. A group of fifteen architectural students' eye movement results would only consist a very small segment of the vast sea of graphical intention and dynamics just from on drawing. The examples discussed in chapter 3.3 would only be a tiny scope into the willful visual powers in architectural sketches. Such connection between eye and mind still requires a far greater amount of work in the future to just glancing into one spot of the full picture.

However, even from this minuscule inspection of the visual power in design sketches, it is possible to make and test how a machine translator will be able to start recognizing and representing the willful observation of human designers meaningfully.

Chapter 4 Creating with Eyes

In this chapter I continue the discussion in chapter two and three on vision and eye-tracking results to propose a novel application: Envisage, using machines learning. Then, I test two prototypes of this framework in representing architectural intentions through a Neural Network algorithm recently developed as "Image-conditional Generative Adversarial Network" (Isola et al., 2018). The first section consists of descriptions of this application. The second section consists of examples and discussions on the two prototypes that will be an attestation on how effectively this framework can become in delivering high-level representation of design intentions from new sketches: one is "Sketch to Intention", and the second "Observation to Gesture". The third section will be a discussion of all the work so far and explain why this application will be a step towards a more dynamic, observer-based design system that encourages visual thinking of ambiguity and intention in machine learning. Furthermore, the discussion section takes a step further from the label-based deterministic calculation into a broader use of image-based thinking exploration stemmed from uncertainty and innate creativity in visual thinking. Finally, this chapter will conclude with future possibilities enabled by vision based communication strategy between designers and machines, and a critical analysis of the work done so far.

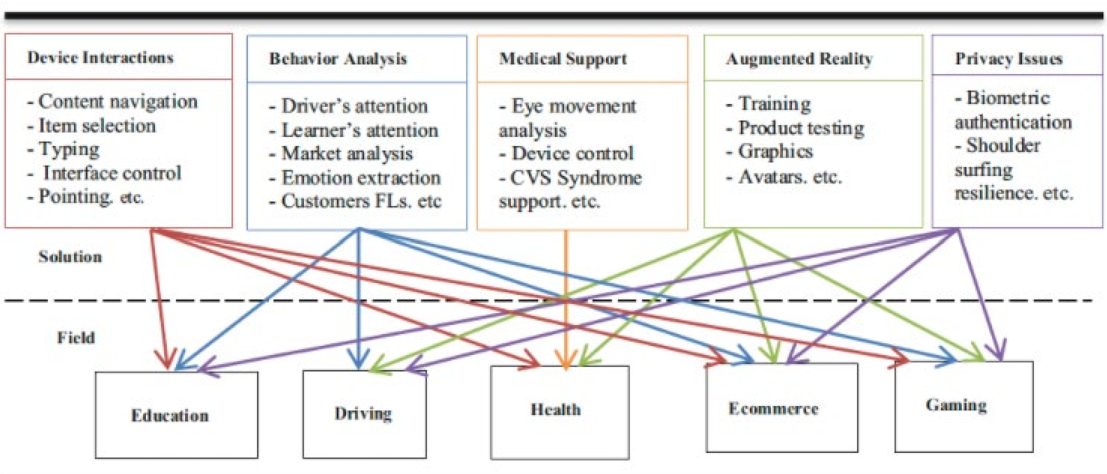


Figure 50. Existing practices in utilizing eye-movement in assisting decision-making (Shehu, I., Wang, Y., Athuman, A., & Fu, X. 2020).

4.1 Sketch to Intention + Observation to Gesture: A Novel Application of Image-conditioned GANs

In this first section I present two applications of image-conditional Generative Adversarial Networks (GAN's) from heat map- sketch datasets collected for chapter three. The major goal to train GAN's is to test on how perception of more abstract intentions can be translated by machine learning algorithms, and therefore the tests conducted will not be conventional application of neural networks to find definitive description or one-to-one translation from one image into another. For the translation problem is between two graphical inputs: heat maps and sketches, I applied the algorithm of an Image-conditional Generative Adversarial Network (GANs) proposed by Isola et al. in 2015[16], which is currently an open-sourced code accessible through Google Colab environment in the TensorFlow examples called "Pix2Pix."

For GAN's has been discussed in the third section of chapter one, in this section I will start with description of specifics that are related to the new applications (figure 51.).

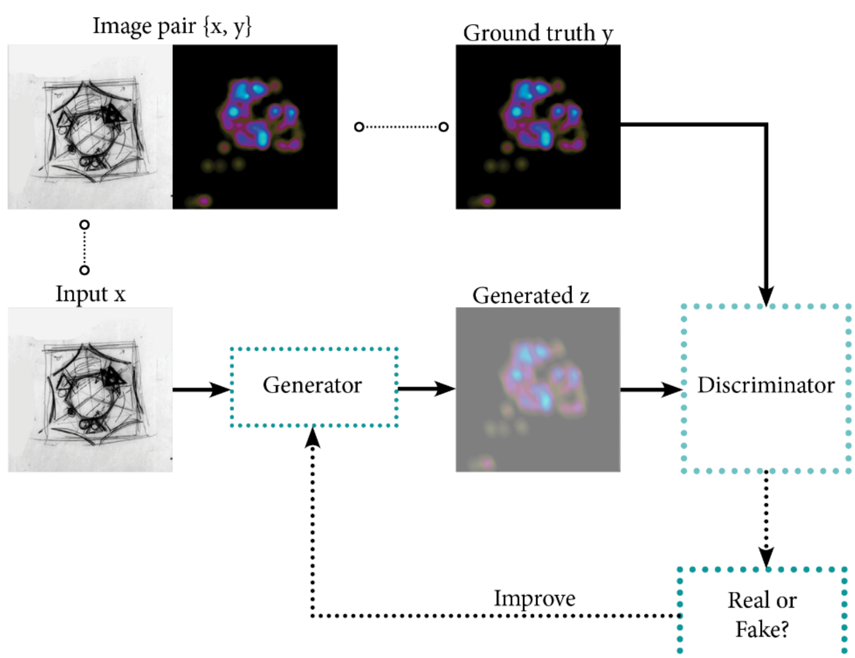


Figure 51. Workflow of GAN's, after Isola et al., 2018.

[16] The original paper was published in 2015, the reference used is the third version published in 2018.

4.1.1 "Sketch to Intention" GANs

i. Data Preparation

The first "Sketch to Intention" GAN's is a direct image-to-image translation by assigning sketches as input and heat maps of viewing patterns as prediction. The training and testing image sets consist of sketch-heat map pairs that obtained through the eye-tracking tests in chapter three. The original rainbow-colored heat maps, which were obtained from GazeRecorder, were processed in Photoshop through an invert image filter to prepare for training set images, assigning the background as black.

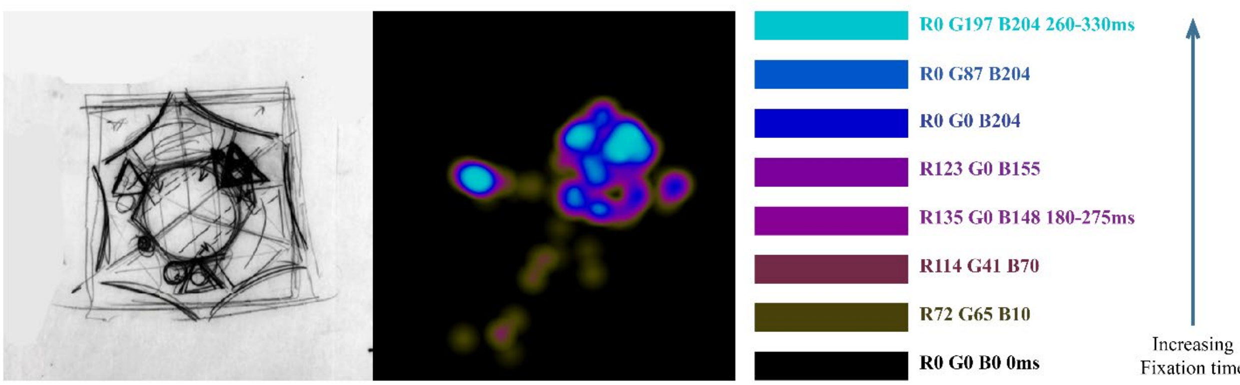


Figure 52. Input pair of a sketch (left) and a heat map (right), and color codes for the heat map.

"Labeling" process in the preparation of training images was to combine the invert-colored heat map to corresponding sketches for all three intention. Two types of measurements from the heat maps were taken into consideration: fixation durations which were color-coded accordingly, and shapes of these colored regions (figure 52.). Both the sketches and heat maps were resized to 256 x 256 pixels to fit the kernel operation in neural net layers for regression (need clear use of terms). For training each intention's corresponding GANs, one-hundred-eighty pairs of sketch-heat maps were assigned to training set, and thirty pairs were assigned to testing set (table 5). Due to the data size, a typical prediction from this GAN's is a vaguer presentation comparing to actual heat maps from human data. However, they are salient enough to cover and indicate potential areas of "intention" overlapping the sketches, following discovered characteristics of graphical clues in chapter three[17].

GANs Properties	Sketch to Intention	Sketch to Gesture
Input	Louis Kahn's plan sketches	Human observation heat maps
Ground Truth	Human observation heat maps	Louis Kahn's plan sketches
# of images in Training set (each "intention" GAN)	180	180
# of images in Testing set (each "intention" GAN)	30	30

Table 5. Training and testing data properties of both GANs.

ii. Training

Following the collection of training and testing data, I trained the network using sketches as input and invert-colored heat maps as output. The trained network is capable of generating a "heat map-like" color-coded representation indicating an imaginative human observer's viewing pattern in terms of fixation time and areas. The whole training process for each GANs was processed in the Google Colab cloud environment. It took 15 seconds for one epoch with 180 images, and a total of 120 epochs took an average of 30 minutes for one network.

A major goal of this training of three separate GANs is to predict individual heat maps for each intention and assign different weights as a combination of quantitative and graphical representation of machines' viewing patterns. These view patterns were "learned" through human data on the same visual tasks. More specifically, based on the different gaze-tracking tasks done by human participants, three GANs were trained separately for each of the three "intentions": shape, composition, and circulation. The following figure 53 provides two set of examples of predicted heat maps.

[17] In Huang and Zheng, 2018, the "plan-to-map" training sets contained 115 pair of images.

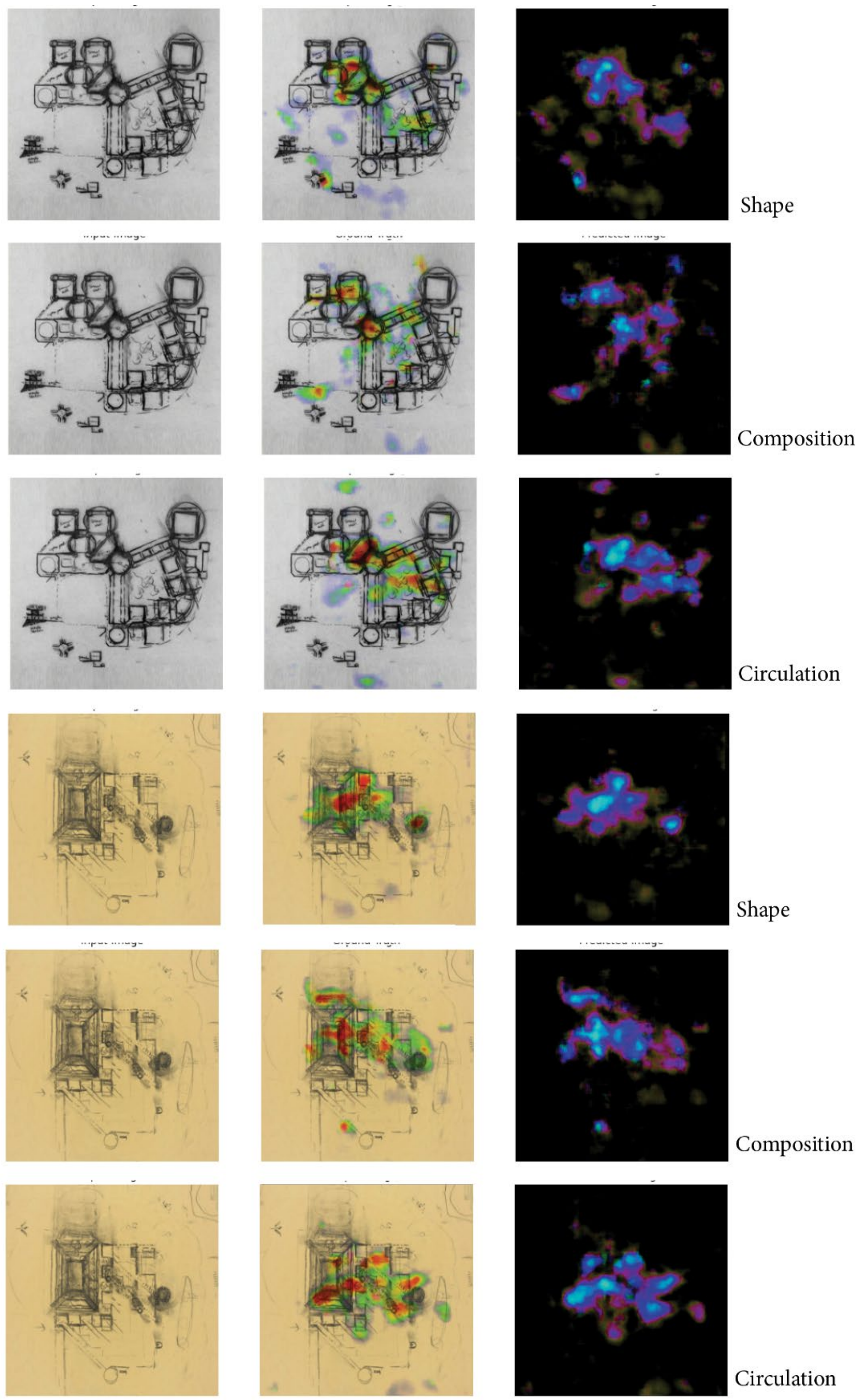


Figure 53. Predicted heat maps from three trained GANs on the three intentions. From left to right: original sketch as input; predicted heat maps overlaying on corresponding sketches; and predicted heat maps.

iii. Result Examples

Trained GAN's prediction provides evidence to the three graphical characteristics discussed in chapter three:

1. Affinity between mechanical eye movements and the observational goals of architects.
2. Goal-oriented observation results in different interpretation of a single design sketch.
3. Viewing patterns reflect meaningful design information visually.

The top three images consist of predicted heat maps (right) and overlaying of these heatmaps on the original study sketch of Louis Kahn for President's estate at the First Capital of Pakistan in 1966. In the first row of shape predictions, the trained network performed well in recognizing the hinge areas that are composed by one circular and a trapezoidal shape. The composition of two squares and one circular shape in the upper left corner of the sketch were recognized by assigning high fixation time areas (in color red) to overlap with one quarter of these symmetrical shapes. This continuous red area implied a consistent reading of these partial contours to extract graphical clues of reconstructing the shapes. In the upper right corner, a small area overlapping a corner of the square inscribed in a circle. Towards the bottom right area another patch of heat map overlaps with two of the four symmetrically arranged rectangles. This prediction is consistent with the graphic features discussed in section 3.3 on how shape is represented through identified distinctive shape features and utilize quick scans to complete the reading.

The second row provided result from the networks predicting composition. It is obvious that the high fixation areas were connected more closely than that in the shape prediction. Similarly, to the previous result, major forms such as the circle close to the center of sketch and the upper left corner were recognized. Additionally, this predicted heat map implies a connectivity from the central circular shape to the upper right area, and to the bottom right area. More interestingly, the predicted heat map clearly indicates a symmetrical reflection to the second half of the kite-shaped composition.

The third row shows prediction on circulation from the corresponding trained GANs. This heat map's representation of circulation is also consistent with discoveries in the first chapter. The areas of high fixation time are even more tightly connected and overlapping with insides of enclosed spaces, such as a courtyard like space inside the kite composition. Meanwhile, this heat map prediction seems to "prioritize" symmetrical and graphically cleaner areas to imagine circulation patterns, such as the four edges of the kite shape. Such "prioritizing" move would very likely due to a learned bias from human data in the training process.

iv. Weighting the Predicted Heat Maps

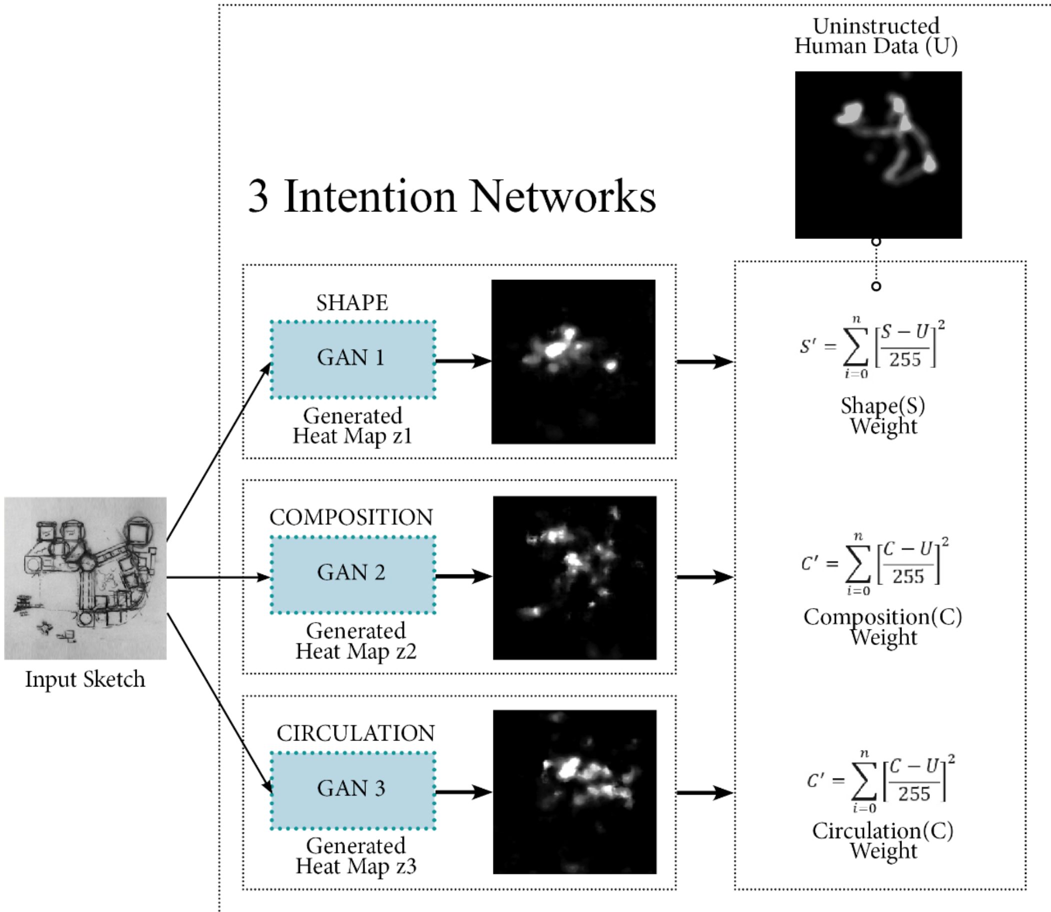


Figure 54. Weighting pipeline of the three networks' predictions on shape, composition, and circulation.

In the test of this GAN, a participant who had taken the previous test sets on the three intentions was asked again to observe ten new sketches by Louis Kahn which were not presented in either training sets or previous visual tasks. The participant was given no specific instructions on what intention categories he needed to focus on in a ten-seconds time frame. Instead, he was instructed to have fifteen seconds to observe and describe what the sketch is about in terms of architectural design language to the author via phone call after the task. The generated heat maps from the fifteen seconds tasks were processed using a custom black-and-white filter in Photoshop to obtain a more optimal grayscale image for intensity (figure 55.). Intensity of pixels are used in calculation of the average of Euclidean distance in intensity value (0 - 100) between each pair of pixels in the human result and GANs prediction to obtain raw difference scores (S', C', and N') for each prediction-uninstructed heat map pair. Adjusted weights (WS, WC, and WN) are calculated by dividing raw difference scores from a constant D (if every Euclidean distance between every pair of pixels is exactly 0.5), then take the power of 2. The smaller the differences are, the more likely that the corresponding "intention" weights more in the uninstructed viewing pattern. Therefore, the higher the corresponding adjusted weight will become.

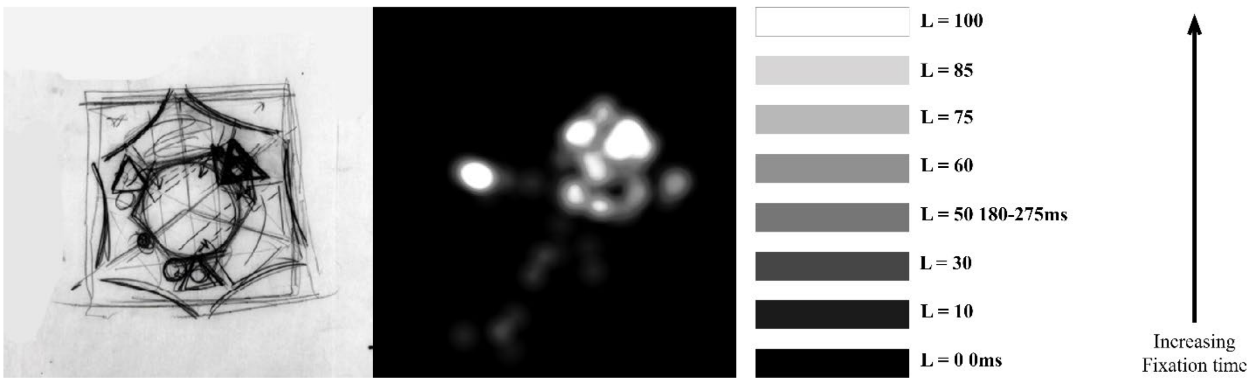


Figure 55. Pair of a sketch (left) and a heat map (right), and Black-and-white intensity codes for the heat map.

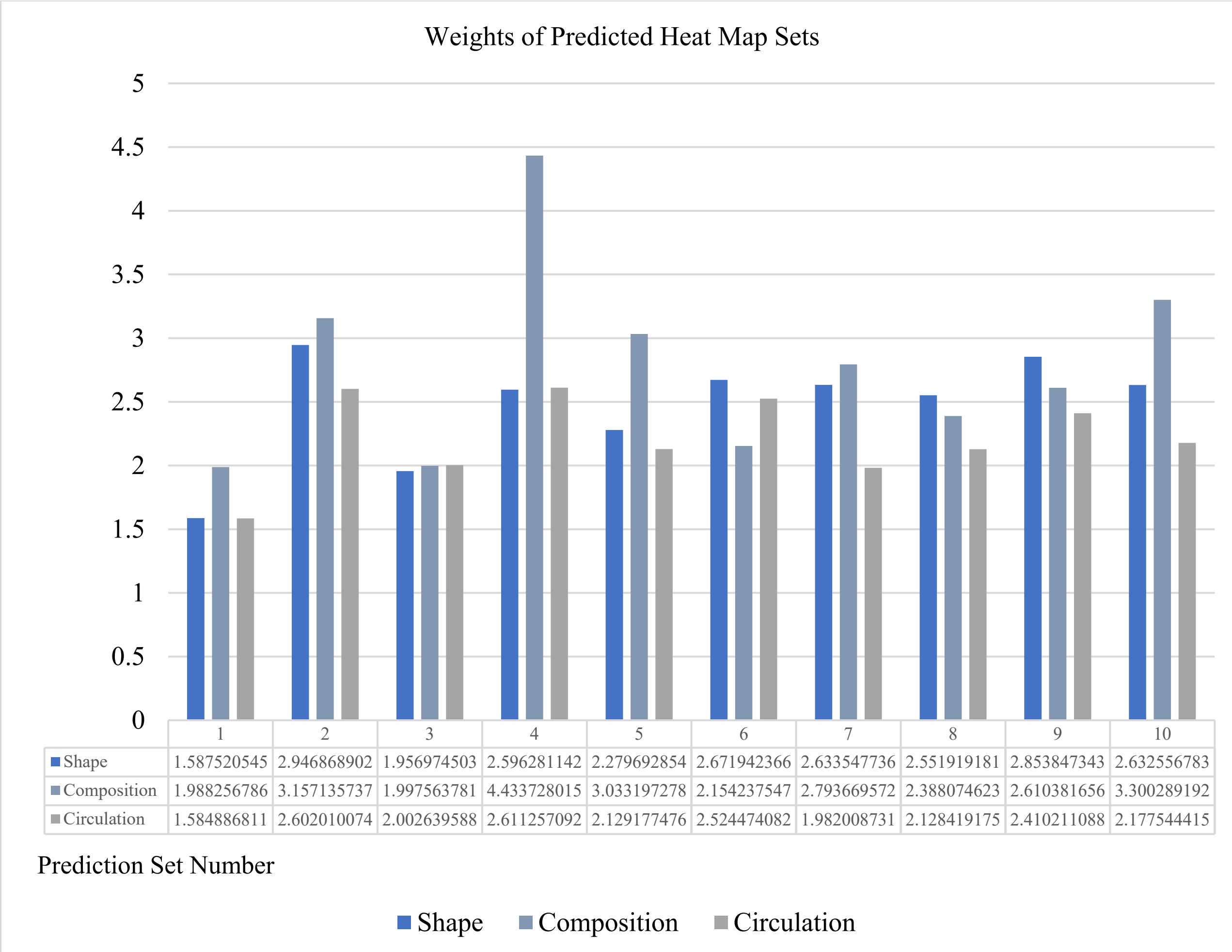
$$S' = \sum_{i=0}^{n} \left[\frac{S-U}{255}\right]^2 \qquad C' = \sum_{i=0}^{n} \left[\frac{C-U}{255}\right]^2 \qquad N = \sum_{i=0}^{n} \left[\frac{N-U}{255}\right]^2$$

Equation 1. Calculating raw difference score

$$WS = \left(\frac{D}{S'}\right)^2 \qquad WC = \left(\frac{D}{C'}\right)^2 \qquad WN = \left(\frac{D}{N'}\right)^2$$

Equation 2. Calculating adjusted weights.

Using this set of three GANs, I collected their prediction result as from "imaginative observers" and compared the similarity between human participants heat map result of uninstructed task on 10 new sketches that were absent in training and testing sets. The resulting heat map prediction and weighting to the uninstructed human data is presented in the following table 6:



	1	2	3	4	5	6	7	8	9	10
Shape	1.587520545	2.946868902	1.956974503	2.596281142	2.279692854	2.671942366	2.633547736	2.551919181	2.853847343	2.632556783
Composition	1.988256786	3.157135737	1.997563781	4.433728015	3.033197278	2.154237547	2.793669572	2.388074623	2.610381656	3.300289192
Circulation	1.584886811	2.602010074	2.002639588	2.611257092	2.129177476	2.524474082	1.982008731	2.128419175	2.410211088	2.177544415

Table 6. Weights of Predicted Heat Map Sets.

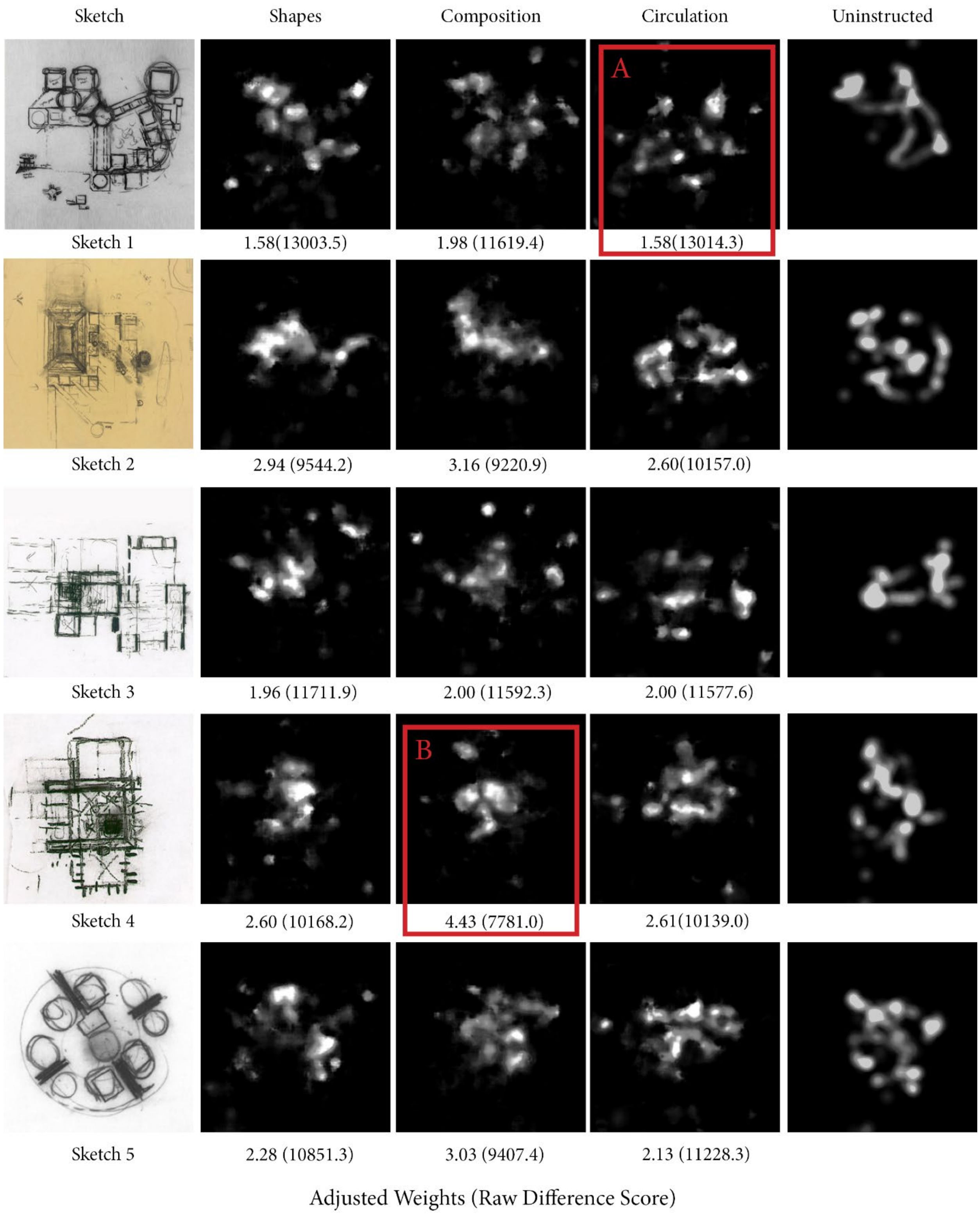


Figure 56. Predicted heat maps (No. 1 – No. 5) and weights comparing to an uninstructed observation data.

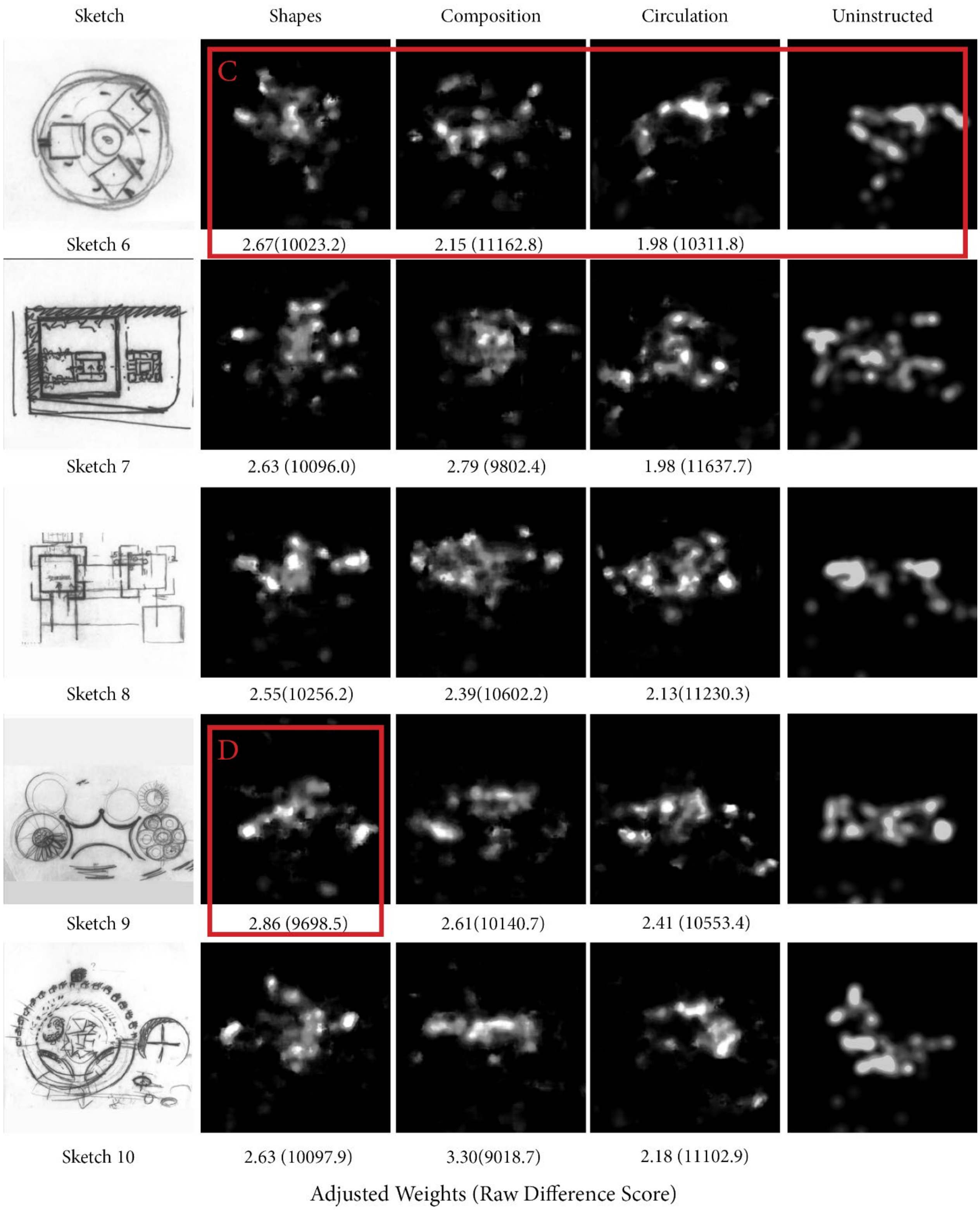


Figure 57. Predicted heat maps (No. 6 – No. 10) and weights comparing to an uninstructed observation data.

v. Analysis of Predicted Heat Maps and Their Weights

Predicted heat maps from the three GANs on shape, composition, and circulation were processed and evaluated using equation 1 and 2. In this section I present four examples marked figure 56 And 57 To explain in detail about potential features in predicted heat maps that affect the weighting.

Heat map A: Circulation prediction for sketch 1, lowest weight (1.58) comparing to the rest of predictions.

This heat map attribute high fixation areas in the sketch to corridor-like areas in the kite-shaped arrangement, which is consistent with the graphical clues found in human heat map data under the intention of circulation. However, the uninstructed result from the human observer instead emphasizes these three main areas: the upper left composition of two squares and a circle, the area between this composition and the upper right circle, and the bottom right triangular shape. These areas are connected through a trail of low fixation areas, and thus contribute to a higher weighting on the composition prediction.

Heat map B: Composition prediction for sketch 4, highest weight (4.43) comparing to the rest of predictions.

This heat map shows a very high, continuous fixation area for the central rectangular area marked with a dark spot. A smaller fixation area appears towards the top of this heat map, corresponding to the upper left corner of the upper rectangle on top of the central shape. In addition, a small tail-like fixation area extends from the center and points towards the lower rectangle. In the uninstructed heat map from the human observer, both the upper and the lower rectangles receive attention and therefore the composition prediction most closely resemble the human result by mainly indicating a vertical composition with main space in the center.

Heat map set C: Shape, composition, and circulation prediction for sketch 6.

This set of heat maps present a new observation in the weighting calculation comparing to appearances of predicted results. At a first glance, the predicted heat map of circulation resembles the most to the uninstructed one due to an extending gesture towards the right of the images. However, when taking the results of shape and composition, it is clearly shown that both heat maps are able to show the other high-fixation areas in the uninstructed result which is pointing towards lower right corner of the image. The shape prediction also indicates the triangular composition of the central circle and the three squares around it, and a small segment of the inscribing circle. Both of them are presence in the human data.

Heat map D: Shape prediction for sketch 9, highest weight (2.86) comparing to composition (2.61) and circulation (2.41) predictions of the same sketch.

This heat map presents another observation between heat map appearance and numeric values from weighting calculation. Again, at a first glance that all three predictions look very similar to that of the human data: a horizontally, semi-symmetrical arrangement of circular shapes. However, the human data has an emphasis on the rightmost circular shapes that only the shape prediction is able to show. On the contrary, the composition prediction emphasized on the left circular shape and the symmetrical aspect of the drawing; and the circulation prediction gives more fixation areas to interior spaces.

vi. Discussion of the Prediction and Weighting Results:

<u>Intention: composition weights the highest among three intentions</u>

The previous four examples provide evidence for an effective method of predicting human view patterns (heat maps) and evaluate their contribution to a real human data graphically. From table xx it is clear that in the ten examples presented, it is more likely for predictions of shape (avg. 2.47) and composition (avg. 2.78) to have higher weights than that of circulation (avg. 2.21). Composition predictions have the highest average weights, and they are the highest weight in sketches 1, 2, 3, 4, 5, 7, and 10 (70% of the sketches). Shape prediction received the highest weights in the rest of the sketches: 6, 8, and 9 (30% of the sketches). One potential explanation behind composition's high weighting is designer' learned tendency to think about overall arrangement when looking at an architectural plan, and such arrangement is better perceived by identifying shapes and their interrelations. On the other hand, circulation is comparatively a more abstract intention that requires further interpretation of the plan sketch by locating possible interior spaces indicated through graphics: as less "visible" comparing to the former two.

<u>Problems and improvements</u>

However, both the weighting calculation and the predictions are in their primitive forms. Calculating the weights is currently limited to pixel-based intensity difference in Euclidean distances. One improvement will be to find a more effective weighting process that takes the areas of different fixation intensity as well as connectivity between same level of fixation areas. Another improvement will be to introduce probabilistic models for measuring how the different intensity areas are likely to be shaped and connected according to different intentions.

If future work continues, it will be important to control the representational appearance of the sketches, for all fifty-five sketches used in the study are not of consistent graphic styles. Some sketches have dark yellow backgrounds while some have light grey backgrounds; line qualities ranged from light construction lines to that of heavy hatches. It will be interesting to explore how the variety of line quality affect both the perception and prediction of heat maps.

4.1.2 "Observation to Gesture" GANs

The second "Observation to Gesture" GANs predicts a sketch-like representation of gestures from lively captured heat maps. This set of GANs reversely generate a sketch-like interpretation of what has been learned by machines to be the most "truthful" result of the image translation. The three networks: shape, composition, and circulation, were trained using heat maps as inputs while the corresponding sketches as prediction. In the limited testing of this network, I used a live-captured heat map frames in a ten-second observation session as input to the three networks and examined how the sequential predictions evolve and form a graphical representation. To give a hint of what this network could achieve, I selected several predicted "gesture" results in the next page (figure 59.).

In the 10s prediction of the shape intention, an arrangement of three larger circular forms is seen in the graphics, with a bilateral symmetry centered on the middle circle. The shape predictions started from rectangular forms on the upper left and gradually morphed into cleaner and definitive circular forms. On the other hand, 10s predictions for composition and circulation have more complex representation of gestures.

More interestingly, when given a blank input as a 256 x 256-pixel white square, the three networks also produce predictions (figure 58). Not only the predicted images have distinct background color, but they also indicate a tendency of global symmetry in all three graphics. For composition and circulation, a "gesture" of a centralized form with emphasis on either top or bottom of the canvas is especially obvious. What these three predictions from a blank input might suggested a "blank" state of visual imagination in plan sketches?

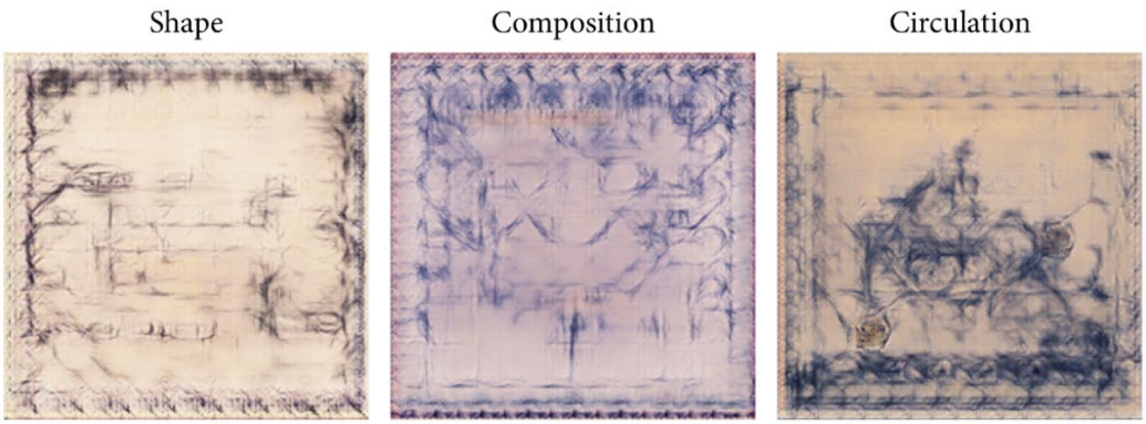


Figure 58. Three examples of prediction from a blank square: shape, composition, and circulation.

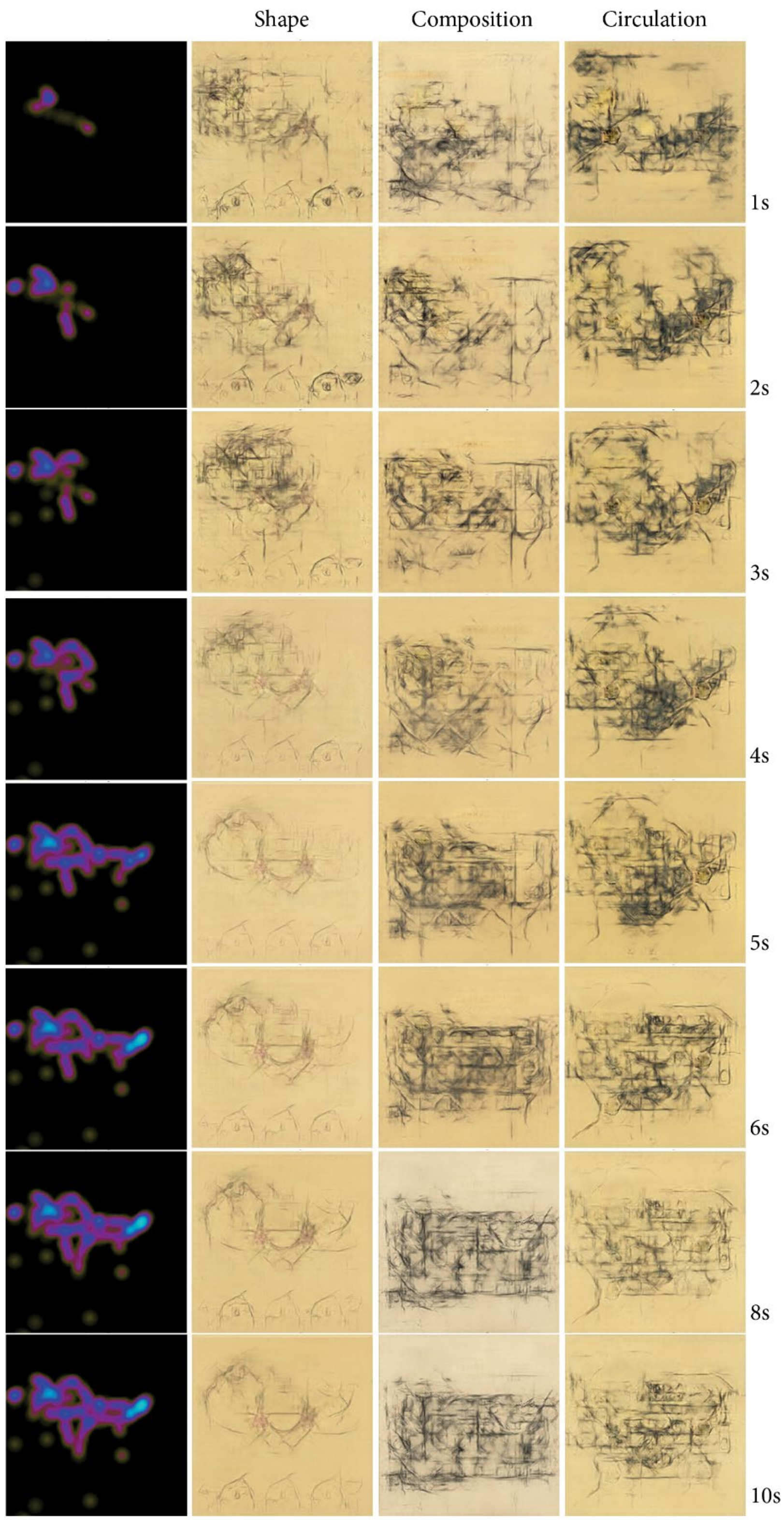


Figure 59. Examples of "Observation to Gesture": generated "gestures" under each intention for the same 10s heat map in at eight time stamps.

4.2 A "Super Network" of Predicting Designers' Viewing Patterns of Intentions

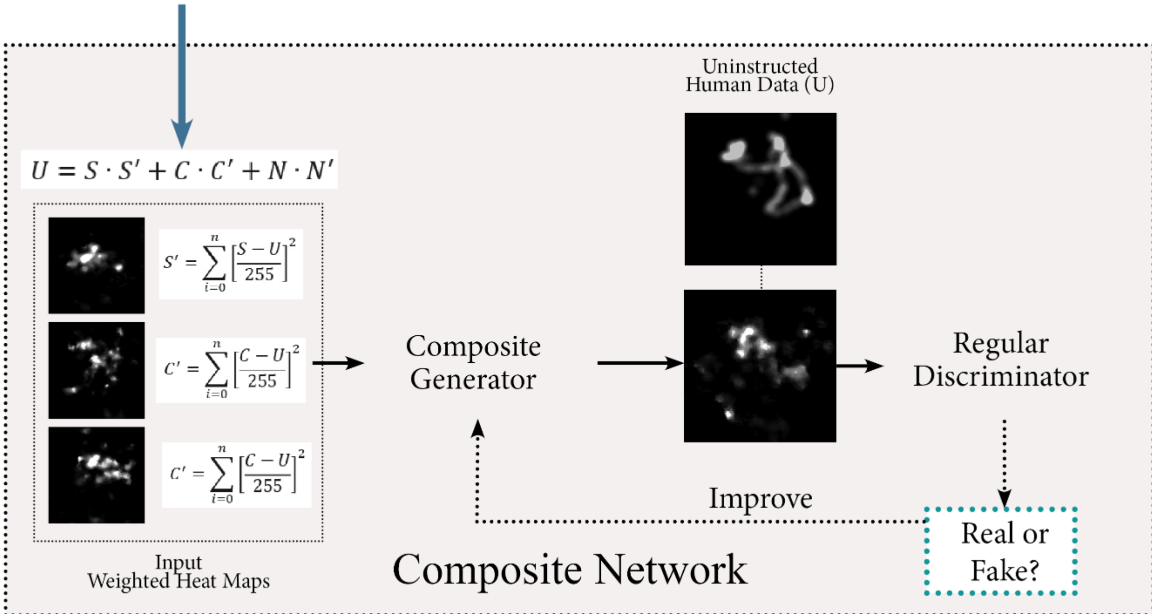


Figure 60. Training pipeline of the "super network" based on Generative Adversarial Networks (Isola et al., 2018).

Following the study of the three separate GANs in predicting and weighting human view patterns, I propose a composite network that can be training through a combination of weighting evaluation and heat map predictions from individual intention networks to produce a human-like interpretation of a given plan sketch. The training data of this "super network" will be a set of weighted predictions from individual intention GANs. These weighted predictions will be processed through a composite generator to compose a heat map using these weights, and to compare with a human observers' heat map result. Then a discriminator will determine whether this composed "human-like" heat map is real or fake, and improve the composite generator based on the result (figure 60.). The composed heat map prediction can be expressed as the following, where S, C, and N are heat map graphics, and S', C', and N' are weights that will be improved through the composite generator.

Due to the time limit of the thesis, I was not able to proceed with this design of a "super network". If future opportunity allows, it will be an interesting exploration to the machines' learning and compositional power which can be obtained through human observational data. The next section consists of a theoretical explanation for this network.

4.3 Architectural Intentions in Visual Representation: An Alternative "Seeing Machine"

> *"The elusive quality of such experiences is hard to capture with our language which commonly describes objects by their tangible, material dimensions. But it is a quality invaluable for abstract thought in that it offers the possibility of reducing a theme visually to a skeleton of essential dynamic features, none of which is a tangible part of the actual object." -- Rudolf Arnheim, "Visual Thinking". 1969.*

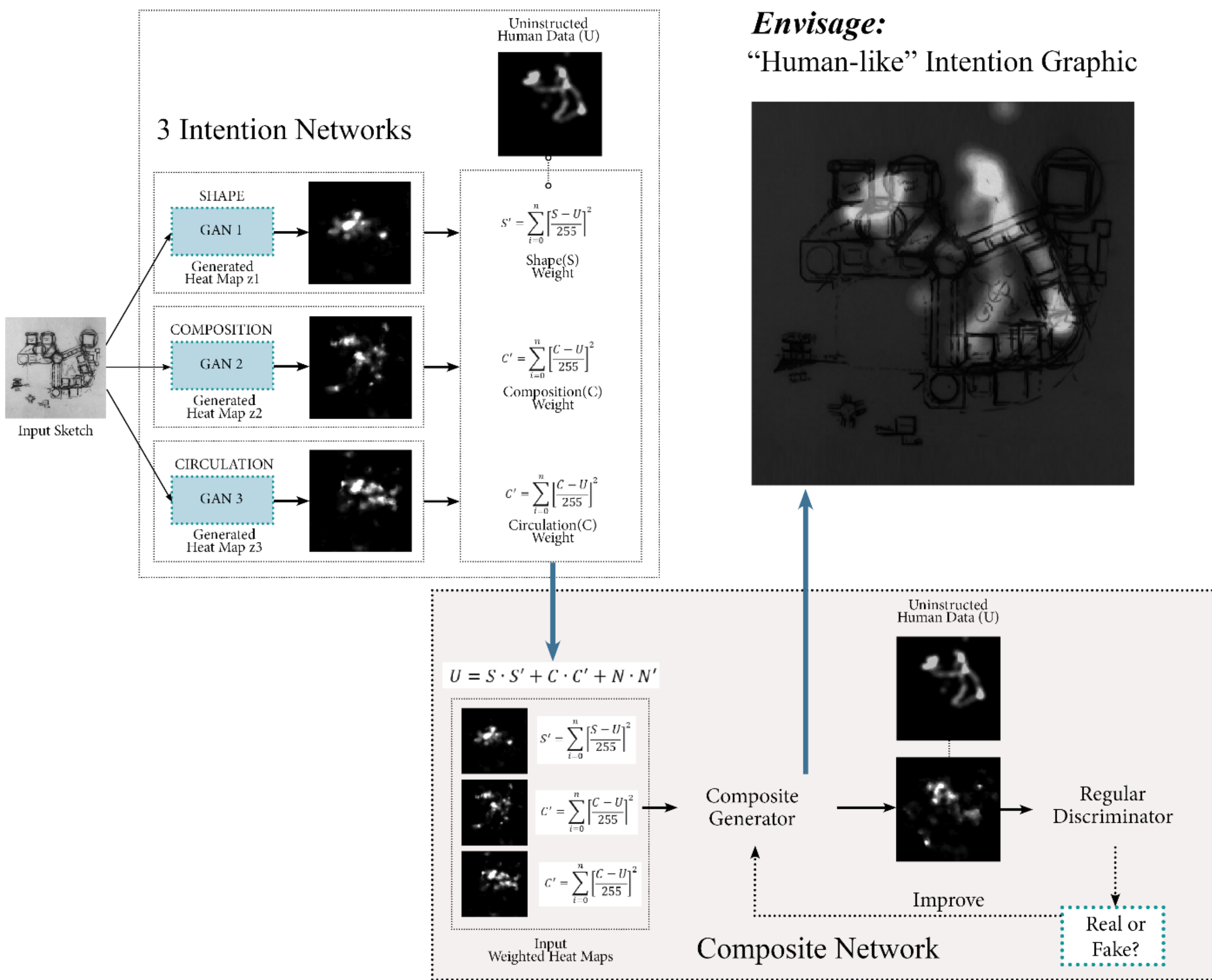


Figure 61. Conceptual diagram of the "super network" – "Envisage"

Reflecting on chapter two's discussion of machines' `eyes", such "elusive quality" of visual experience in design has not been widely acknowledged due to its mysteriously nature and ambiguity. As the adage once said: “A picture is worth a thousand words", not only an image will consist of many recognizable visual elements, but also they are observed and combined together for further imagination by individuals. While how machines will design and what does design mean when encountering any Boolean operation and numeric calculation have been a persistent exploration from both sides -- artists and scientists, one possibility to connect creative observation is to link mechanical movement to potential graphical representation as in the same dimension as visual stimulus. Following the path of this idea, I approach this problem of translation from the angle of extracting abstract representations of architectural intentions in plan sketches. I collected and tested eye-tracking results using Louis Kahn’s sketches from professionally trained architecture students. Meanwhile I examined these sketches and their corresponding heatmaps using both quantitative data analysis and architect’s visual interpretation, supported by the existing connection between deliberate eye movements and goal-oriented visual observational process. From the analysis between scene perception areas indicated by heat maps and the actual sketch segments that facilitates the comprehension of these graphical clues, I summarized how the plan drawings, when seen with different dynamics, are able to gather and recreate corresponding representations for each intention.

At the same time, this connection is indeterministic, meaning a less of one-to-one identification of “correct answer”, but a “some-to-more” extension of possibilities. Therefore, I propose a framework named “Envisage”, which means to not only a retrieval of information using eyes, but a forward-looking representation of graphic potentials. In this framework machines learns through human designers view patterns to assess and determine a graphical representation of design intentions and thus communicate visually using layers of weighted predictions.

i. A Metaphor: Pond, Pebble, and the One Who Throws the Pebble.

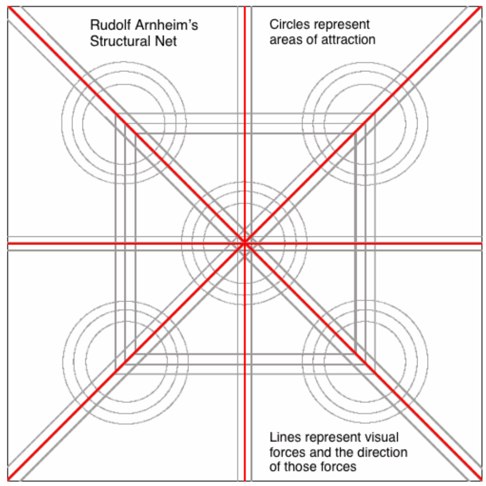


Figure 62. Diagram of Arnheim's magnetic forces on a structural net (Bradley S., 2014).

Returning to the points of visual focus and its derived dynamics to compose architectural intentions, this framework will be built on taking in images as communicators for the next step in design processing. As heat maps and gaze tracking results bring up a skeletal structure of containing associated design information, how this structure is constructed and how it will affect greater utilization in communicating design intentions can be summarized as the following: Partially echoing Arnheim's "elusive quality" (1954), this framework is a "pond" of captured visual entities that serve as "rocks" that are tossed into the water of design and give ripples of design representation. Designer's eye movements are encoded as "rocks" and the water dynamics are parallel to the "intentions" being produced. Progressive interaction among the ripples forms the image of the whole. Instead of asking "what the shape is" for each ripple, the water surface asks for how the ripples collide and morph. The next location for a pebble to be thrown into the pond is decided by the one who observes the pond and throws the pebble.

ii. Input and Output

For both human visual system and machines' operating system's sake, inputs and outputs of this framework are defined as the following: What is the input of calculation here will be human eye movement, whose mechanical nature can be much more precisely captured and delivered by mechanical sensors. The machine takes in these temporal-spatial features of eyes in the observation process. Then these coordinates and directions are processed by a translator to produce real-time graphical representation, such as heat maps into a continuously evolving graphic form of colors and areas. These heat maps become "translators". Potential output will be graphical translations of the input graphics, in a form of designers' choice: whether to reveal intention or generate live "sketches". The translator can be based on detecting low-level features of the input and calculating these features to synthesize outputs, which can be controlled by minimizing losses in most algorithms.

4.4 Discussion: Behind "Envisage": Architectural Design Intention in Humans and Machines

"Envisage" describes both a potential of machines to learn to observe visual design representations from human exports, and a creative process facilitated through image-based generative algorithm. Through the work of chapter three and four, I probed into the mechanical characteristics of human eye movements with perception of architectural intentions from plan sketches by Louis Kahn. I proposed an alternative approach to explain visual perception done by trained architecture professionals by incorporating graphical analysis with visual behaviors determined by fixation time and scanpaths. I demonstrated that a translation of how visually we interpret a design drawing for its innate expressive power of intentions can be assisted and represented through another graphical translation of our eye movement patterns. Such a graphic-to-graphic translation is important because it raises the one-dimensional labeling process of machines' "eyes" to a level of two-dimensional images. Machine is then able to learn how to probe a design image from trained human experts, and potentially delve into further opportunities of communicating design intentions with human partners.

By developing experimentation using image conditional GAN's, I provided evidences of how machines learn from human behavior patterns to produce their "own" interpretation of a design sketch. Although the predictions in the study are rather ambiguous and unrefined, it is more valuable in the sense of visual design that has been an under-appreciated aspect in design representation. In the spirit of designers, and their inherited practice of seeing in a layered, unfeigned way, I advocate for a braver embrace of abstractness and dynamics for the eyes of machines to see and create beyond what is recognizable and what is not. The following two sections provides a detailed explanation for this advocation.

i. Use of Non-Descriptive, Layered Information.

Information processing of visual perception is not limited to distinctive and descriptive qualities such as labels and names, but many times includes a layered composition ("multilevel interrelationships") of less-descriptive visual stimuli. Particularly in the architectural design and education, how to effectively compose and conduct visual information has always been a work of art and science in many generations of professionals and students. Seeking for possible methodologies of transferring designers' visual thinking process to forms which are suitable for computer processors will be a part of the solution. As Alexander Koutamanis (1994) wrote in his paper "Recognition and Retrieval in Visual Architectural Databases" that for digitizing visual qualities and expressions, how effective, efficient and reliable the end result are depends a lot more on human operators' "choices" and "interpretations" of the documents, texts and pictures alike.

Meanwhile, Koutamanis clearly states that "Of architectural documents drawings require particular attention. One reason is that they pose the most intricate interpretation-related problems in computerization" (P17). Another being that "Drawings are probably the most informative sources on the built environment in terms of measurability and accuracy," for they are what "we employ to understand or envisage a building". The verb "envisage" therefore echoes with this critical point for its indication of "imagining something new from things seen". How such "things" are and will be expressed in a digitized manner to reflect a designer's visual thinking process will be studied by three comparative studies using machine learning. However, Koutamanis' progress takes the symbol aspect of visual stimuli to parameterizable planar elements such as doors and walls.

Additionally, Goldschmidt's proposition declares that "visual displays" are an important part of information retrieval for designers via "scrutinizing, intentionally or unintentionally". More importantly, "[b]ringing additional and new pictorial information into the problem space has the potential of restructuring the problem representation." The author then proposed two methods to utilize the visual displays: one to improve "cognitive design operations" by supplying sorted collections of visual display; and the other to weave in the "ill-structuredness mechanisms" to further enhance the information. What "Envisage" suggests an incorporation of these two methods.

ii. Brittleness as Opportunities

Designer's visual process can be partially uncertain, brittle, and abstract. However, these seemly disadvantages are actively facilitating imagination through "visual imagery" (Koutamanis, p. 55). Visual imagery is a cognitive apparatus well-suited to reasoning in tasks of a high degree of novelty (Finke et al.,1992; Kaufmann, 1980). As a tangible, solidified version of visual imagery, design sketches facilitate these functions in enhancing exploration, state transition and transformations.

Uncertainty and brittleness store energies as an ever-flowing database and prepare "for designers to pull the changes and leaps at proper times" (Vermeulen, 1995). These characteristics of designers' visual power have been built upon a complex system for imagination and creativity, and they will do so for designers' computerized counterpart to model and process. Machines usually stop at identifying and calculating problems (firmitas + utilitas), while on the more "aesthetic" (venustas) they have little authority or input. If we can incorporate the emergent quality of visual perception and the sheer number of possibilities and varieties resulted from personal aesthetic experiences into the perceptive ability of machines, it would be thinkable to design a type of system by which human and machine designers exchange and reflect about high-level intentions using visual interactions – a prerequisite for understanding visible beauty of creativity and mutual appreciation.

4.5 Future Work

Further improvements in both the eye-tracking tests will allow for a more comprehensive examination of the view patterns and the resulting heat maps' graphical implication to design intentions. In my study the participants were all professionally trained architecture graduates who have been influenced by a shared educational approach to observe and interpret design sketches. If the view patterns I concluded for these tasks are unique to an expert group, it will be necessary to collect data from non-professional participants to compare and isolate distinctive visual behaviors for perceiving architectural intentions. In addition, there exist many more expressions of architectural intention such as lighting, rhythm, transparency, etc., and even more graphical conventions than those adopted by Louis Kahn to represent building plans. Therefore, another meaningful addition for eye-tracking studies will be to compare differences in fixation points and scanpaths of a variety of design representations of one intention category. For example, use of watercolor sketches, pen drawings, or even primal massing models in contrast with the pencil sketches. Such exploration will likely reveal if the eye movement are rather general across different design representations, or contextual with how intentions are designed to be represented.

Weakness of human visions will be another remaining question to the design of using eye-tracking as a base to deliver graphical interpretations of visual thinking process. Human eyes, although powered by the processing power of our brains, possess many inadequacies when compared to machine vision: such as the small high-resolution area on the retina due to the limitation of fovea, or weak three-dimensional perception when given a two-dimensional stimulus. In the tests of GANs the predictions from machines are rather unfeigned and ambiguous due to the uncertainty in human viewing patterns as a general graphical input, which is performing against the segmentation process in the neural network learning process. Considering the architectural design is essentially a three-dimensional problem, whether the discoveries based on two-dimensional interfaces can be adaptable to a higher dimensional environment requires further studies. Some questions would be: how human designers imagine a 2D scene differently than a 3D scene? What would the equivalent of ambiguity in a 2D sketch in a 3D environment since 3D environment is generally consisted of defined objects? Whether designers' vision still plays a similar role in a 3D environment, and if machines essentially perform better instead?

The strength of machine visions cannot be overlooked. Although a majority of machine vision tasks has been focusing on precisely and actively identifying and segmenting objects or their attributes from a given scene, it will be extremely hasty to conclude that vision tasks empowered by these segmenting algorithms will always lack of power to be creative. Creativity has been appraised in human perception of beauty and teleology that ascend our impression of our species to be somehow special and hard to be surpassed or redefined. Especially in the rising era of Artificial Intelligent and human-machine partnership, it is even more crucial than before to inquire and communicate between the two sides. Emotive or rational, experiential, or experimental, our creativity and its meaning to us and machines alike might be continuously challenged. However, from the discourses we (humans and machines) always find collaborative opportunities. Especially in the field of architectural design, it is time now to communicate further between to two sides on what aspects of our perceptive process shapes our creativity, and how we move further as mutual-understanding companions. Afterall, our future requires us to "envisage" together with beautiful intentions.

4.6 Contributions

1. Identified a problem between recognition of labels and units in design and freer and dynamic graphic expression of design intentions and proposed a solution through examine observational processes.

The first chapters investigate both from a historical and designer's perspective about the innovative and dynamic observational process enabled by active, goal-oriented observation in the perception and expression of architectural intentions. I provided a retrospective of the methods and graphical representations mastered by great architects in assisting of their creative power: both for recognizing design elements and associating personal experiences. I compared how the rise of scientific tools such as photographing techniques has broadened the mindset of recording and capturing objectivity; and how the last century's progression on machine intelligence expanded the pursuit of "true-to-nature" and the subsequent search-based design-solving methods -- such as BIM and Artificial Intelligence. To explore a potential link between the dynamic intention and recognizable representations in design sketch, I proposed a method of translating active observational process of designers into a graphical representation which would allow a type of communication to be less relied on exact labels, but from direct perceptions done by goal-oriented observation behaviors.

2. Conducted an eye-tracking study which produced meaningful results in:

Identifying intention-specific graphical characteristics in heat maps from professional architecture students' observation responses of three architectural intentions: shape, composition, and circulation (navigation).

Providing evidence in graphically meaningful clues which facilitate translation of the abstract intentions through view patterns: how differently the reading of graphical clues can imply the formation of intentions from a single sketch; and how mechanical viewing patterns built up areas of interests which are architecturally expressive.

The goals of chapter three are to both examine and attest how willful observation could serve as a translator between a visual stimulus and its stored visual information as perceived as architectural intention. I introduced an alternative perspective to interpret heat maps from gaze-tracking as a visual expression in graphic form. Designers' observations possess an active force that pull the embedded information from images through how and where they would like to see and probe. Such visual power functions beyond identifying objects to a level of mind-eye interaction that facilitate creative perception and imagination.

3. Demonstration of an initial step towards utilizing graphical representation of human observation behavior to perceive or extract abstract visual information in design. Envisaged an application that take advantage of the dynamic ambiguous graphical translation obtained from eye-tracking tests.

In the fourth chapter I inquired on the question of visual perception and its extension into a reciprocal, human-machine linkage problem. I propose a prototype application to enable a translation between intentional observation and expressions of abstract architectural qualities in shape, composition, and circulation. Using a machine learning algorithm: image-conditional Generative Adversarial Networks, I trained three separate networks to predict each intention tested in chapter three from collected human participants' data; and examined their similarity and compositional weights against uninstructed human heat map data. The three trained networks performed decently considering training and testing data's limitation in resolution and accuracy: under the perspective of "Envisage", they predict distinct heat maps which shared graphical characteristics discovered in chapter three. Some predictions even present a high similarity to human data. Although the weighting algorithm was rather primitive and what the heat maps were really indicating beyond pure graphical clues still needs further investigation, the results from these three "translators" were inspiring in the sense of bringing more willful and magical human behaviors into the learning process of image-based machine algorithms. To conclude chapter four, I envisioned a framework of future to promote visual communication between humans and machines on another level of perceiving intentions not through matching of patterns and symbols, but through the matching of eyes.

Bibliography

Ackerman, J. S. (2016). *Origins, invention, revision: Studying the history of art and architecture.* Yale University Press.

Adler, A., & Davis, R. (2007). Speech and sketching: An empirical study of multimodal interaction. *Proceedings of the 4th Eurographics Workshop on Sketch-Based Interfaces and Modeling*, 83–90. https://doi.org/10.1145/1384429.1384449

Ahlquist, S. (2020). Negotiating human engagement and the fixity of computational design: Toward a performative design space for the differently-abled bodymind. *International Journal of Architectural Computing*, *18*(2), 174–193. https://doi.org/10.1177/1478077120919850

Ahuja, S., & Chopson, P. (2020). Automation and Machine Learning in Architecture: A new agenda for performance-driven design. *Architectural Design*, *90*(2), 104–111. https://doi.org/10.1002/ad.2553

Alberti, L. B. (1988). *On the art of building in ten books (T. Rykwert, J., Leach, N. & Tavernor, R. Trans.).* The MIT Press. (Original work published 1485)

Andy J. K., Nadine Bol, R. Glenn C. & Kevin K. J. (2019). Improving visual behavior research in communication science: An overview, review, and reporting recommendations for using eye-tracking methods. *Communication Methods and Measures*, 13:3, 149-177, DOI: 10.1080/19312458.2018.1558194

Arnheim, R. (1954). *Art and visual perception: The psychology of the creative eye.* University of California Press.

Artemel, AJ. (2021. April 20). The Master composer: 17 collages and drawings by Ludwig Mies van der Rohe. *Architizer.* https://architizer.com/blog/inspiration/industry/mies-van-der-rohe-collages/

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, 61 (3), 183-193.

Baker, G. H. (2001). *Le Corbusier - The Creative Search: The Formative Years of Charles-Edouard Jeanneret.* Taylor & Francis.

Blum, H. (1967). A transformation for extracting new descriptors of shape. In: Wathen-Dunn, W. (Ed.). *Models for the perception of speech and visual form.* The MIT Press, Cambridge, pp 362-380

Bradley, S. (2014, December 12). Design Principles: Visual Weight and Direction. *Smashingmagazine*. https://www.smashingmagazine.com/2014/12/design-principles-visual-weight-direction/#:~:text=Visual%20weight%20is%20a%20measure,the%20greater%20its%20visual%20weight

Brillhart, J. (2018, October 6). Drawing on the road: The story of a young Le Corbusier's travels through Europe. *Archdaily.* https://www.archdaily.com/784616/drawing-on-the-road-the-story-of-a-young-le-corbusiers-travels-through-europe

Brothers, C. (c2008.). *Michelangelo, drawing, and the invention of architecture*. Yale University Press.

Buckinghan-Hisao, R. (2018). Drawn words: Pictographs in the Chinese language and visual culture. *Drawing: Research, Theory, Practice*, 3(2), pp. 229-240. https://doi.org/10.1386/drtp.3.2.229_1

Burch, M., Chuang, L., Fisher, B., Schmidt, A., & Weiskopf, D. (2015). *Eye tracking and visualization: Foundations, techniques, and applications.* Springer.

Carte, B. T., & Luke, S. G. (2020). Best practices in eye tracking research. International Journal of Psychophysiology, 155, 49-62. https://doi.org/10.1016/j.ijpsycho.2020.05.010

Chaillou, S. (2020). ArchiGAN: Artificial intelligence x architecture. In: Yuan P., Xie M., Leach N., Yao J., Wang X. (Eds). *Architectural Intelligence.* Springer, Singapore. https://doi.org/10.1007/978-981-15-6568-7_8

Da Vinci, L. (1955). *Das lebensbild eines genies.* Emil Vollmer Verlag, Wiesbaden Berlin.

Davis, R. (1998). What are intelligence? And why? 1996 AAAI presidential address. *AI Magazine*, 19(1), 91–110.

Dickinson, S. J., & Pizlo, Z. (Ed.). (2013). *Shape perception in human and computer vision: an interdisciplinary perspective.* Springer.

Duchowski, A.T. (2017). *Eye tracking methodology* [eBook edition]. Springer. https://link.springer.com/book/10.1007%2F978-3-319-57883-5
Dulaney, R., & Lyn, F. (2010). Representational craft and production: Comparison of the value of hand drawing and digital media in architectural academies and practices. *Design Principles & Practice: An International Journal*. https://lib.mit.edu/record/asu/66386392
Eastman, C, M. (1969). Cognitive processes and ill-defined problems: A case study from design. *IJCAI'69: Proceedings of the 1st international joint conference on Artificial intelligence*. ACM Digital Library. https://dl.acm.org/doi/10.5555/1624562.1624622
Elgammal, A., Mazzone, M., Liu, B., Kim, D., & Elhoseiny, M. (2018). The shape of art history in the eyes of the machine. *ArXiv,* 1801.07729 [Cs]. http://arxiv.org/abs/1801.07729
Eye Movement in Reading. (2020, August 30). In *Wikipedia.* https://en.wikipedia.org/w/index.php?title=Eye_movement_in_reading&oldid=975808659
Frey, B. J. (1998). *Graphical models for machine learning and digital communication*. The MIT Press.
Galle, P. (1994). Computer support of architectural sketch design: A matter of simplicity? *Environment and Planning B: Planning and Design,* 21(3), 353–372. https://doi.org/10.1068/b210353
Gallov, G. (2019). Observations on drawing: The art of architecture. *Architectural Design.* https://doi.org/info:doi/10.1002/ad.2501
GazeRecorder. (2021). *FAQ.* GazeRecorder. https://gazerecorder.com/faq/
Gil-Mastalerczyk, J. (2015). The significance of drawing and painting in architectural design (as exemplified by Le Corbusier's sacred architecture).
Goel, V. (1988). Complicating the 'logic of design.' *Design Studies*, 9(4), 229–234. https://doi.org/10.1016/0142-694X(88)90008-7
Google. (2021, March 6). *How search by image works* [Video]. YouTube. https://www.youtube.com/watch?v=keTZaJg0784.
Gresleri, G. (2000). *Le Corbusier: Carnets, les voyages d'allemagne voyage D'orient*. Electa.
Halpern, O., & Halpern, O. (2014). *Beautiful data: A history of vision and reason since 1945* [EBook].. https://lib.mit.edu/record/nlebk/991585
Harris, S. (2015). Le Corbusier between sketches: A graphic analysis of the Acropolis sketches. *LC2015 – Le Corbusier, 50 years later*. DOI:10.4995/LC2015.2015.911
Huang, W., & Zheng, H. (2018). Architectural drawings recognition and generation through machine learing. *ACADIA 2018*. http://papers.cumincad.org/data/works/att/acadia18_156.pdf
Ilyas, A., Santurkat, S., Tsipras, D., Engstrom, L., Tran, B., & Madry, A. (2019). Adversarial examples are not bugs, they are features. *arXiv:1905.02175 [stat.ML]*
Isola, P., Zhu, J., Zhou, T., & Efros, A. A. (2018). Image-to-image translation with conditional adversarial networks. *arXiv:1611.07004[cs.CV]*. https://arxiv.org/abs/1611.07004
Jahanian, A., Keshvari, S., & Rosenholtz, R. (2018). Web pages: What can you see in a single fixation? *Cognitive Research: Principles and Implications,* 3, 14. https://doi.org/10.1186/s41235-018-0099-2
Jensen, M. B., Foged, I. W., & Andersen, H. J. (2020). A framework for interactive human–robot design exploration. *International Journal of Architectural Computing*, *18*(3), 235–253. https://doi.org/10.1177/1478077120911588
Joachim, H. (1981). The Drawings of Andrea Palladio. *Bulletin of the Art Institute of Chicago* (1973-1982), 75(3), 8-9. Retrieved May 14, 2021, from http://www.jstor.org/stable/4104008
Karstworlds. (2018) Lascaux cave: History. *www.karstworlds.com*. Retrieved 26 November 2018. http://www.karstworlds.com/2010/12/lascaux-cave-history.html
Kepes, G. (1965). *Education of vision.* New York: G.Braziller.
Kilian, A. (2000). Defining digital space through a visual language. [Master's thesis, Massachusetts Institute of Technology]. MIT DSpace.
Knight, T., & Stiny, G. (2015). Making grammars: From computing with shapes to computing with things. *Design Studies*, 41(A), 8-28. https://doi.org/10.1016/j.destud.2015.08.006

Koile, K. (1997). Design conversations with your computer: Evaluating experiential qualities of physical form. In R. Junge (Ed.), CAAD futures 1997 (pp. 203–218). Springer Netherlands. https://doi.org/10.1007/978-94-011-5576-2_16
Koile, K. (2004). An intelligent assistant for conceptual design. In J. S. Gero (Ed.), *Design Computing and Cognition '04* (pp. 3–22). Springer Netherlands. https://doi.org/10.1007/978-1-4020-2393-4_1
Koile, K. (2006). Formalizing abstract characteristics of style. *AI EDAM*, *20*(3), 267–285. https://doi.org/10.1017/S0890060406060203
Koutamanis, A. (1994) Recognition and retrieval in visual architectural databases. *CUMINCAD.*
Koutamanis, A., Timmermans, H., & Vermeulen, I. (1995). Visual database in architecture: Recent *advances in design and decision making.* Avebury.
Lafuente Sánchez V.A., López Bragado D. (2019) A new approach to architectural representation according to the principles of Gestalt perception. In: Marcos C. (eds) *Graphic Imprints.* EGA 2018. Springer, Cham. https://doi.org/10.1007/978-3-319-93749-6_120
Li, F., Johnson, J., & Yeung, S. (2017). Lecture 11: Detection and segmentation [Slides]. *Stanford University.* http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture11.pdf
Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology*, 5(5), 552 – 563. https://doi.org/10.1016/S0960-9822(95)00108-4
Lorenzo-Eiroa, P., & Sprecher, A. (2013). *Architecture in Formation: On the Nature of Information in Digital Architecture* [EBook]. Architecture in Formation : On the Nature of Information in Digital Architecture. https://lib.mit.edu/record/nlebk/649844
Lowe, D. G. (2019). Object recognition from local scale-invariant features. *Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece*, vol.2, 1150-1157. doi: 10.1109/ICCV.1999.790410.
Lyokoï88. (2015). Français : Gros plan de la statue de la victoire. *Wikimedia Commons.* https://commons.wikimedia.org/wiki/File:Victoire_de_Samothrace_-_vue_de_trois-quart_gauche,_gros_plan_de_la_statue_(2).JPG
Marinčić, N. (2019). Computational models in architecture: Towards communication in CAAD. Spectral characterization and modelling with conjugate symbolic domains. *Birkhäuser.* https://www-degruyter-com.libproxy.mit.edu/view/title/542043
Marbleicons (n.d.). [Your story immortalized in marble]. Retrieved May 13, 2020, from https://www.marbleicons.com
Morris, C. W. (1939). Esthetics and the theory of signs. *The Journal of Unified Science*, 8, 131-150.
Merrill. (2010). *Louis Kahn: On the Thoughtful Making of Spaces: The Dominican Motherhouse and a Modern Culture of Space.* Lars Müller Publishers.
Mcgrath, B., & Gardner, J. (2008). Cinemetrics: Embodying Architectural Representation in the Digital Age. *Architectural Theory Review*, *13*, 29–51. https://doi.org/10.1080/13264820801915104
Negroponte, N. (1970). *The Architecture Machine: Toward a More Human Environment.* MIT Press.
Orquin, J. L., Ashby, N. J. S., & Clarke, A. D. F. (2016). Areas of interest as a signal detection problem in behavioral eye-tracking research. *Journal of Behavioral Decision Making*, 29, 103-115.
Papadopoulou, A. (2014). *Perceptual prototypes: Towards a sensory pedagogy of space* [Thesis, Massachusetts Institute of Technology]. https://dspace.mit.edu/handle/1721.1/91411
Park, J. H. (2015). *Synthetic tutor: Profiling students and mass-customizing learning processes dynamically in design scripting education.* [Thesis, Massachusetts Institute of Technology]. https://dspace.mit.edu/handle/1721.1/101544
Pauly, D. (2018). *Le Corbusier: Drawing as Process*. Yale University Press.
Peng, W. (2018). Machines' perception of space. [Master's thesis, Massachusetts Institute of Technology]. MIT DSpace.
Photohistoryreview. (2008, September 17). *Invention.* Photography Project. http://photohistoryreview.blogspot.com/2008_09_01_archive.html

Pilsitz, M. (2017). Drawing and Drafting in Architecture Architectural History as a Part of Future Studies. *Periodica Polytechnica: Architecture*. https://doi.org/info:doi/10.3311/PPar.11310
Quotes of Michelangelo. (n.d.). Michelangelo. Retrieved April 13, 2021, from https://www.michelangelo.org/michelangelo-quotes.jsp#:~:text=The%20marble%20not%20yet%20carved,thought%20the%20greatest%20artist%20has.%20%E2%80%9D&text=Trifles%20make%20perfection%2C%20and%20perfection%20is%20no%20trifle.%20%E2%80%9D&text=A%20beautiful%20thing%20never%20gives,to%20hear%20and%20see%20it.%20%E2%80%9D
Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception and visual search. *The Quarterly Journal of Experimental Psychology*, 62 (8), 1457 - 1506.
Reisner, Y. (2019). Abstraction and Informality Generate a New Aesthetic: An Interview with Kazuyo Sejima. *Architectural Design*, *89*(5), 30–37. https://doi.org/10.1002/ad.2476
Reverse Image Search. (2021, February 21). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Reverse_image_search&oldid=1008027369
Rezanejad M., & Siddiqi K. (2013). Flux graphs for 2D shape analysis. In: Dickinson S., Pizlo Z. (Eds.) *Shape perception in human and computer vision. advances in computer vision and pattern recognition* (pp. 41-54). Springer, London. https://doi.org/10.1007/978-1-4471-5195-1_3
Roberts, L. G. (1961). Machine perception of three-dimensional solids. [Master's thesis, Massachusetts Institute of Technology]. MIT DSpace.
Rodriguez, M. (2014). Transnational architecture and new femininity. *Open Cities: The New Post-Industrial World Order.* https://www.acsa-arch.org/chapter/transnational-architecture-and-newfemininity
Sung, W. (2013). Sketching in 3D: Towards a fluid space for mind and body. [Master's thesis, Massachusetts Institute of Technology]. MIT DSpace.
Sánchez, V. A. F., & Bragado, D., L. (2019) A new approach to architectural representation according to the principles of gestalt perception. In Marcos, C.L. (Ed.). *Graphic Imprints: The influence of representation and ideation tools in architecture,* Springer*,* pp. 1449 – 1450.
Sandoval, C., Pirogova, E., & Lech, M. (2019). Two-Stage Deep Learning Approach to the Classification of Fine-Art Paintings. *IEEE Access*, *7*, 41770–41781. https://doi.org/10.1109/ACCESS.2019.2907986
Schlemmer, O. (1971). Dairy entry dated February 1925. In Kuckling, H. (Ed.). *Man: Teaching Notes from the Bauhaus.* The MIT Press.
Semmelmann, K., & Weiglt, S. (2017). Online webcam-based eye-tracking in cognitive science: A first look. *Behavior Research Methods,* 50, 451 - 465. https://doi.org/10.3758/s13428-017-0913-7
Shehu, I., Wang, Y., Athuman, A., & Fu, X. (2020). Paradigm shift in remote eye gaze tracking research: highlights on past and recent progress. In Kacprzyk, J. (Ed.), *Advanced in Intelligent Systems and Computing,* Springer Press, Switzerland, pp. 171.
Smith, S. M., Ward, T. B., & Finke, R. A. (2009). *The Creative Cognition Approach*. MIT Press.
Stiny, G. (1981). A Note on the Description of Designs. *Environment and Planning B: Planning and Design*, 8(3), 257–267. https://doi.org/10.1068/b080257
Stiny, G. (2006). *Shape: Talking about Seeing and Doing*. The MIT Press. https://doi.org/10.7551/mitpress/6201.001.0001
Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVP*R), 2818–2826. https://doi.org/10.1109/CVPR.2016.308
Tamke, M., Nicholas, P., & Zwierzycki, M. (2018). Machine learning for architectural design: Practices and infrastructure*. International Journal of Architectural Computing,* 16(2), 123–143. https://doi.org/10.1177/1478077118778580

Tausky, D., Labahn, G., Lank, E., & Marzouk, M. (2007). Managing ambiguity in mathematical matrices. *Proceedings of the 4th Eurographics Workshop on Sketch-Based Interfaces and Modeling,* 115–122. https://doi.org/10.1145/1384429.1384454

Wagemans J. (2013) Two-dimensional shape as a mid-level vision gestalt. In: Dickinson S., Pizlo Z. (Eds.). *Shape Perception in Human and Computer Vision. Advances in Computer Vision and Pattern Recognition.* Springer, London. https://doi-org.libproxy.mit.edu/10.1007/978-1-4471-5195-1_6

Wikipedia, (n.d.). [The Atlas Slave]. Retrieved May 13, 2020, from https://en.wikipedia.org/wiki/Atlas_Slave#/media/File:Michelangelo_-_Atlas.jpg

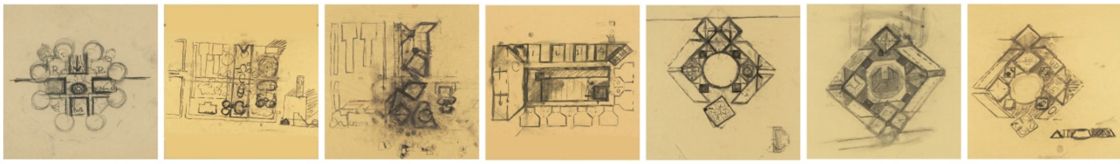
Wild, A. (2011, August 16). Google's Reverse Image Search*. Scientific American*. https://blogs.scientificamerican.com/compound-eye/googles-reverse-image-search/

Vlavianos, N. (2016). Shape Grammar Reality (SGr): Computing in the real world. [Master's thesis, Massachusetts Institute of Technology]. MIT DSpace.

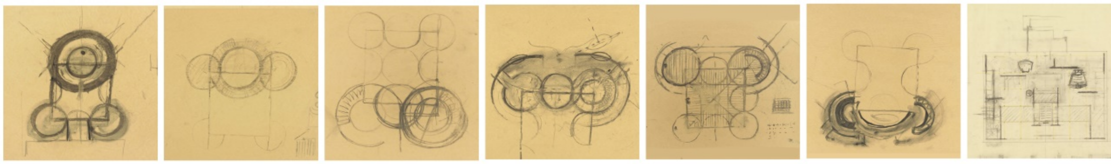
Vidler, A. (2000). Diagrams of Diagrams: Architectural Abstraction and Modern Representation. *Representations*, 72, 1–20. JSTOR. https://doi.org/10.2307/2902906

Appendix I: Catalog of Louis Kahn’s Sketches Used Eye-tracking and GANs Studies

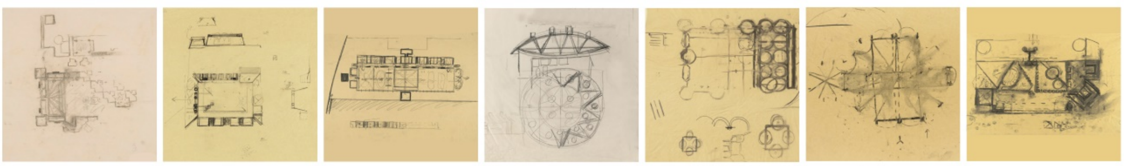
Images



01 02 03 04 05 06 07

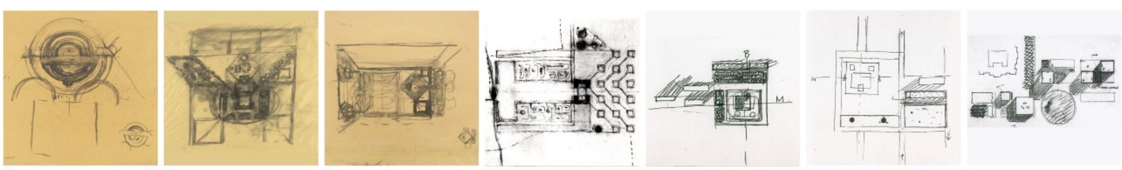


08 09 10 11 12 13 14

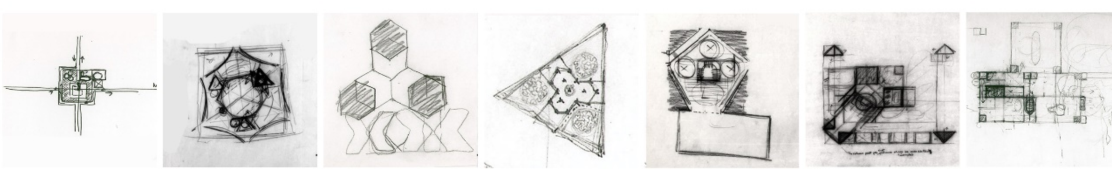


n01 n02 n03 n06 n07 n08 n09

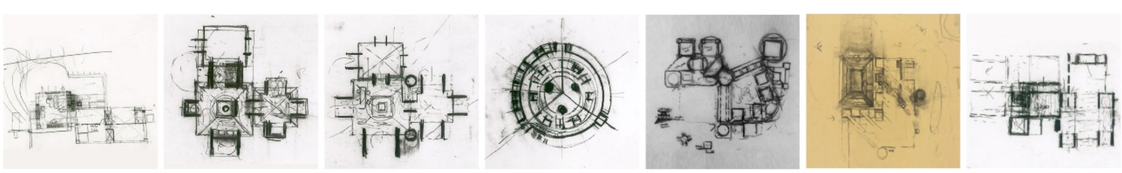
n10 n11 n12 n13 n14 n15 n16



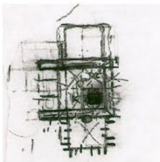
n17 n18 n19 p01 p02 p03 p04



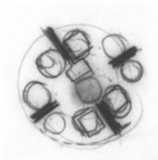
p05 p06 p07 p08 p09 p10 p11



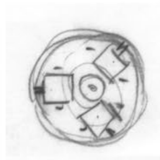
p12 p13 p14 p15 pl1 pl2 pl3



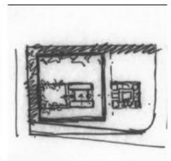
pl4



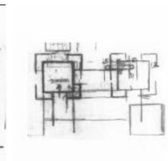
pl5



pl6



pl7



pl8

pl9

pl10

Image Titles

Image Name	Sketch Information
1	Eleanor Dannelley erdman Hall, Bryn Mawr College, Bryn Mawr, Pennsylvania, Plan sketches
2	Philadelphia College of Art, project, Philadelphia, Pennsylvania, Site-plan and elevation sketches. 1965
3	Philadelphia College of Art, project, Philadelphia, Pennsylvania, Site-plan sketches. 1965
4	Indian Institute of Managementm, Ahmedabad, India, Plan sketch of classroom building. 1964
5	Sher-e-Bangla Nagar, Capital of Bangladesh, Dhaka, Bangladesh, National Assembly Building: plan sketch. 1963
6	Sher-e-Bangla Nagar, Capital of Bangladesh, Dhaka, Bangladesh, National Assembly Building: plan sketch. 1964
7	Sher-e-Bangla Nagar, Capital of Bangladesh, Dhaka, Bangladesh, National Assembly Building: plan sketch. 1963
8	Sher-e-Bangla Nagar, Capital of Bangladesh, Dhaka, Bangladesh, Prayer Hall: plan sketch. 1965
9	Sher-e-Bangla Nagar, Capital of Bangladesh, Dhaka, Bangladesh, Prayer Hall: plan sketch. 1965
10	Sher-e-Bangla Nagar, Capital of Bangladesh, Dhaka, Bangladesh, Prayer Hall: plan sketch. 1965
11	Sher-e-Bangla Nagar, Capital of Bangladesh, Dhaka, Bangladesh, Prayer Hall: plan sketch. 1965
12	Sher-e-Bangla Nagar, Capital of Bangladesh, Dhaka, Bangladesh, Prayer Hall: plan sketch. 1965
13	Sher-e-Bangla Nagar, Capital of Bangladesh, Dhaka, Bangladesh, Prayer Hall: plan sketch. 1965
14	Ludwig Mies van der Rohe. IIT Master Plab, Chicago, Illinois, Student Union building. Plan. Partial. 1939
n01	Plan of the Buildings of the Indian Institute of Management, Ahmedbad
n02	Plan of the Buildings of the Indian Institute of Management, Ahmedbad, plan and elevation
n03	Tribune Review Publishing Company Building, Greensburg, Pennsylvania, Plan sketch1958
n06	Louis I. Kahn .First Unitarian Church and School, Rochester, New York, Plan and elevation sketches1959
n07	Louis I. Kahn. Mikveh Israel Synagogue, project, Philadelphia, Pennsylvania, Plan sketches1963
n08	Louis I. Kahn. Mikveh Israel Synagogue, project, Philadelphia, Pennsylvania, Ceiling sketch1965
n09	Louis I. Kahn. Fine Arts Center, School, and Performing Arts Theater, Fort Wayne, Indiana, Site-plan sketch1963
n10	Louis I. Kahn. Fine Arts Center, School, and Performing Arts Theater, Fort Wayne, Indiana, Site-plan and elevation sketches1963
n11	Louis I. Kahn. Fine Arts Center, School, and Performing Arts Theater, Fort Wayne, Indiana, Site-plan sketch1963

n12	Louis I. Kahn. Indian Institute of Management, Ahmedabad, India, Plan sketches of clssroom building1963
n13	Louis I. Kahn. Indian Institute of Management, Ahmedabad, India, Plan sketches of clssroom building1963
n14	Louis I. Kahn. Sher-e-Bangla Nagar, Capital of Bangladesh, Dhaka, Bangladesh (Plan sketch)1963
n15	Louis I. Kahn. Sher-e-Bangla Nagar, Capital of Bangladesh, Dhaka, Bangladesh, National Assembly Building: plan sketch1963
n16	Louis I. Kahn Fine Arts Center, School, and Performing Arts Theater, Fort Wayne, Indiana, Site-plan sketch 1963
n17	Louis I. Kahn. Sher-e-Bangla Nagar, Capital of Bangladesh, Dhaka, Bangladesh, Prayer Hall: plan sketch1965
n18	Louis I. Kahn. Sher-e-Bangla Nagar, Capital of Bangladesh, Dhaka, Bangladesh (Plan sketch)1963
n19	Louis I. Kahn. Indian Institute of Management, Ahmedabad, India, Plan
p01	Project: Abbasabad Development
p02	Penn Center Studies
p03	Penn Center Studies
p04	Penn Center Studies
p05	Penn Center Studies
p06	City Tower Project
p07	City Tower Project, Floor plan, section
p08	City Tower Project, Floor plan, section
p09	Peabody Museum, alterations and additions (Hall of Ocean Life)
p10	Kansas City Office Building
p11	Adler Residence, Plan studies
p12	Adler Residence, Plan studies
p13	Morris Residence, detail plan
p14	Morris Residence, detail plan
p15	Civic Center Studies
pl1	President's Estate at the First Capital of Pakistan. Islamabad, 1966.
pl2	Indian Institute of management, Ahmedabad, India, Plan sketch of classroom building, 1963
pl3	Morris Residence
pl4	Morris Residence
pl5	Salk Institute for Biological Studies, Site plan study
pl6	Salk Institute for Biological Studies, Site plan study
pl7	United States Consulate and Residence, Site and floor plan of chancellery, section
pl8	Erdman Hall, Plan, various studies
pl9	General Motors Exhibit, 1964 World's Fair, Diagrammatic site plan, elevation
pl10	General Motors Exhibit, 1964 World's Fair, Partial floor plan

Image Sources

Image Name	Source link
1	Louis I. Kahn. Eleanor Donnelley Erdman Hall, Bryn Mawr College, Bryn Mawr, Pennsylvania, Plan sketches. 1960-61 | MoMA
2	Louis I. Kahn. Philadelphia College of Art, project, Philadelphia, Pennsylvania, Site-plan and elevation sketches. 1965 | MoMA
3	https://www.moma.org/collection/works/626
4	Louis I. Kahn. Indian Institute of Managementm, Ahmedabad, India, Plan sketch of classroom building. 1964 | MoMA
5	https://www.moma.org/collection/works/518
6	Louis I. Kahn. Sher-e-Bangla Nagar, Capital of Bangladesh, Dhaka, Bangladesh, National Assembly Building: plan sketch. 1964 | MoMA
7	Louis I. Kahn. Sher-e-Bangla Nagar, Capital of Bangladesh, Dhaka, Bangladesh, National Assembly Building: plan and elevation sketches. 1963 | MoMA
8	Louis I. Kahn. Sher-e-Bangla Nagar, Capital of Bangladesh, Dhaka, Bangladesh, Prayer Hall: plan sketch. 1965 | MoMA
9	Louis I. Kahn. Sher-e-Bangla Nagar, Capital of Bangladesh, Dhaka, Bangladesh, Prayer Hall: plan sketch. 1965 | MoMA
10	Louis I. Kahn. Sher-e-Bangla Nagar, Capital of Bangladesh, Dhaka, Bangladesh, Prayer Hall: plan sketch. 1965 | MoMA
11	Louis I. Kahn. Sher-e-Bangla Nagar, Capital of Bangladesh, Dhaka, Bangladesh, Prayer Hall: plan sketch. 1964 | MoMA
12	Louis I. Kahn. Sher-e-Bangla Nagar, Capital of Bangladesh, Dhaka, Bangladesh, Prayer Hall: plan sketch. 1965 | MoMA
13	Louis I. Kahn. Sher-e-Bangla Nagar, Capital of Bangladesh, Dhaka, Bangladesh, Prayer Hall: plan sketch. 1965 | MoMA
14	Ludwig Mies van der Rohe. IIT Master Plan, Chicago, Illinois, Student Union building. Plan. Elevation. Two exterior perspectives. 1939 | MoMA
n01	https://www.sothebys.com/en/auctions/ecatalogue/2018/boundless-india-in1801/lot.6.html
n02	https://www.moma.org/collection/works/596
n03	https://www.moma.org/collection/works/414?artist_id=2964&page=1&sov_referrer=artist
n06	https://www.moma.org/collection/works/442
n07	https://www.moma.org/collection/works/489
n08	https://www.moma.org/collection/works/503
n09	https://www.moma.org/collection/works/529
n10	https://www.moma.org/collection/works/534
n11	https://www.moma.org/collection/works/485
n12	https://www.moma.org/collection/works/592
n13	https://www.moma.org/collection/works/592
n14	https://www.moma.org/collection/works/507
n15	https://www.moma.org/collection/works/543
n16	https://www.moma.org/collection/works/485
n17	https://www.moma.org/collection/works/584
n18	https://www.moma.org/collection/works/512
n19	https://www.moma.org/collection/works/600

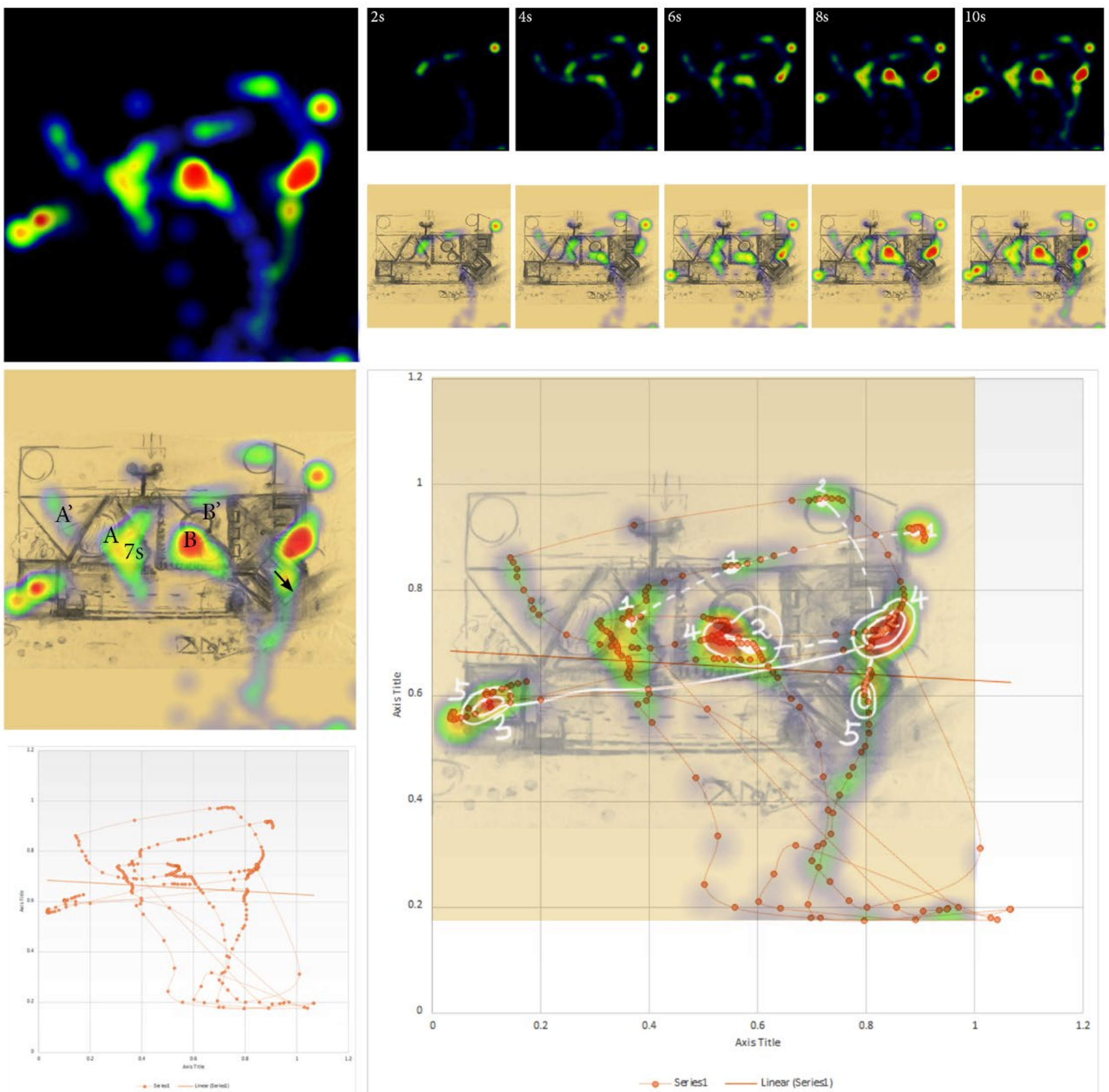
p01	https://www.philadelphiabuildings.org/pab/app/im_display.cfm?RecordId=C3DFC2E4-E427-435D-9BEF01334451EB7A&ArchitectId=&ProjectId=&Authorize=&CFID=58780236&CFTOKEN=92fab03503112d51-0F3FF387-155D-0A04-06ECBEA2B1F45139
p02	https://www.philadelphiabuildings.org/pab/app/im_display.cfm?ImageId=B3187672-C153-4211-974DFF194D0C32E2
p03	https://www.philadelphiabuildings.org/pab/app/im_display.cfm?ImageId=2391FD8E-4570-4E52-A8A7A3CDA53BB954
p04	https://www.philadelphiabuildings.org/pab/app/im_display.cfm?ImageId=7A2CF06A-B31C-4B51-89D70D50E042ACF4
p05	https://www.philadelphiabuildings.org/pab/app/im_display.cfm?ImageId=0CFA7081-01BA-4659-A57CACA36AA06667
p06	https://www.philadelphiabuildings.org/pab/app/im_display.cfm?ImageId=DE940BB7-B0C2-4DFF-872EDD81B17758F7
p07	https://www.philadelphiabuildings.org/pab/app/im_display.cfm?ImageId=3FC00C17-5C35-40EA-A536FADD5B76B197
p08	https://www.philadelphiabuildings.org/pab/app/im_display.cfm?ImageId=C629C168-3B40-4155-B6A8B3CACBA40535
p09	https://www.philadelphiabuildings.org/pab/app/im_display.cfm?ImageId=228F190F-73FC-4487-826EA5894C4E3A4B
p10	https://www.philadelphiabuildings.org/pab/app/im_display.cfm?ImageId=D25E4F4A-9714-4488-9F3FA8C334CE00E3
p11	https://www.philadelphiabuildings.org/pab/app/im_display.cfm?ImageId=707D6300-34B9-487F-B9EDBA4DB4CE6ED4
p12	https://www.philadelphiabuildings.org/pab/app/im_display.cfm?ImageId=55EE1145-D511-4AE2-9EF734E7036B7705
p13	https://www.philadelphiabuildings.org/pab/app/im_display.cfm?ImageId=15A9F765-BFAC-4BEE-8BD4A53F656DE5EA
p14	https://www.philadelphiabuildings.org/pab/app/im_display.cfm?ImageId=BDAF9467-5621-4B31-ACCED330ACF696A3
p15	https://www.philadelphiabuildings.org/pab/app/im_display.cfm?ImageId=3A8A2A3D-8117-4012-BA13E40420CE674F
pl1	https://architexturez.net/doc/az-cf-166214
pl2	Louis I. Kahn. Indian Institute of Management, Ahmedabad, India, Plan sketch of classroom building. 1963 | MoMA
pl3	https://www.philadelphiabuildings.org/pab/app/im_display.cfm?ImageId=88395823-D099-42CB-9DA8E4E037FB19EF
pl4	https://www.philadelphiabuildings.org/pab/app/im_display.cfm?ImageId=64E874DA-EA0E-41F5-B1D2986ABD0A7664
pl5	https://www.philadelphiabuildings.org/pab/app/im_display.cfm?ImageId=DF8A6B67-BAD2-485D-8CA82EA1032268CA
pl6	https://www.philadelphiabuildings.org/pab/app/im_display.cfm?ImageId=DF8A6B67-BAD2-485D-8CA82EA1032268CA
pl7	https://www.philadelphiabuildings.org/pab/app/im_display.cfm?ImageId=2535A560-EEB2-4AEF-8F2EB8BB4A4533A2
pl8	https://www.philadelphiabuildings.org/pab/app/im_display.cfm?ImageId=138AE6EB-6805-403B-9E0A7D35FDFAE70F
pl9	https://www.philadelphiabuildings.org/pab/app/im_display.cfm?ImageId=3B032FF8-3B30-46D2-BBE7FEA15C926CD4

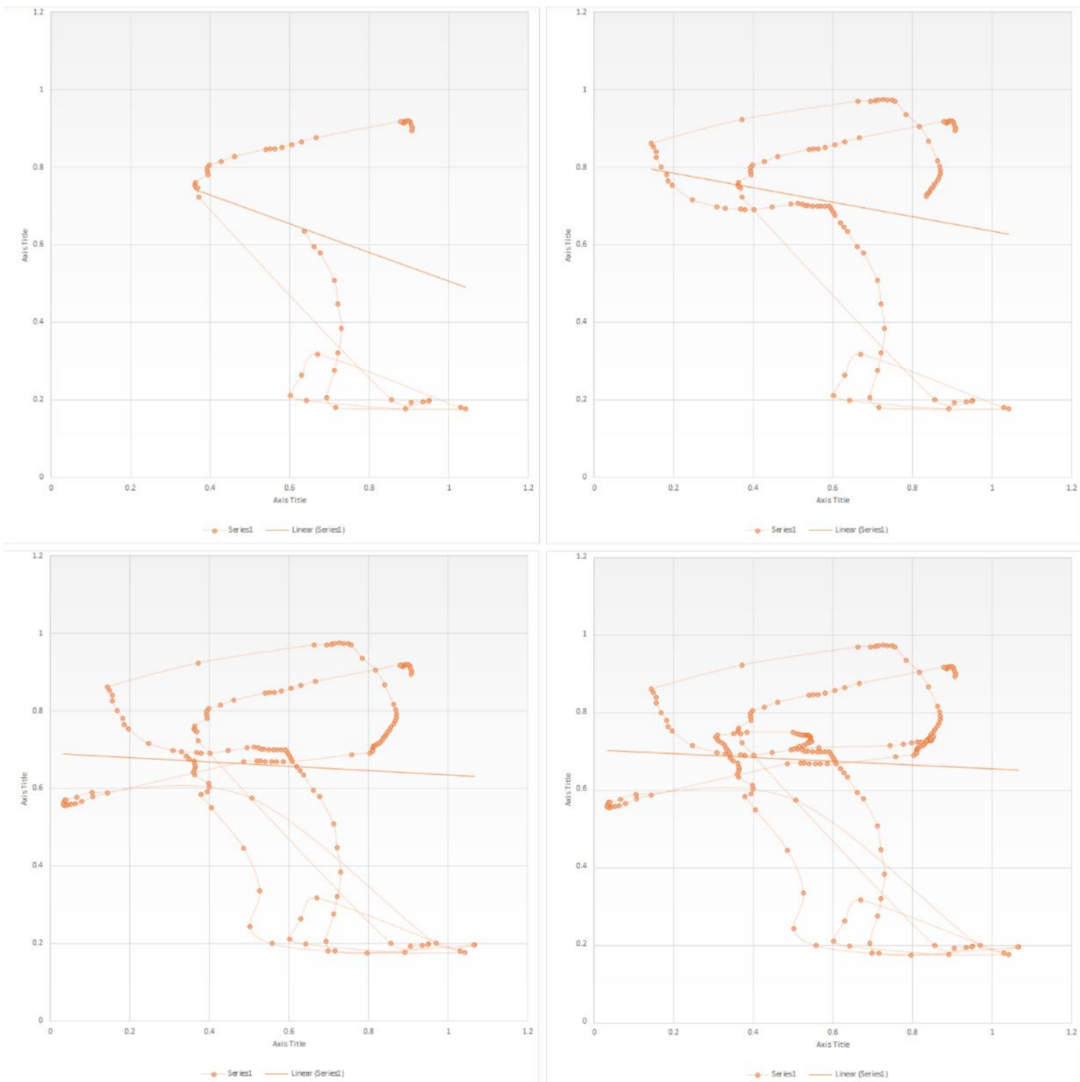
pl10	https://www.philadelphiabuildings.org/pab/app/im_display.cfm?ImageId=021BD5E9-29D6-45B9-962F06F0E312AE36

Appendix II: Heat Maps Examples Used for Average Time Calculation in Chapter 3.3.1

Selected Examples from Eye-Tracking Tests in 3.3.1

Sporadic: Shapes

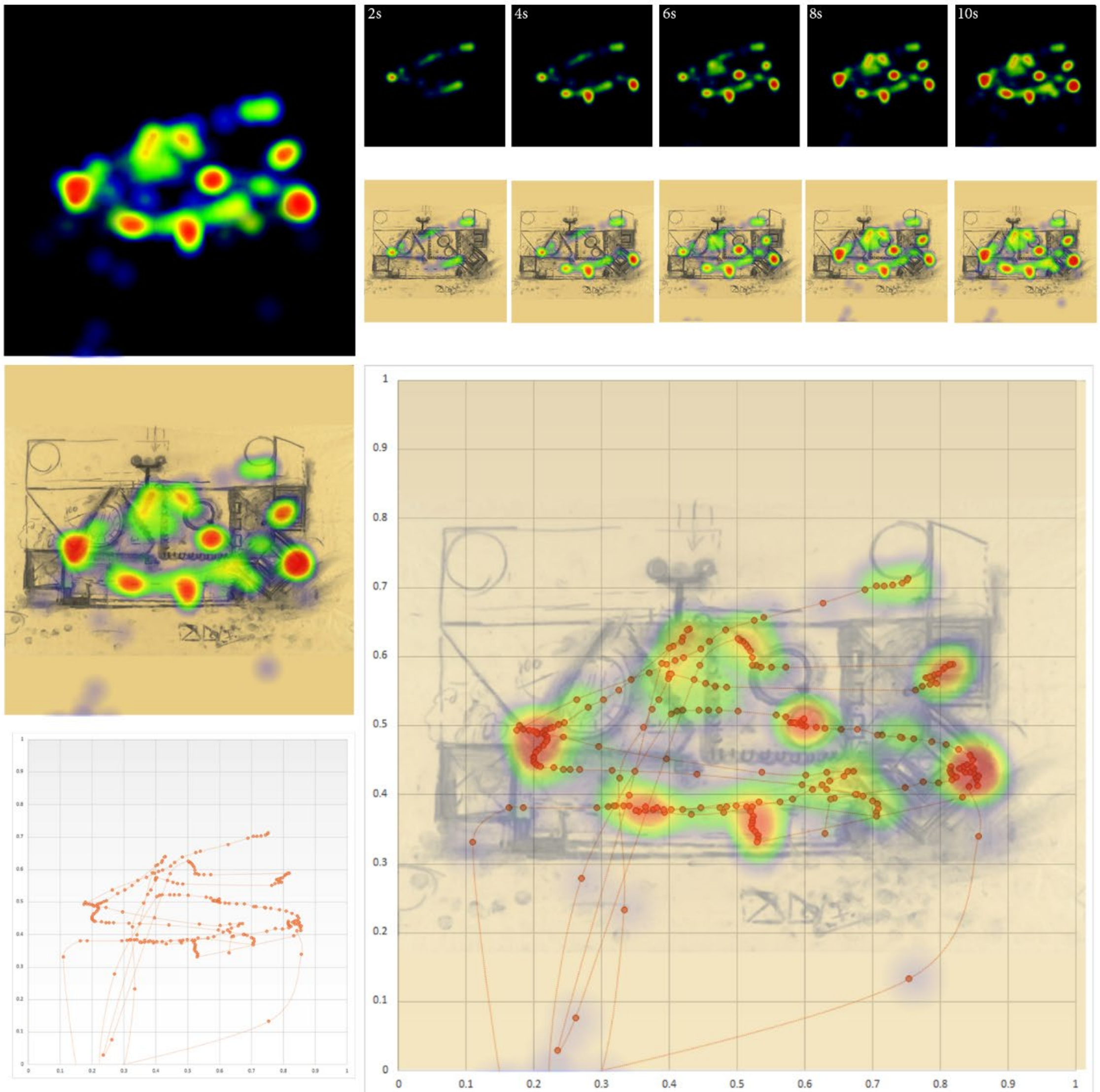


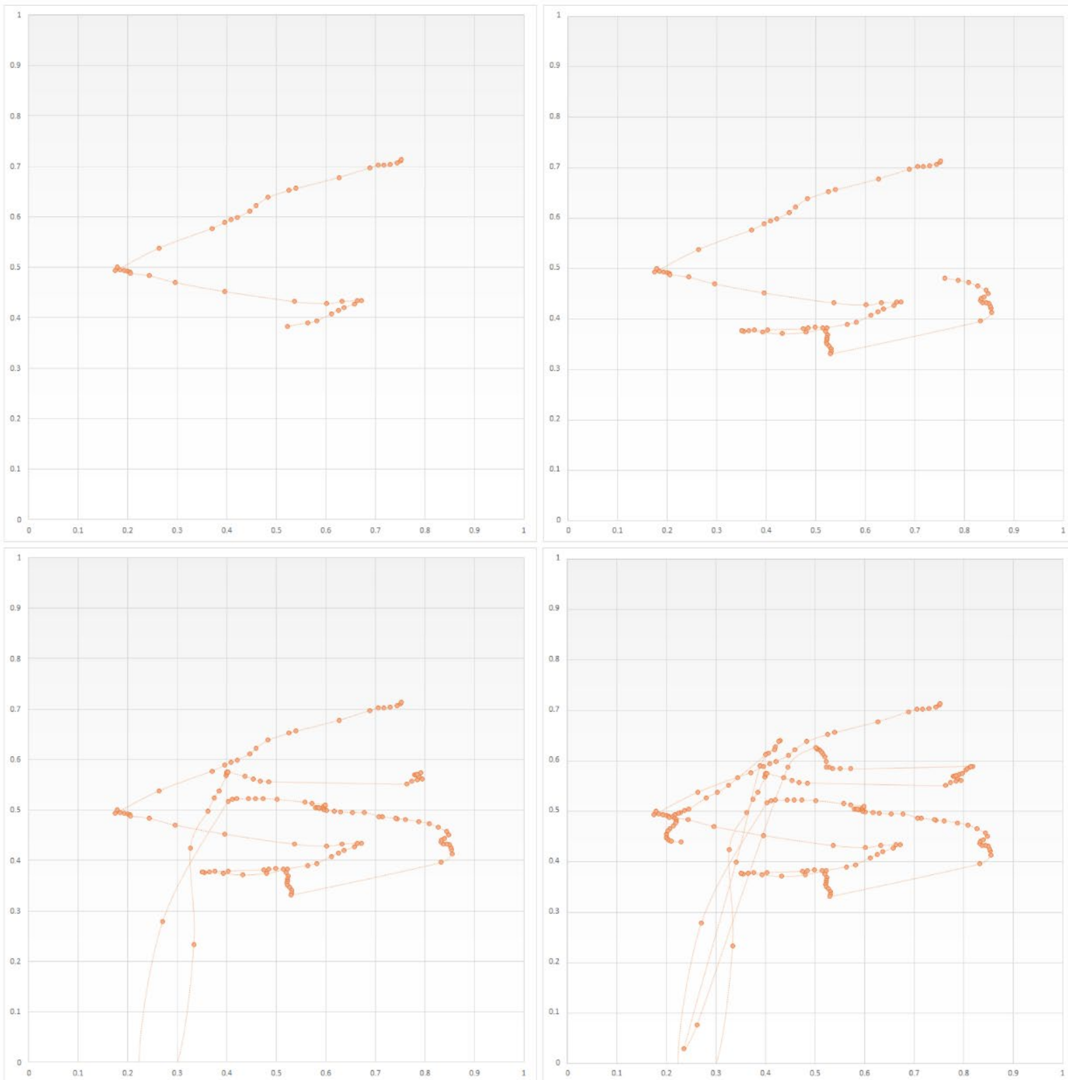


Axis Title
Axis Title
Series1
Linear (Series1)

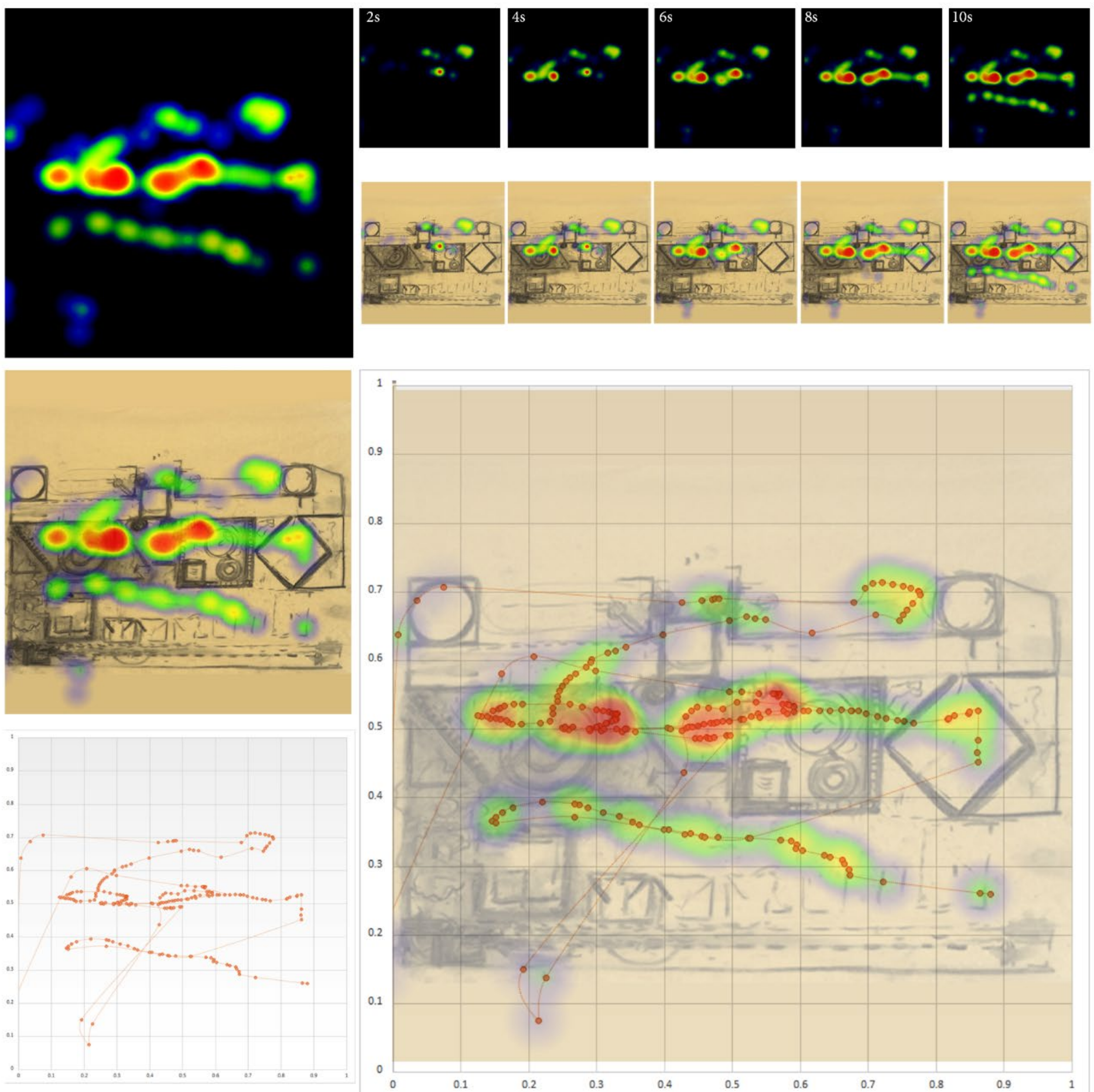
Selected Examples from Eye-Tracking Tests in 3.3.1

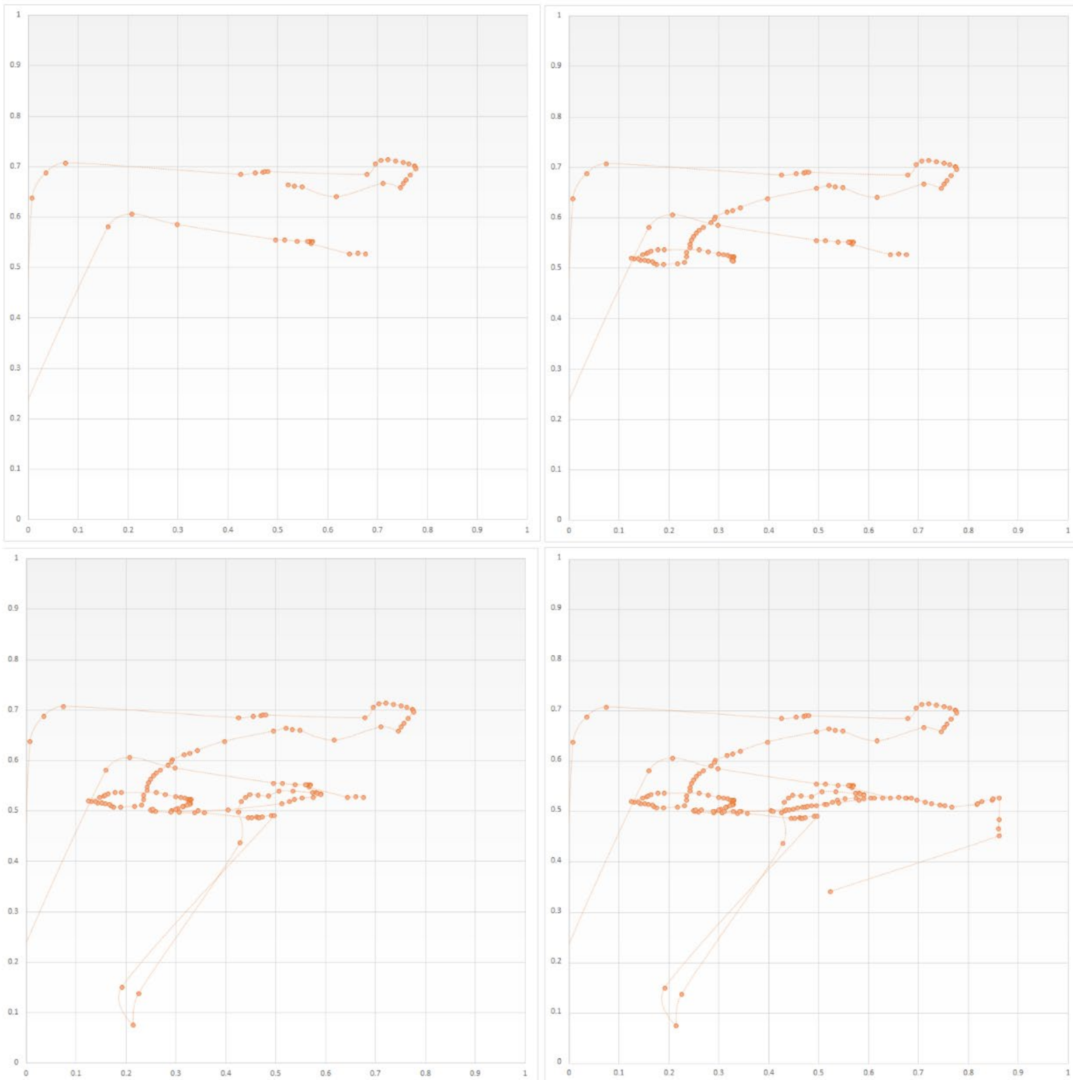
Sporadic: Composition



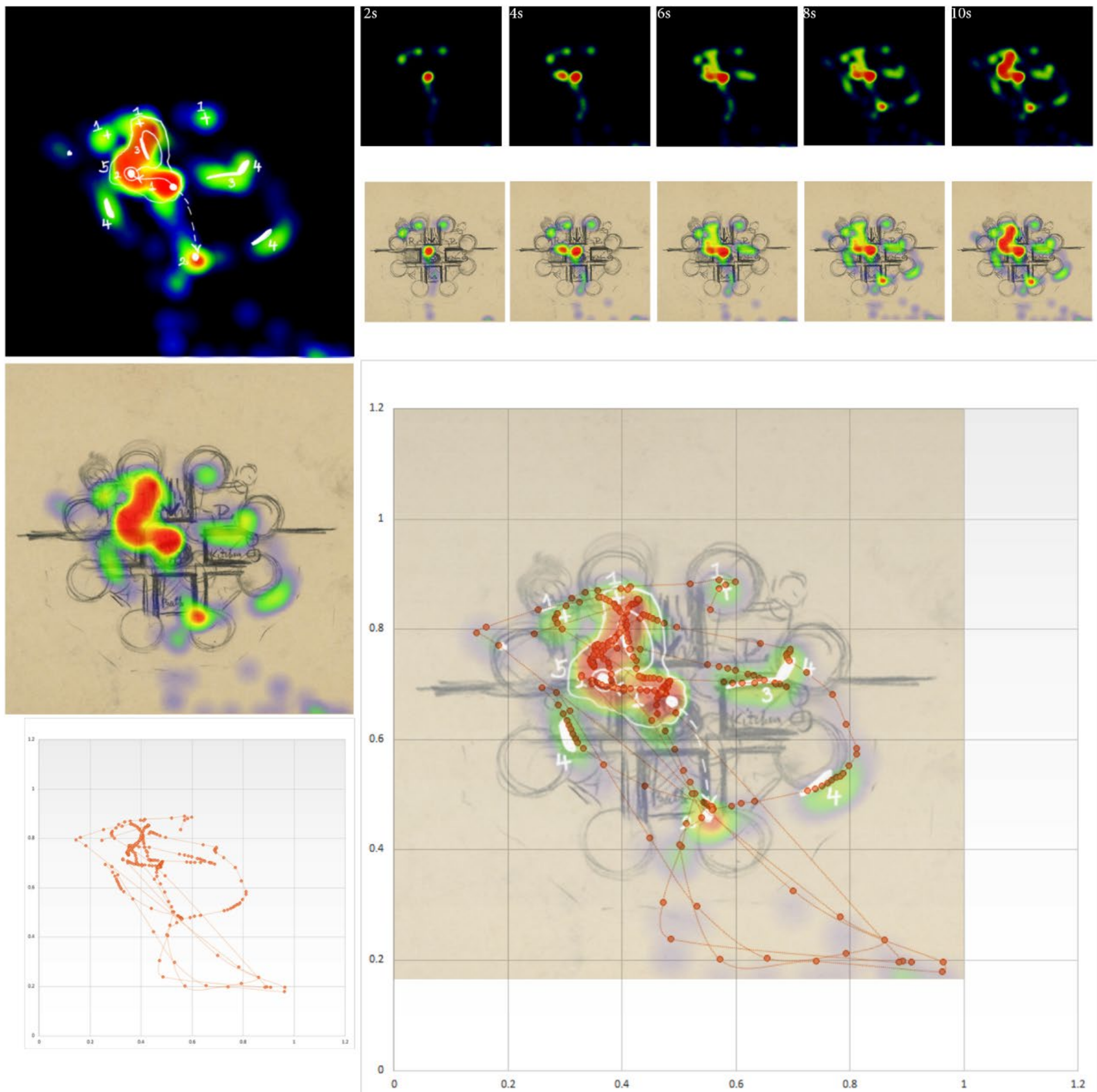

1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

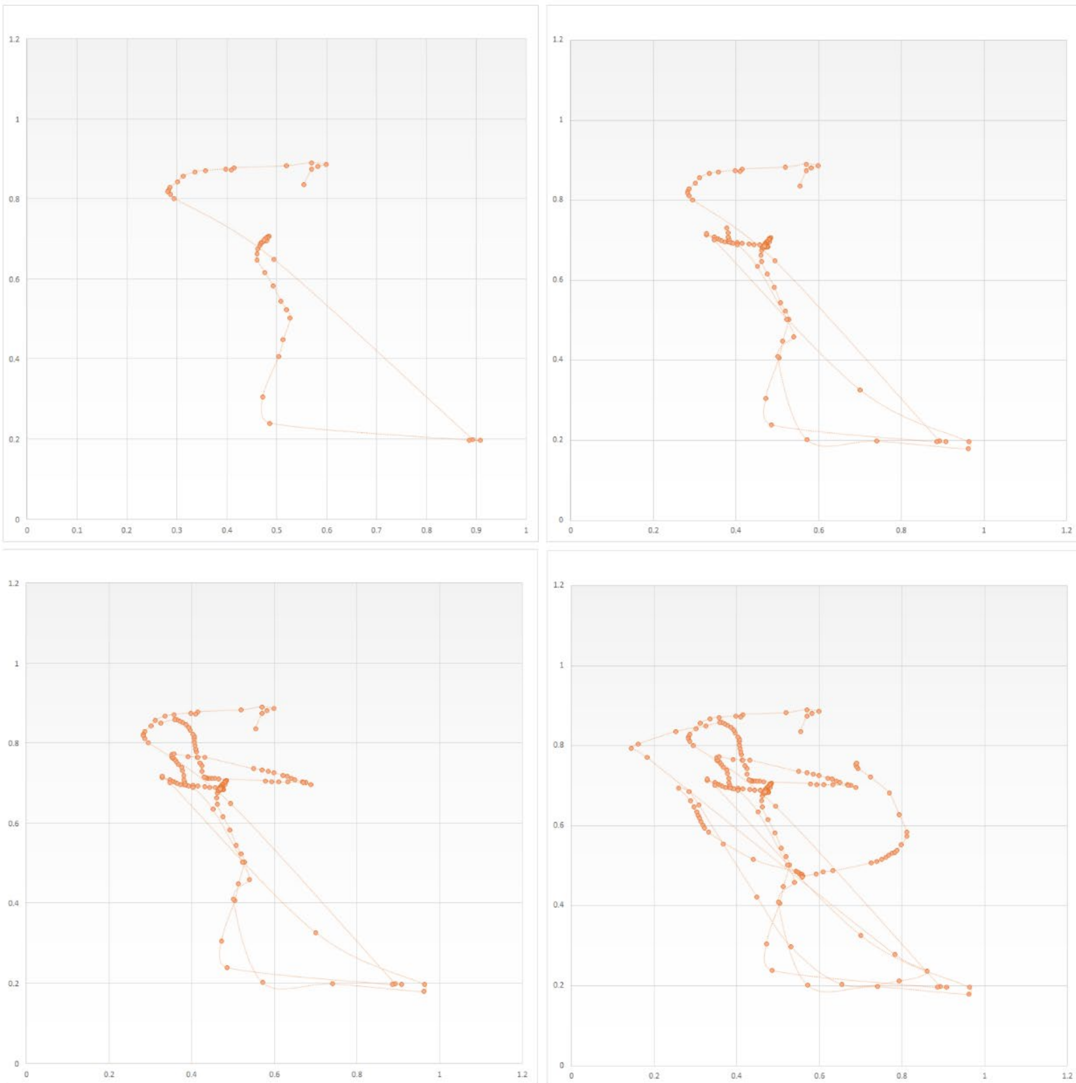
Selected Examples from Eye-Tracking Tests in 3.3.1
Sporadic: Circulation



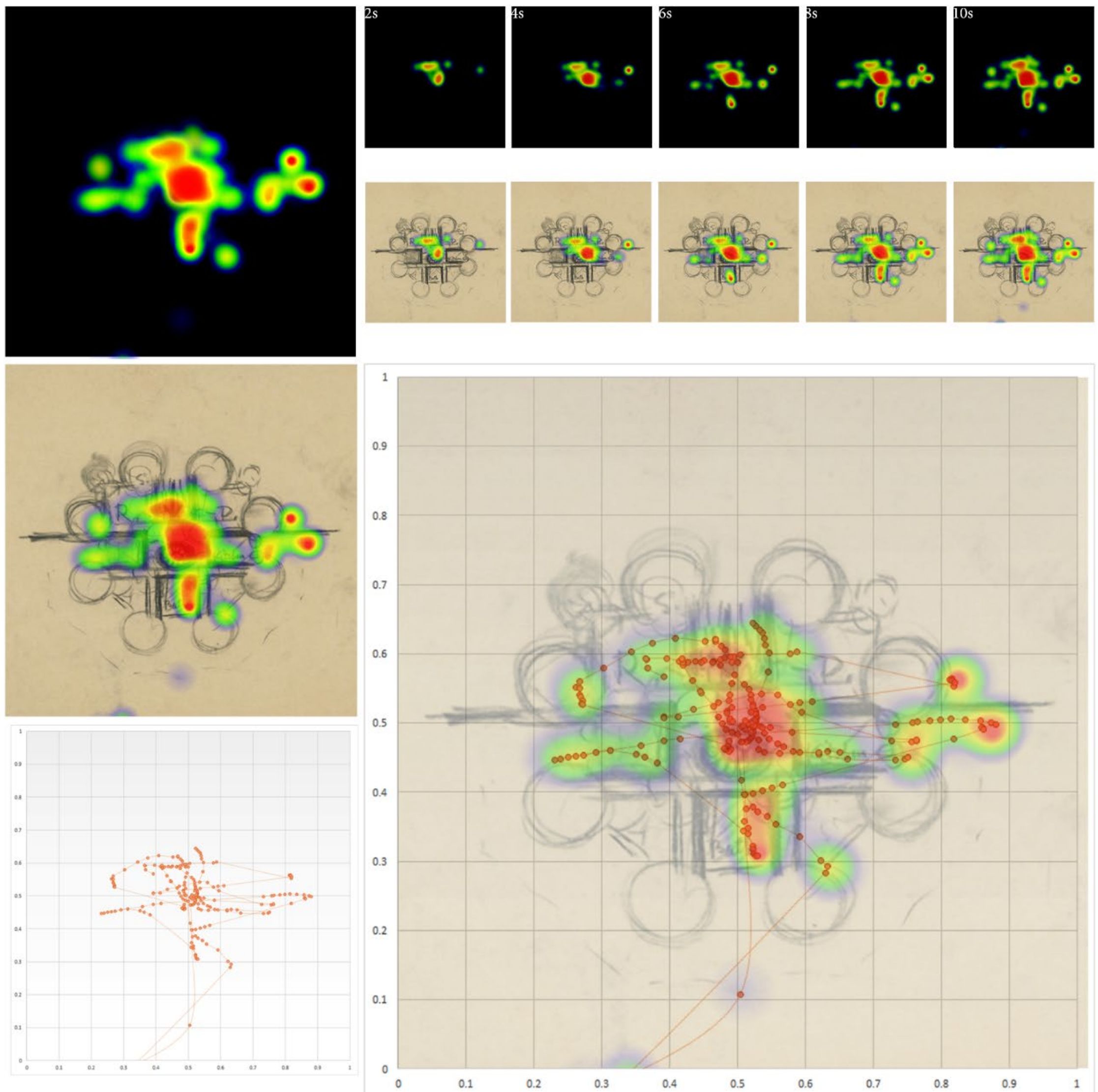

1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

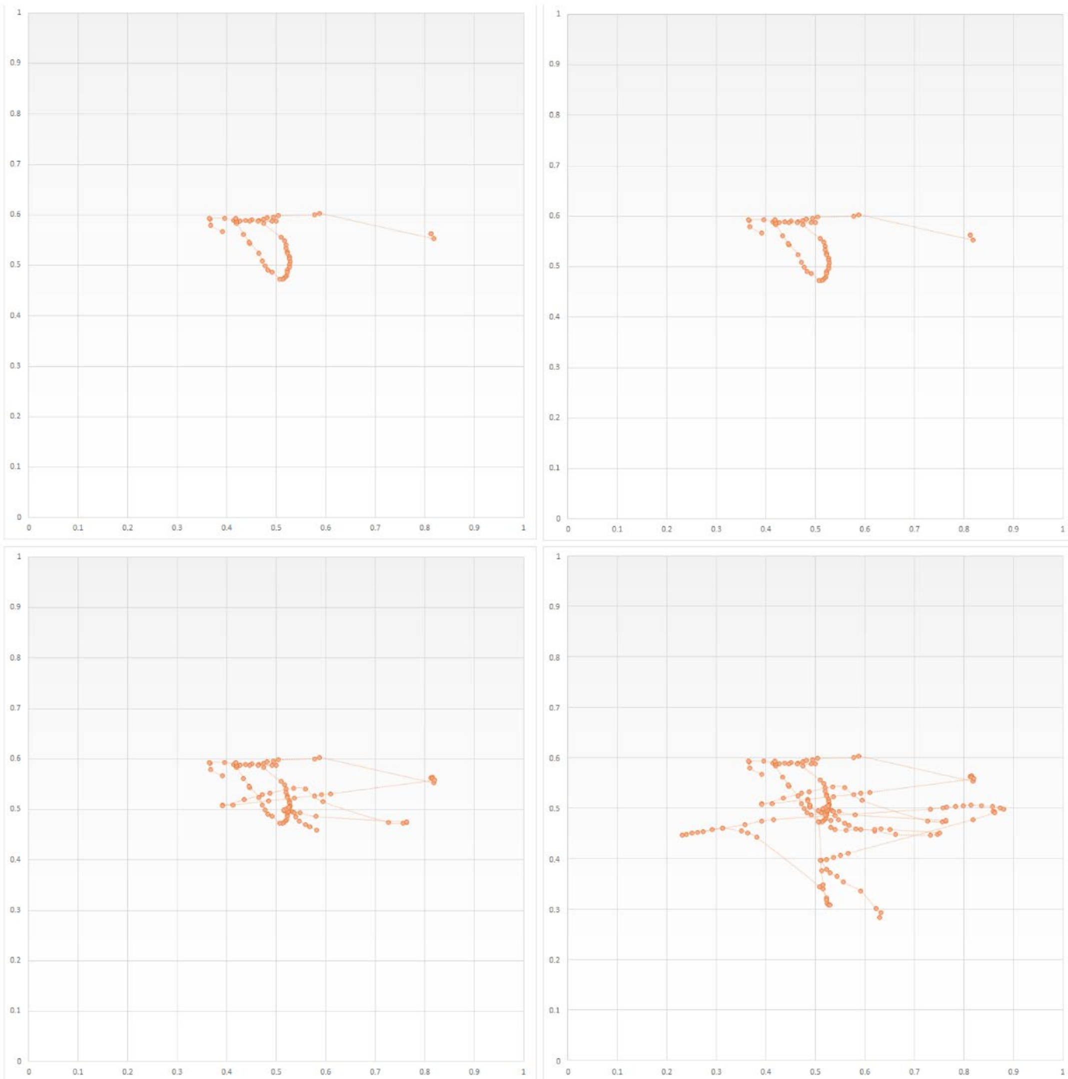
Selected Examples from Eye-Tracking Tests in 3.3.1
Centralized: Shapes





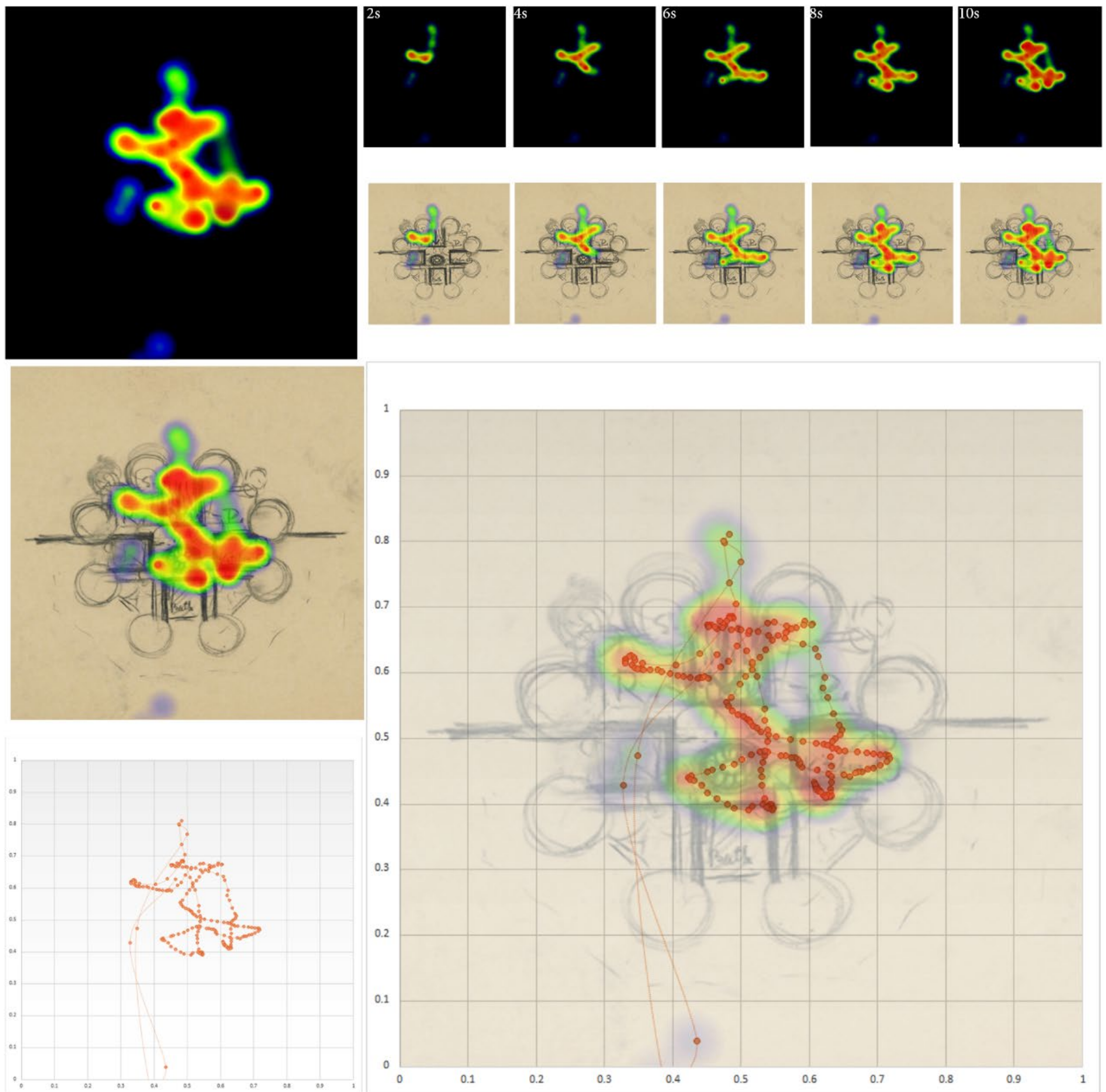
Selected Examples from Eye-Tracking Tests in 3.3.1
Centralized: Composition





1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Selected Examples from Eye-Tracking Tests in 3.3.1
Centralized: Circulation



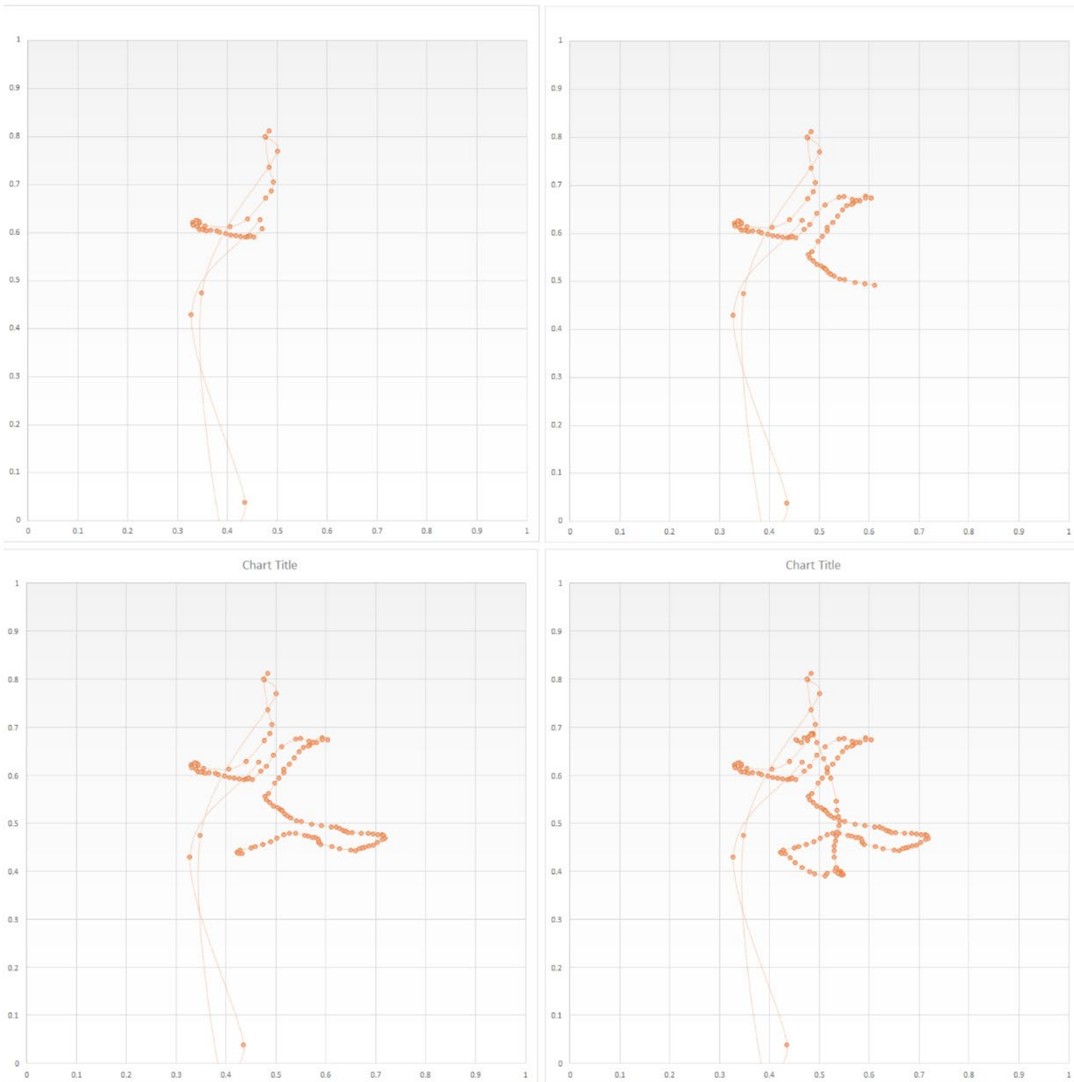
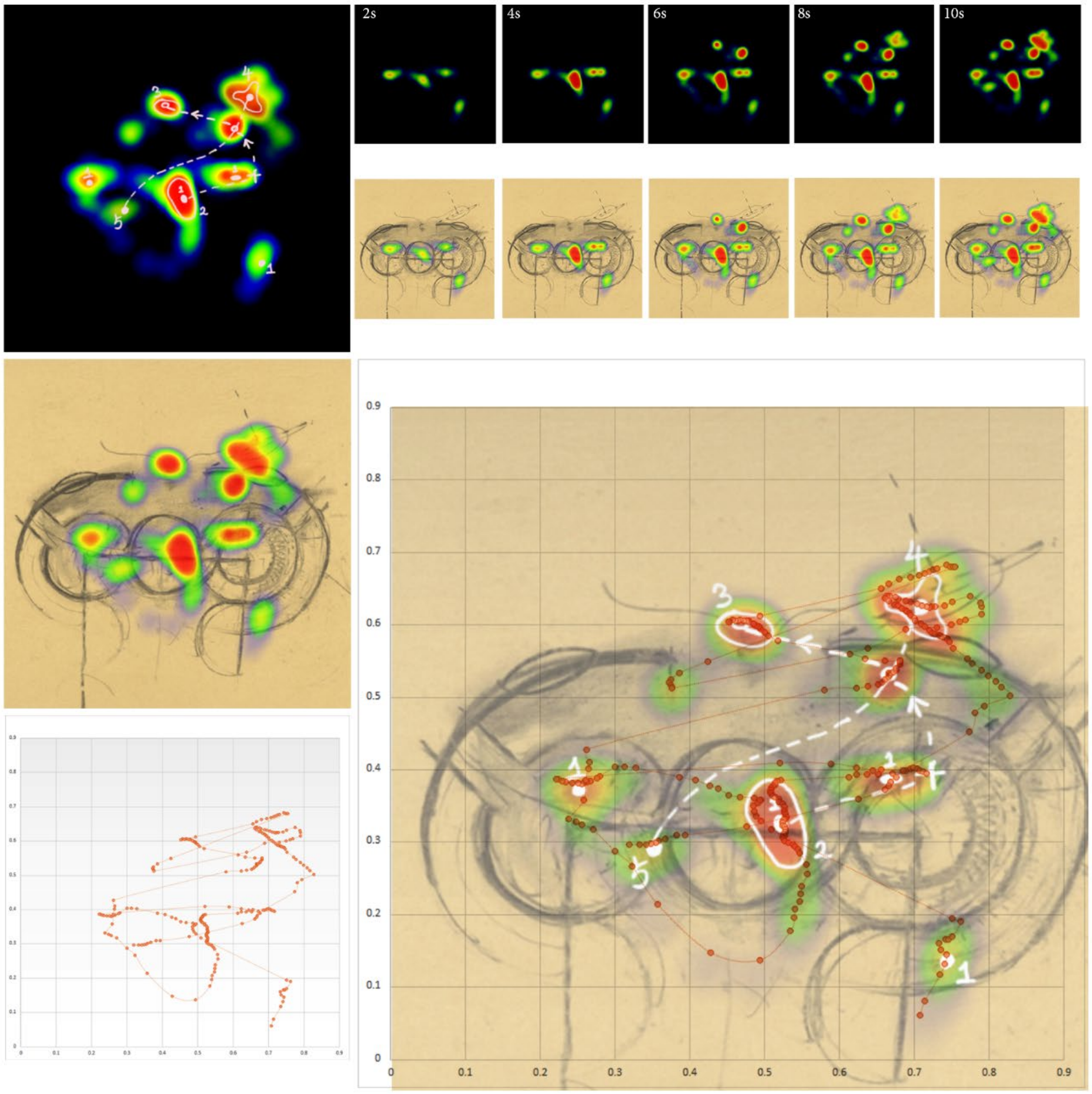


Chart Title
Chart Title

Selected Examples from Eye-Tracking Tests in 3.3.1
Bilateral Symmetry: Shapes



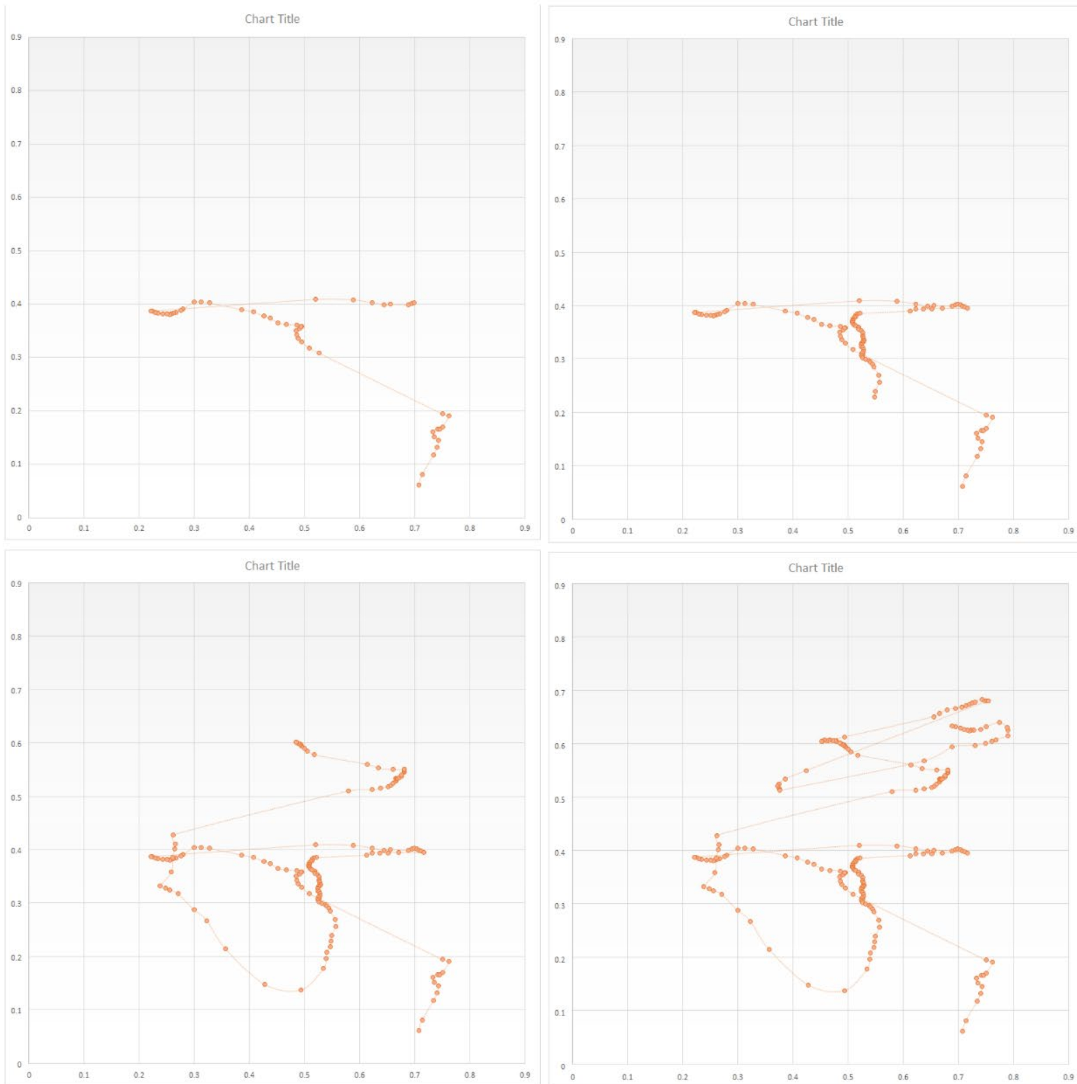
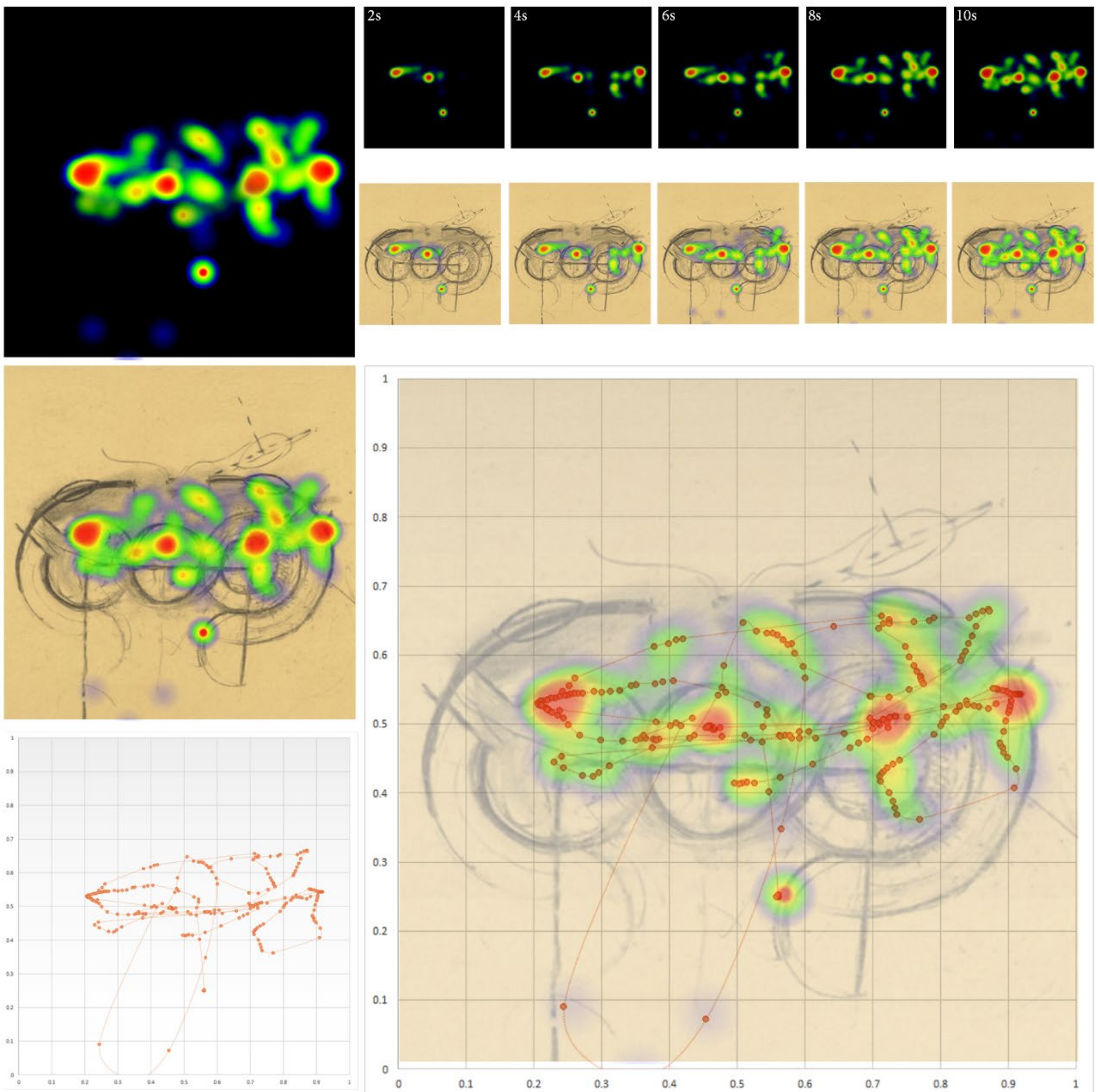
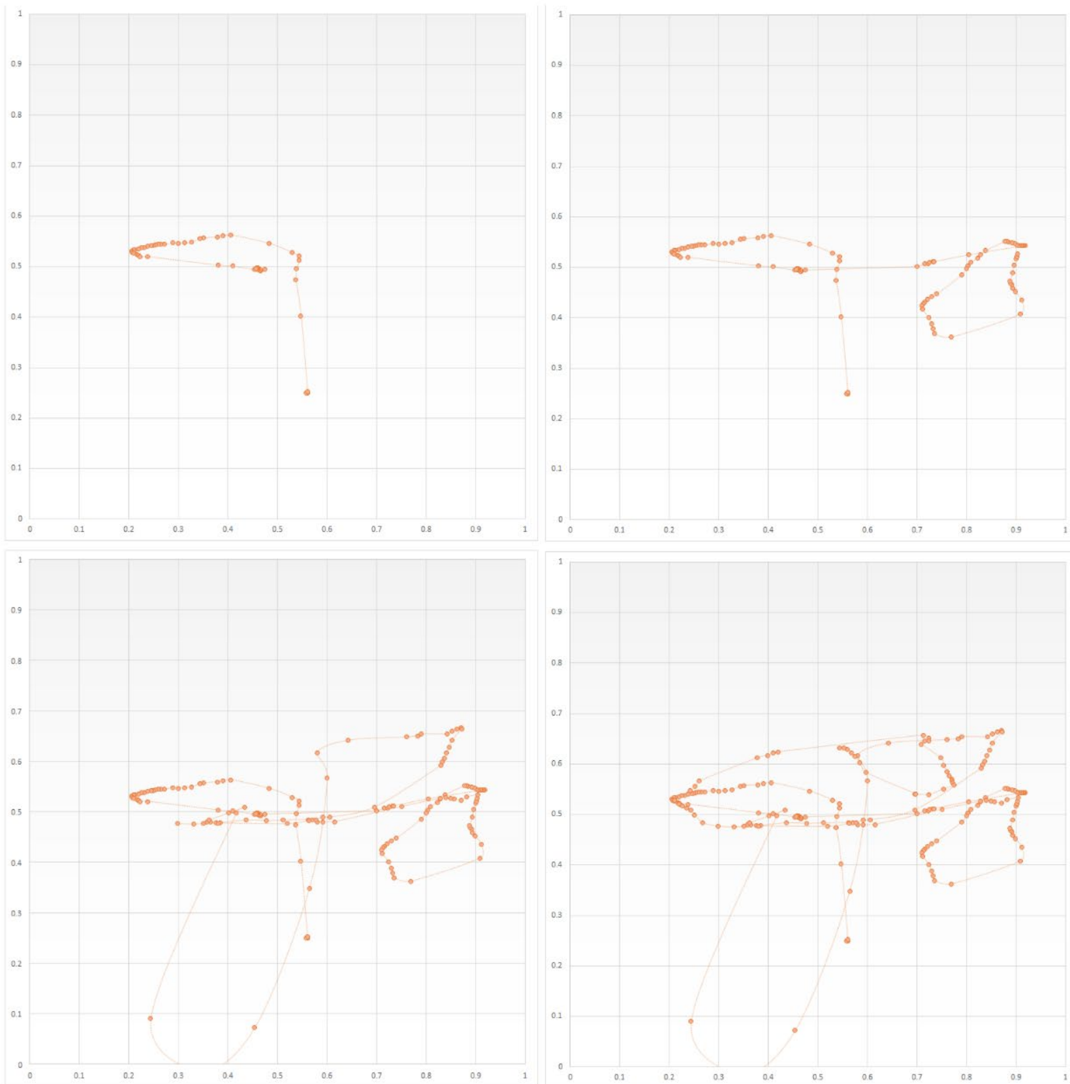


Chart Title
Chart Title
Chart Title
Chart Title

Selected Examples from Eye-Tracking Tests in 3.3.1
Bilateral Symmetry: Composition




0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Selected Examples from Eye-Tracking Tests in 3.3.1
Bilateral Symmetry: Circulation