Patterns of Moments

Reasoning about Space Video via Pattern Language of Human Behavior by Extracting Multi-Action Activities via Machine Learning Video

by

Ngai Hang (Charles) Wu

Bachelor of Architecture
Tsinghua University, 2017

Master of Architectural Design
University College London, 2018

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

MAY 2022

Signature of Author: ____

Department of Architecture
MAY 6, 2022

Certified by: ____

Takehiko Nagakura
Associate Professor of Design and Computation
Thesis Supervisor

Accepted by: ____

Leslie K. Norford
Professor of Building Technology
Chair, Department Committee on Graduate Students

Thesis Supervisor

Takehiko Nagakura, MArch, PhD
Associate Professor of Design and Computation

and readers

Terry Knight, PhD
William and Emma Rogers Professor of Design and Computation

Axel Kilian, PhD
Visiting Assistant Professor of Architecture and Urbanism

Patterns of Moments

Reasoning about Space Video via Pattern Language of Human Behavior by Extracting Multi-Action Activities via Machine Learning Video

by

Ngai Hang (Charles) Wu

Submitted to the Department of Architecture on
May 25, 2022
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Architecture Studies

ABSTRACT

Architecture shapes our perception of space through scale, material, shape, and structure. The design of these elements convinces us of certain behaviors within and around the space, and it plays a significant role in our everyday lives. Experience oriented spatial design provides support for sustainable development, and improves people's material and physical satisfaction, well-being, and overall quality of life. It is contemptuous to think of architecture as a mere visual subject, but rather a medium where purposeful design of stimulus can be set up to lead to specific social behaviors in humans.

This thesis investigates the relationship between the built environment and human behavior through a data-driven method using on-location videos and machine learning. It is intended to provide a crucial means to understand the future opportunities that lie within responsive architecture and human-centered design. Human-centered design is conventionally a top-down approach that is highly dependent on architects' subjective pedagogy and experience of a specific space and their dwellers' and passengers' immediate needs. For example, Christopher Alexander published a collection of design patterns that promotes everyday users to become consciously aware of their living patterns around specific architectural setups. However, his prescriptive proposal outlines only his empirical insight, without further exploration into the dimension of culture, community, and time. The ability to understand human activities more thoroughly in space is lacking.

The research method is to observe and quantify human events and the types of spaces accommodating them and compare the behavioral difference within various spatial settings through short video clips. Initially, field data is collected by observing and recording human behaviors in public. Data-driven Computer Vision techniques are adopted, such as event recognition, scene attribute extraction, and dynamic analysis. Low-level features of human actions such as typing, drinking, stirring, and chewing are recognized, as well as the features of the surrounding space such as greenery, traffic, and enclosure. These low-level understandings discover behavioral patterns in different spaces with various features, providing insights into high-level human-centered spatial design.

After tests and analysis of a case conducted on street café designs, certain correlations between the properties of built environments and user behaviors were discovered. This case study demonstrated the adequacy of the proposed methodology to understand human behavior in space with the help of data-driven machine learning models. It can potentially be used to build a computational human-centered design system that designs by experience. For instance, such a system can help refitting a residential space to better-fit home office for work during pandemic situations.

Thesis Supervisor: Takehiko Nagakura
Title: Associate Professor of Design and Computation

ACKNOWLEDGMENTS

I would like to express my deep and sincere gratitude to Prof. Takehiko Nagakura, Prof. Axel Kilian, and Prof. Terry Knight for all the support and advice for my research. Thank you for your invaluable guidance throughout the challenging two years.

Apart from the committee members, I would also like to thank Dr. Daniel Tsai and Prof. Guzden Varinlioglu for their feedback and support. And also the masters and PhDs of the Computation & Design group for help and comments.

I am extremely grateful to my parents for their love, caring, and sacrifices in educating and preparing me for my future.

More than anybody else, thank you, Lucy.

This thesis was finished during the COVID-19 pandemic. All the motivations and inspirations of this thesis were to push the limit of technology to create changes in the world and the better place it will become. To those who lost their loved ones during this difficult period of time. To my grandmother RuiQing.

CONTENT

1 | INTRODUCTION

Architecture shapes our perception of space through scale, material, shape, and structure. The design of these elements convinces us of certain behaviors within and around the space, and it plays a significant role in our everyday lives. Experience oriented spatial design provides support for sustainable development, and improves people's material and physical satisfaction, well-being, and overall quality of life. It is contemptuous to think of architecture as a mere visual subject, but rather a medium where purposeful design of stimulus can be set up to lead to specific social behaviors in humans.

This thesis investigates the relationship between the built environment and human behavior through a data-driven method using on-location videos and machine learning. It is intended to provide a crucial means to understand the future opportunities that lie within responsive architecture and human-centered design. Human-centered design is conventionally a top-down approach that is highly dependent on architects' subjective pedagogy and experience of a specific space and their dwellers' and passengers' immediate needs. For example, Christopher Alexander published a collection of design patterns that promotes everyday users to become consciously aware of their living patterns around specific architectural setups. However, his prescriptive proposal outlines only his empirical insight, without further exploration into the dimension of culture, community, and time. The ability to understand human activities more thoroughly in space is lacking. A research question is asked, "Can machines provide a relatively bottom-up approach to understanding the relationship between the built environments and their users?"

Perhaps designers and researchers of complex problem spaces like this should not attempt to design the holistic system itself, but rather approach the problem from a collective perspective - the resultant pattern should be logically concluded from a series of existing interrelated systems.

For the human-centered design problem space that involves multitudes of interactions between different mediums, designers should not place themselves at the center of the system, but see themselves as "participants" in recognizing ways that can contribute to it.

The research method is to observe and quantify human events and the types of spaces accommodating them and compare the behavioral difference within various spatial settings through short video clips. Initially, field data is collected by observing and recording human behaviors in public. Data-driven Computer Vision techniques are adopted, such as event recognition, scene attribute extraction, and dynamic analysis. Low-level features of human actions such as typing, drinking, stirring, and chewing are recognized, as well as the features of the surrounding space such as greenery, traffic, and enclosure. These low-level understandings discover behavioral patterns in different spaces with various features, providing insights into high-level human-centered spatial design.

Figure 1. The relationship between the built environments and their users

After tests and analysis of a case conducted on street café designs, certain correlations between the properties of built environments and user behaviors were discovered. This case study demonstrated the adequacy of the proposed methodology to understand human behavior in space with the help of data-driven machine learning models. It can potentially be used to build a computational human-centered design system that designs by experience. For instance, such a system can help refitting a residential space to better-fit home office for work during pandemic situations.

2 | BACKGROUND

2.1 Space and User

The term "user" was invented as part of the design vocabulary around the 1950s. The origin of the term corresponds to "consumer" during the introduction of welfare state programs in Western Europe at the time, referring to a person who has "consumptive behavior" in a buyer-seller relationship.[1] The more contemporary word "users", which denotes both "consumptive behavior" and "usage", expands the timeline to include the post-consumption process where they would continue to engage with the subject product in a way corresponding to the design of the system.

The meaning of "user" in architecture is often assumed to be the person expected to occupy space. Yet, the term “user”, comparable to “client”, “inhabitants” or “occupants”, is traditionally not someone who would not normally be expected to contribute to the actual design process.[2] The concept of "user" appears after the development of space. However, as more products from other domains begin to find and research users before launching, architecture has begun to consider the needs of prospective users as well.

Since the early 1960s, the term “user” has been debated and explored by researchers and designers. ‘User’ is a recurrent term in Dutch architect Herman Hertzberger’s publications, and he emphasizes often on his way of defining the purpose of architectural design as a way to enable ‘users to become inhabitants’, and to allow for ‘the users the freedom to decide for themselves how they want to use each part, each space’.[3] Measure of the worth of an architectural creation or the success of the architect themself for Hertzberger revolves around

---

[1] Adrian Forty. *Words and buildings: A vocabulary of modern architecture*. Vol. 268. London: Thames & Hudson, 2000.
[2] Adrian Forty. *Words and buildings: A vocabulary of modern architecture*. Vol. 268. London: Thames & Hudson, 2000.
[3] Herman Hertzberger, ‘Flexibility and Polyvalency’, Forum, vol. 16, no. 2, February-March 1962, pp. 238-39; ‘Architecture for People’, A+U, March 1977, pp. 124-46

way spaces are used, the diversity of activities which they foster, and the opportunities the space provides its users for creative reinterpretation.

Henry Swain, an English schools architect, said in 1961: “To evolve techniques to help us to analyze the needs of the users of buildings is the most urgent task of our profession”.[4]

Swain promoted the idea that analyzing user needs would rebirth new architectural solutions diverging from the traditional approach of building blocks.

The use of the term “users” in a historical setting was to sustain architectural practice and design systems during a period of time of massive stylization and competition in the domain. Shortly after the end of the Second World War saw the growth of the welfare state in Western European countries, and of welfarist policies in the USA. Within this political system, designed to stabilize relations between capital and labor but without affecting any major redistribution of the ownership of wealth, architecture was widely adopted by Western governments as an important part of their strategy. Not only was it a matter of providing new schools, housing and hospitals, but of doing so in such a way that those who occupied these buildings would be convinced of their ‘equal social worth’ with all other members of society. For the many architects employed on public-sector projects, it was necessary to convince themselves - and the public at large - that the ‘client’ was not the bureaucracies or elected committees that actually commissioned the buildings, but those who would actually inhabit them. Although these people were almost invariably unknown to the architects, the professional claims of architects to serve the greater good of society depended upon being able to show that the true beneficiaries of the new schools or social housing were indeed those destined to occupy them.

By privileging ‘the user’, it could be claimed the expectations within a welfare state democracy for the disempowered to be treated as citizens of ‘equal social worth’ were being realized.

[4] Henry Swain, ‘Building for People’, *Journal of the Royal Institute of British Architects*, vol.68, Nov. 1961, pp. 508-10

The decline of interest in the 'user' and 'user needs' corresponded to the decline in public-sector commissions in the 1980s. Perhaps another reason for dissatisfaction with the 'user' has been that it is such an unsatisfactory way of characterizing the relationship people have with works of architecture: one would not talk about 'using' a work of sculpture.

The emerging interest in adopting a user-centric approach in architecture has shaped the way architects and engineers approach their work. The term "user" may be understood as a bridging element to the current paradigm of relationships - if a connection was said to exist between the buildings humans live in and their inhabiting humans' social behavior, then it is necessary to have a word to reference the actions of which the buildings should be designed to promote. The "user", therefore, may be seen as a result of the attempt to establish the relationship.

2.2 Human-Centered Design in Space

"Form follows function." Louis Sullivan said, who designed the first steel-frame skyscrapers in the late-19th century, when design was associated with expensive materials and exquisite craftsmanship before Modernism.[5] The Modernists agree with the value of aesthetically appealing, but hold that beauty should come from the pursuit of functionality, rather than simply use decoration or styling.

A function is a reflection of the requirement or purpose of the design which comes from human needs and wants, which can be incredibly complex.

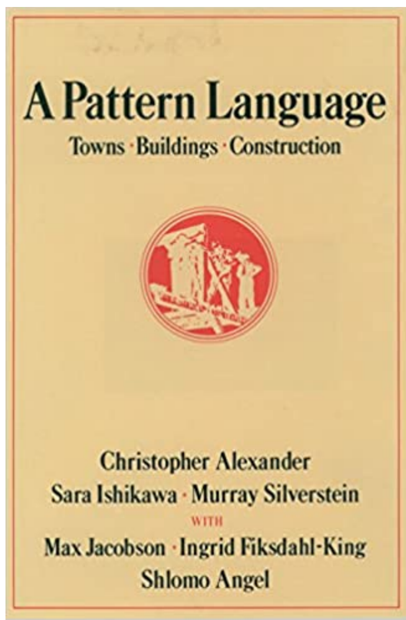


Figure 2. A Pattern Language, by Christopher Alexander (1977)

A Pattern Language, by Christopher Alexander, developed a broad philosophical critique of the modern alienated condition. He argued for an ideal balance between work and family life, suitable public institutions, mixed use in neighborhoods, and rich public spaces for carnivals and other expressions of irrationality. According to Alexander, rich, community-oriented settings go

[5] Louis Sullivan. *The tall office building artistically considered*. 1922.

beyond mere functionality and act as social settings that afford spatial and affective learning.[6] A Pattern Language consists of suggestive diagrams that, in built form, encourage interaction between people and their environments at multiple levels.

He rooted the development of human-centered design. Understanding buildings without paying conscious attention would necessarily involve innate learning and open up fundamental, shared human capacities. Alexander describes this process as a search for the quality of things that is subjective, that cannot be named, and yet has a level of objectivity and precision. He believed that everything has its degree of life and there is a scientific way to measure it. However, Alexander's approach is relatively top-down prescriptive interpretations based on their own empirical experience of lives. It inspired the research question of this thesis - Can machines provide a relatively bottom-up approach to understanding the relationship between the built environments and their users?

---

[6] Christopher Alexander. *A pattern language: towns, buildings, construction*. Oxford university press, 1977.

2.3 Precedent Studies on Analysis of Built Environment and Their Users

Through emerging technologies, the integration between humans and artificial intelligence products begins to expand into the design domains. With the development of building information modeling, data visualization and assistive intelligence technology began to impact the way we perceive and analyze human behavioral patterns within architectural spaces. In order to systematically collect data about the user experience of space, attempts such as path tracking, bioinformatics, sonic analysis, and visual recording, were proposed to have a deeper understanding of how people interact with space via context, behavior, body, and mind.

2.3.1 Path Tracking

Agent-based simulation

There is promising research investigating agent-based simulation in buildings. CityScope Andorra proposed a novel information visualization approach developed and deployed in the state of Andorra.[7] We present a framework to analyze and represent the flow of people through a multi-level interactive and tangible agent-based visualization.

Figure 3. Agent-based modeling for behaviors, building occupancy loads and risk infection at the University of Guadalajara (Larson, 2018)

[7] Kent Larson, et al, *"Travel Demand and Traffic Prediction with Cell Phone Data: Calibration by Mathematical Program with Equilibrium Constraints,"* 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), 2020, pp. 1-8, doi: 10.1109/ITSC45102.2020.9294614.

Exodus suggested a simulation tool that predicted evacuation through available doors and windows.[8] A number of people perish in this simulation due to the rapidly developing fire. These promising simulation approaches are target-specified for system optimization, and individuals are labeled as small dots inside buildings. I wonder if more subtle details about human activities can be captured and analyzed.

Figure 4. Exodus - Fire Escape Simulation (Galea, 2009)

Kenny Cheung's thesis at the Massachusetts Institute of Technology explores the relationship between workplace environment and human behavior as an important concern for architectural designers.[9] The project outlines a pilot study where ubiquitous computing, including a collection of sensors, was utilized to evaluate activity quality in a designed commercial environment. The result was then implemented in a data visualization platform as a means to initiate conversations and discussions revolving around the subject.

---

[8] Owen, Matthew, Edwin R. Galea, and Peter J. Lawrence. “The Exodus Evacuation Model Applied To Building Evacuation Scenarios.” *Journal of Fire Protection Engineering 8*, no. 2 (May 1996): 65–84. https://doi.org/10.1177/104239159600800202.

[9] Cheung, Kenneth Chun-Wai. *“Understanding behavior with ubiquitous computing for architectural design”*. MIT, 2007.

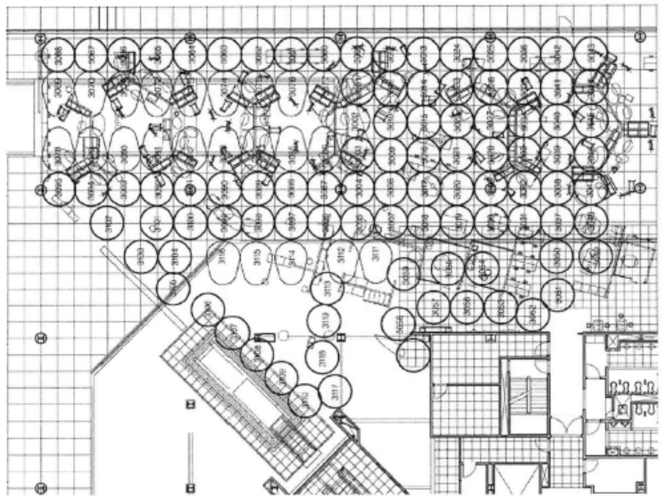


Figure 5. Understanding Behavior with Ubiquitous Computing for Architectural Design, Kenneth Cheung (2007)

2.3.2 Bioinformatics

Research engineers at Chiba University proposed a system for approximating individual emotions based on a set of multimodal indoor environment data for human participants, including both indoor environment data obtained from visual and auditory sensors and the emotions data based on pulse and temperatures gathered via direct skin contact sensors.[10] Through the combination of different sensors, the researchers proposed and built a adoptive human emotion estimation model based on the analyzed data that can predict the emotion of humans within a contained space up to 80% accuracy. The ability to make predictions allows us to envision the future where human emotions and intelligence can be integrated into architectural design. People's need in both the public, communal domain as well as in the personal domain are important, and it is particularly important to engage with the psychological needs of people's behavior when considering design attributes to a space.

[10] Komuro, N., Hashiguchi, T., Hirai, K. *et al.* Predicting individual emotion from perception-based non-contact sensor big data. *Sci Rep* 11, 2317 (2021). https://doi.org/10.1038/s41598-021-81958-2

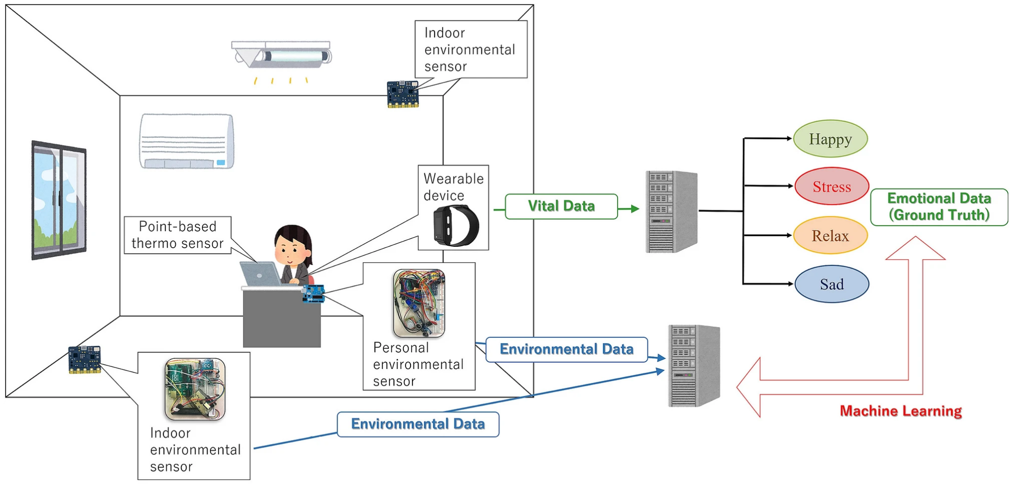


Figure 6. Predicting individual emotion from perception-based non-contact sensor big data (2021)

2.3.3 Sonic Analysis

The Birth of a Word is a project by MIT researcher Deb Roy, who wanted to visualize the way speech intersects space from a developmental perspective.[11] He recorded a continuous dataset of his infant son's journey from blurred vocal sounds to being able to properly pronounce words with multiple syllabo. The data-rich research has great implications and evokes questions around the way spatial cues could be correlated with human development through time. Human language ability has always been the interest of numerous research fields given the challenges we must overcome in order to build artificial intelligence systems that have the same language capability as humans. The mechanisms of language acquisition require a multitude of information input, and Dr. Roy's approach was one that stood out from the way he attempted to find correlation between environmental factors and speech, opening up a new domain of possible explorations regarding how human development could coincide with interaction with spaces.

[11] Roy, Brandon Cain. "The birth of a word." PhD diss., Massachusetts Institute of Technology, 2013.

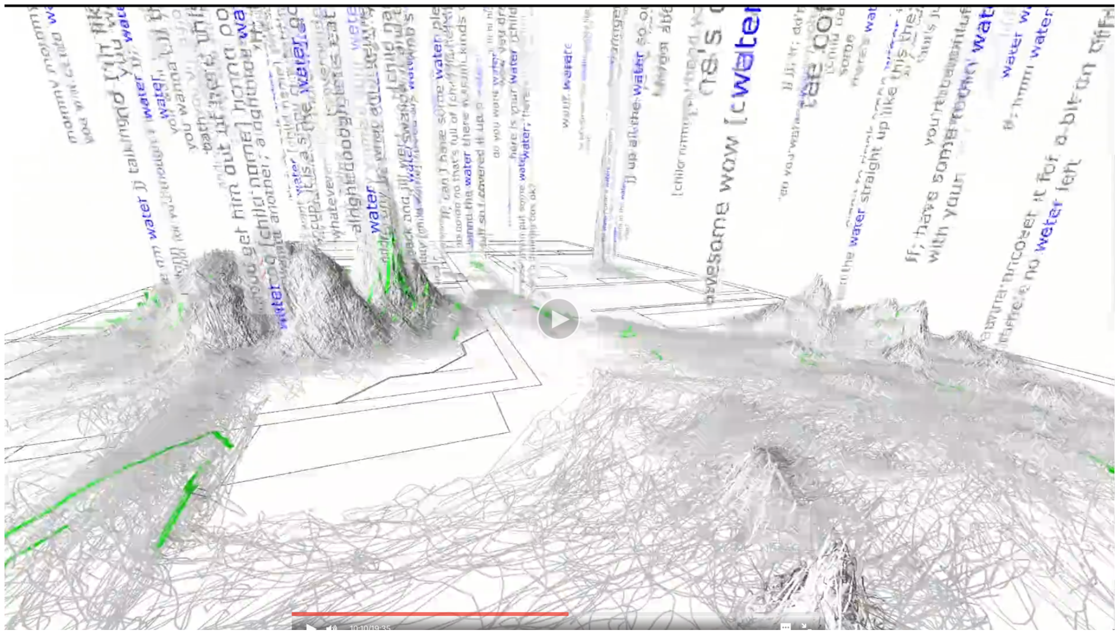


Figure 7. The Birth of a Word spatial visualization by Deb Roy (2011)

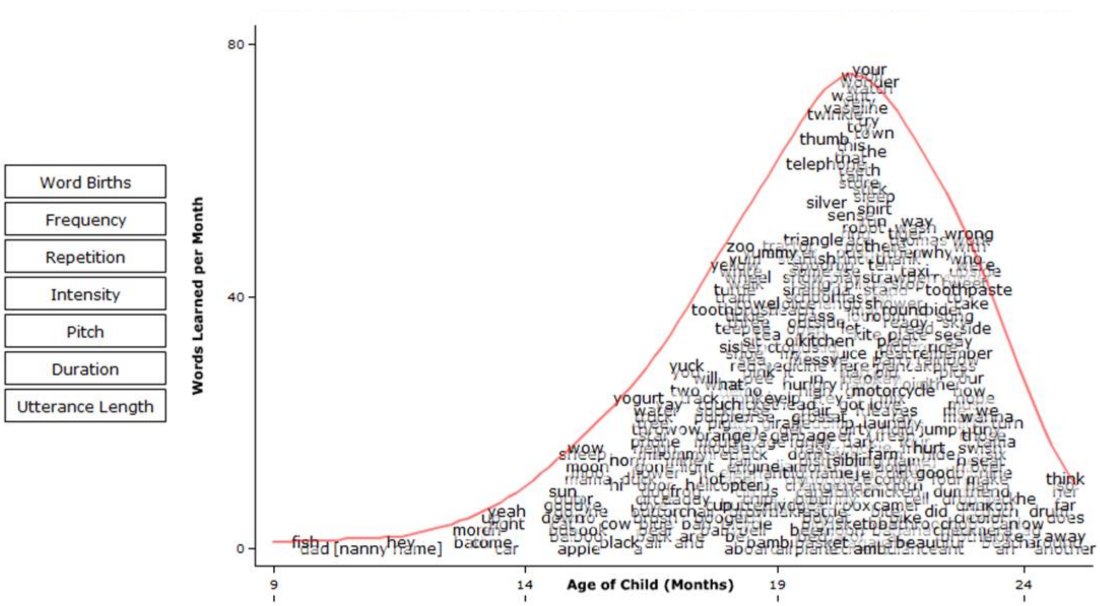


Figure 8. The Birth of a Word visualization by Deb Roy (2011)

2.3.4 Visual Observation

When one considers the adoption of technologies in various industries, the architecture industry may not come to mind, yet that has begun to change, starting with the inclusion of data in the design process. New interest in innovative technology in architecture comes from necessity, and utilization of big data can help industry practitioners become more flexible, efficient and secure in the design build process. From a user-oriented perspective, there are multiple benchmark datasets that focus on analyzing behavior through space. Moments in Time, co-developed by MIT and IBM, is a large-scale dataset for recognizing and understanding motion in videos.[12] The dataset includes a collection of more than one million labeled videos involving people, animals, objects, or natural phenomena, capturing dynamic scenes. It is designed to have a large coverage and diversity of events in both visual and auditory modalities, and can serve as a new challenge to develop models that scale to the level of complexity and abstract reasoning that a human processes on a daily basis. An ever-going part of our consideration in architecture is the "built environment", and it contains a vast digital footprint and visual data if recorded. Effective exploration and analysis of the data requires intelligent tools and supporting benchmarks like Moments in Time, that can help architects, designers and engineers to gain a deeper understanding of how humans and objects interact in distinctive spatial design, and inspire ideas that can be applied to our current practices.

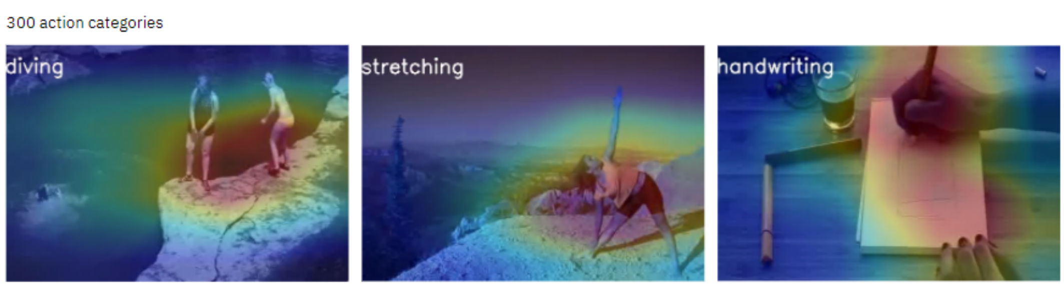


Figure 9. Action Recognition examples in Moments in Time (2017)

[12] Monfort, Mathew, Alex Andonian, Bolei Zhou, Kandan Ramakrishnan, Sarah Adel Bargal, Tom Yan, Lisa Brown et al. "Moments in time dataset: one million videos for event understanding." *IEEE transactions on pattern analysis and machine intelligence* 42, no. 2 (2019): 502-508.

2.4 Computer Vision Techniques

2.4.1 Understanding “Dynamic” in Computer Vision Field

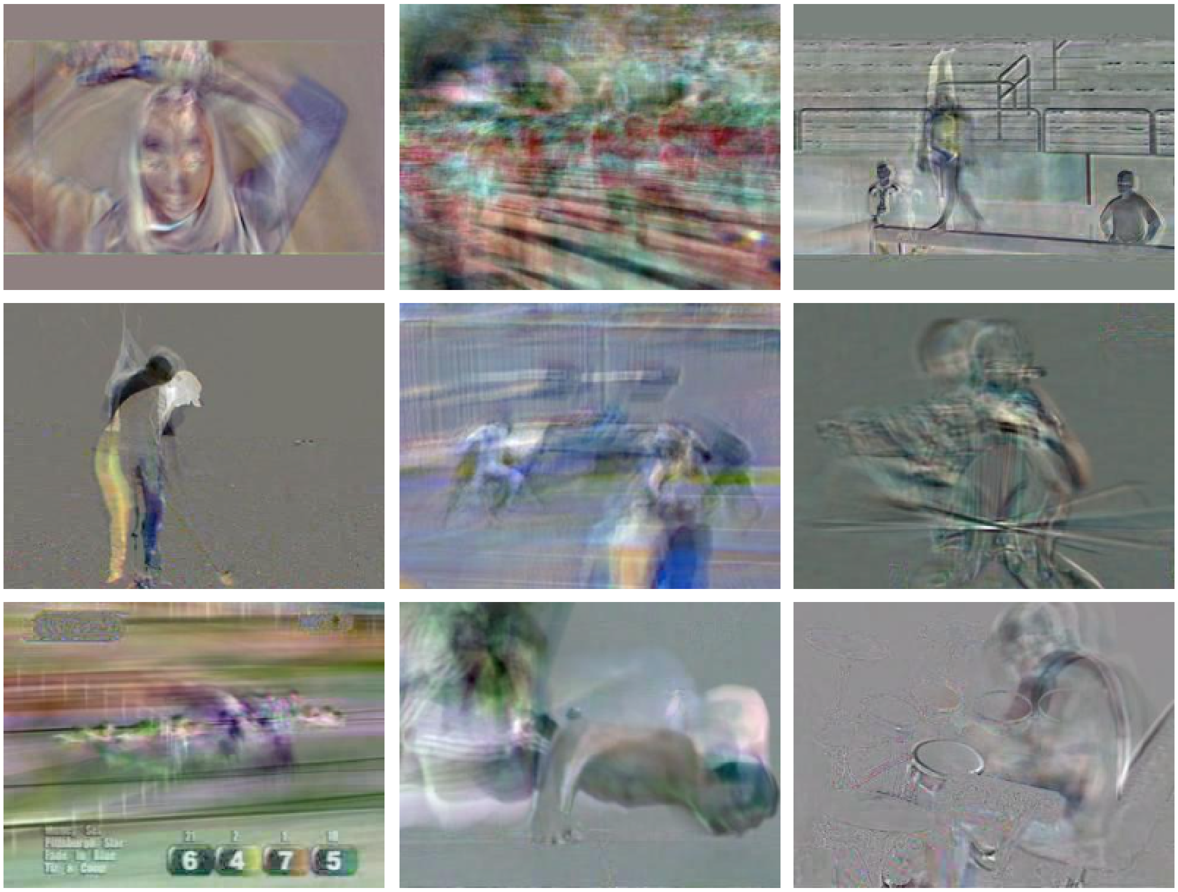


Figure 10. Examples of dynamic images by Hakan Bilen (2016)

The concept of the dynamic image, is a novel compact representation of videos useful for video analysis especially when convolutional neural networks (CNNs) are used.[13] The dynamic image is based on the rank pooling concept and is obtained through the parameters of a ranking machine that encodes the temporal evolution of the frames of the video. Dynamic images are obtained by directly applying rank pooling on the raw image pixels of a video producing a single RGB image per video. This idea is simple but powerful as it enables the use of existing CNN

[13] Bilen, Hakan, Basura Fernando, Efstratios Gavves, Andrea Vedaldi, and Stephen Gould. "Dynamic image networks for action recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3034-3042. 2016.

models directly on video data with fine-tuning. We present an efficient and effective approximate rank pooling operator, speeding it up orders of magnitude compared to rank pooling. Our new approximate rank pooling CNN layer allows us to generalize dynamic images to dynamic feature maps and we demonstrate the power of our new representations on standard benchmarks in action recognition achieving state-of-the-art performance.

Summarizing the video content in a single still image may seem difficult. In particular, it is not clear how image pixels, which already contain appearance information in the video frames, could be overloaded to reflect dynamic information as well, and in particular the long-term dynamics that are important in action recognition.

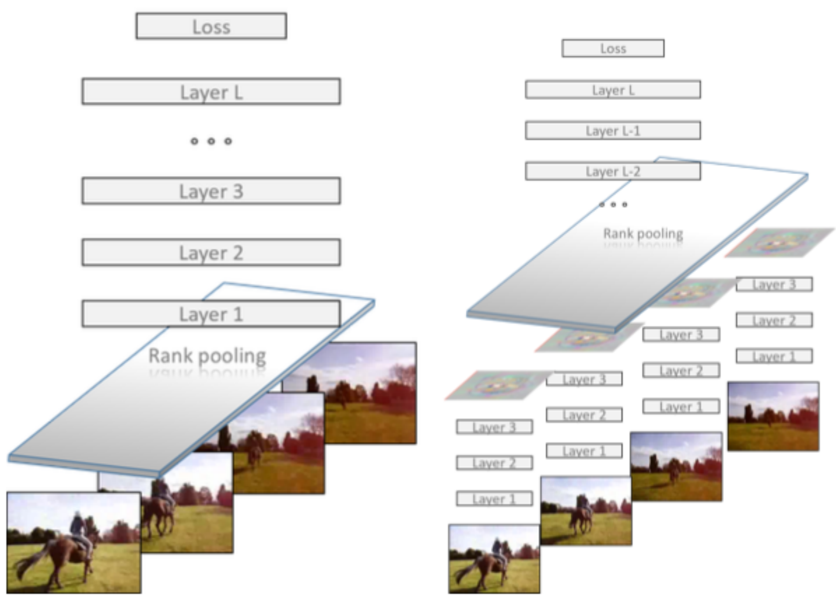


Figure 11. The algorithm of dynamic images by Hakan Bilen (2015)

2.4.2 Object Recognition for Architectural Analysis

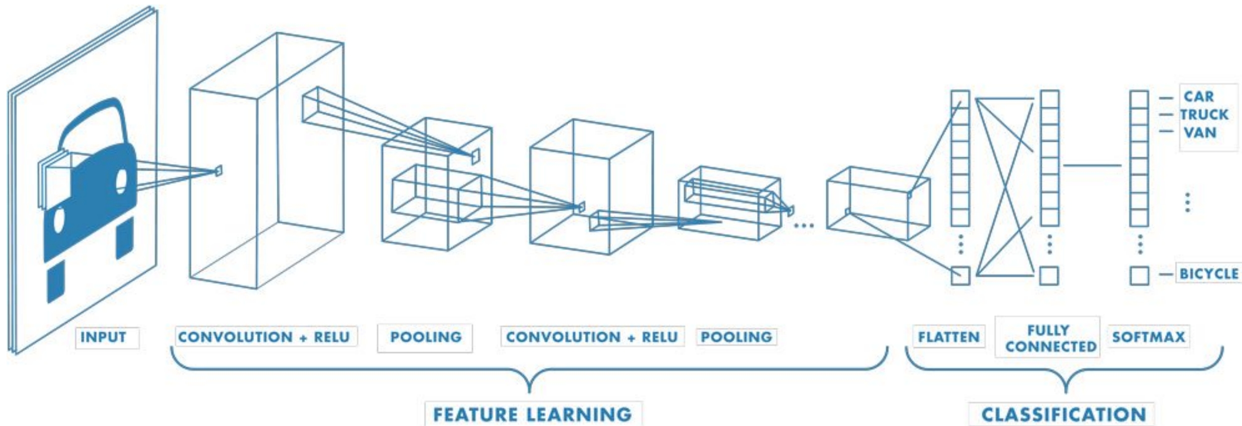


Figure 12. (Source: https://www.mathworks.com/discovery/convolutional-neural-network-matlab.html)

For visual data, convolutional neural networks (CNN), first proposed by Lecun in 1989, have been proved to have excellent performance in feature extraction.[14] The layers in between the network intend to extract and learn features specific to the target from pixels. Convolution, Rectified linear unit (ReLU) , and pooling are three of the most common layers applied in CNN. They are repeated over hundred times in order to identify from low-level features (e.g. eye, nose, ear) to higher-level features (e.g. face).

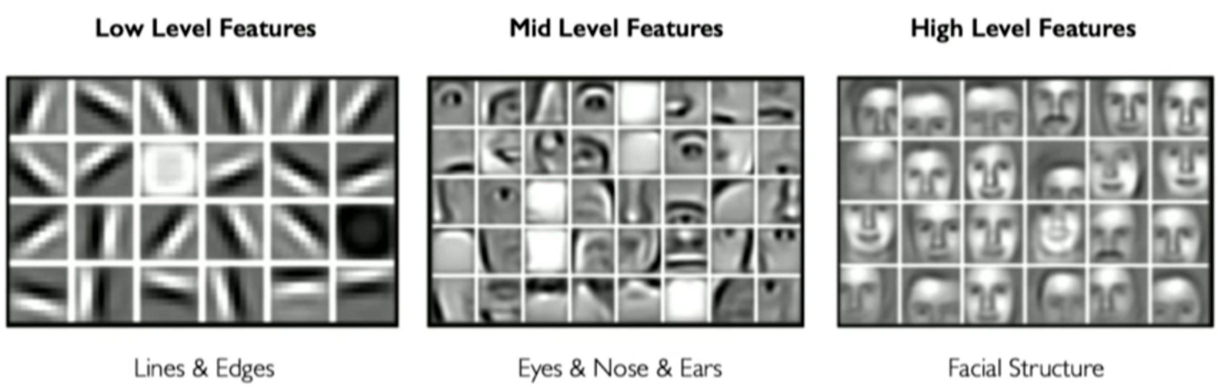


Figure 13. CNN from low-level feature layers to high-level feature layers
(source: http://introtodeeplearning.com/)

[14] LeCun, Yann, et al. “Backpropagation applied to handwritten zip code recognition. “Neural computation 1.4 (1989): 541-551.

In order to measure the spatial properties from single image data, the CNN technique can be used for object recognition and scen attribute extraction.

YOLOv2, by Joseph Redmon, has been widely used for object recognition. CNN architectural based model has phenomenal performance for image classification.

2.4.3 Video-Based Representation of Space

When one considers the adoption of technologies in various industries, the architecture industry may not come to mind, yet that has begun to change, starting with the inclusion of data in the design process. New interest in innovative technology in architecture comes from necessity, and utilization of big data can help industry practitioners become more flexible, efficient and secure in the design build process. From a user-oriented perspective, there are multiple benchmark datasets that focus on analyzing behavior through space. Moments in Time, co-developed by MIT and IBM, is a large-scale dataset for recognizing and understanding motion in videos. The dataset includes a collection of more than one million labeled videos involving people, animals, objects, or natural phenomena, capturing dynamic scenes. It is designed to have a large coverage and diversity of events in both visual and auditory modalities, and can serve as a new challenge to develop models that scale to the level of complexity and abstract reasoning that a human processes on a daily basis. An ever-going part of our consideration in architecture is the "built environment", and it contains a vast digital footprint and visual data if recorded. Effective exploration and analysis of the data requires intelligent tools and supporting benchmarks like Moments in Time, that can help architects, designers and engineers to gain a deeper understanding of how humans and objects interact in distinctive spatial design, and inspire ideas that can be applied to our current practices.

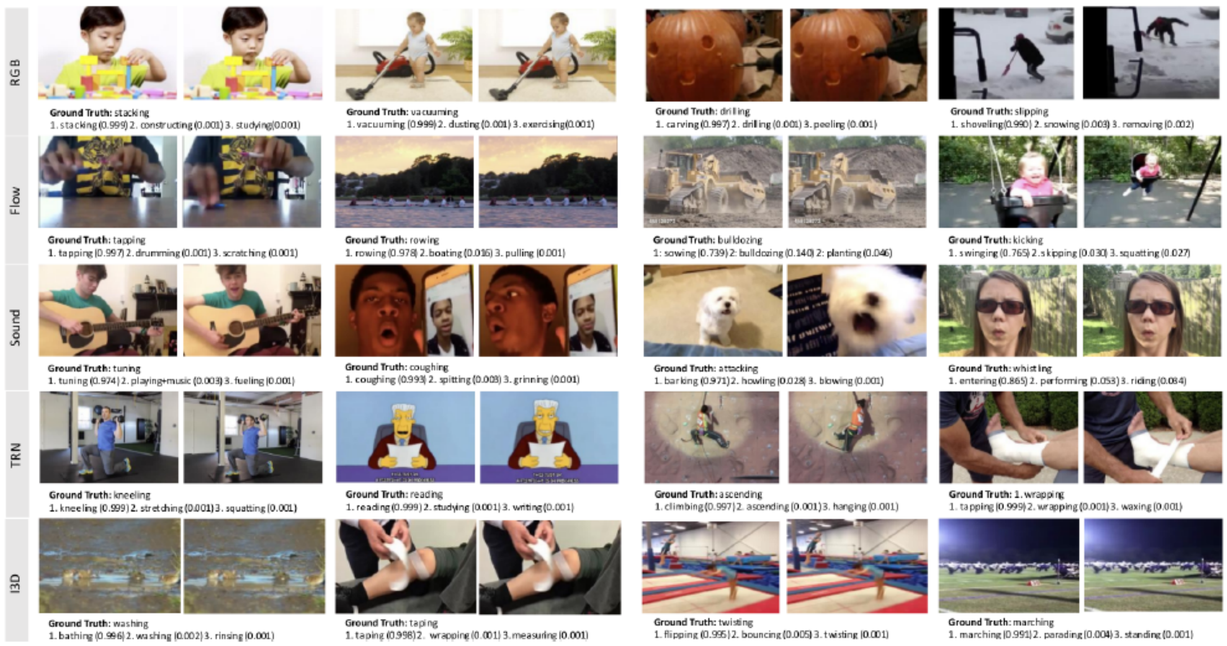


Figure 14. Moments In Time, Aude Oliva (2019)

2.4.4 Scene Attribute Recognition

Figure 15. Places Dataset, Bolei Zhou (2017)

Figure 16. Places Dataset, Bolei Zhou (2017)

3 | INVESTIGATION

3.1 General Pipeline

The hypothesis of the experiment is that video-based data facilitates the analysis of user behaviors in built environments in order to provide insights of correlation.

In order to explore the possible uses of video data, a framework of methodology is proposed to study the relationships between built environments and user behaviors:

1. Take Pattern 88 - Street Cafe as an example for investigation. Retracing one of Alexander's patterns by using data-driven methods provides bottom-up insights of correlations.
2. Collect data from several street cafes for comparisons. Visit street cafes with similar spatial settings as Alexander’s described but with slight variation of details in order to compare the differences of user behaviors in different settings.
3. Analyze user behaviors by using computational models to recognize actions in video data.
4. Analyze spatial settings by using computational models to recognize the spatial properties.
5. Discover correlations between user behaviors and spatial settings and to provide insights for spatial design.

Road Map

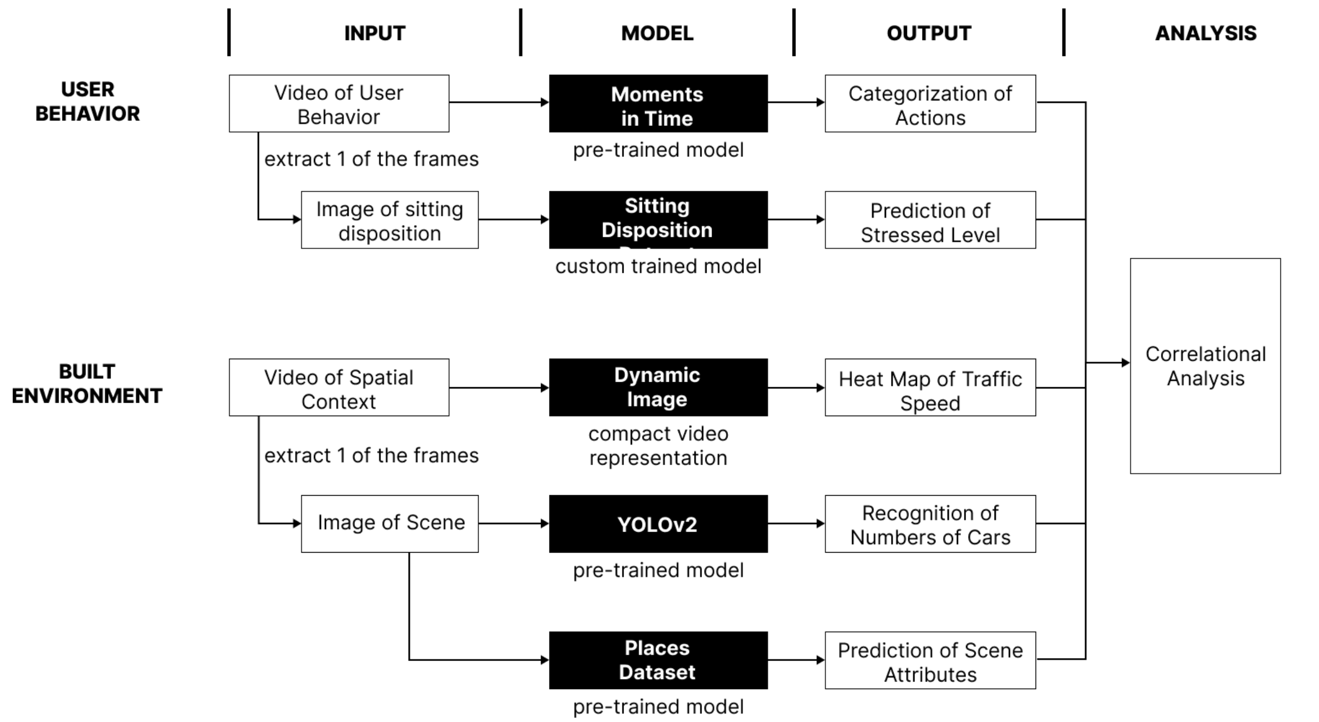


Figure 17. General road map of methodology for analysis, Charles Wu (2022)

3.2 A Case of Street Cafe

A Pattern Language, by Christopher Alexander, specify architectural cases in which to study user actions. Alexander tried to study the relationship between function and patterns of design that form an architectural language. *A Pattern Language* consists of defined architectural design sub-cases where people and architecture are composed.[15]

[15] Alexander, Christopher. *A pattern language: towns, buildings, construction*. Oxford university press, 1977.

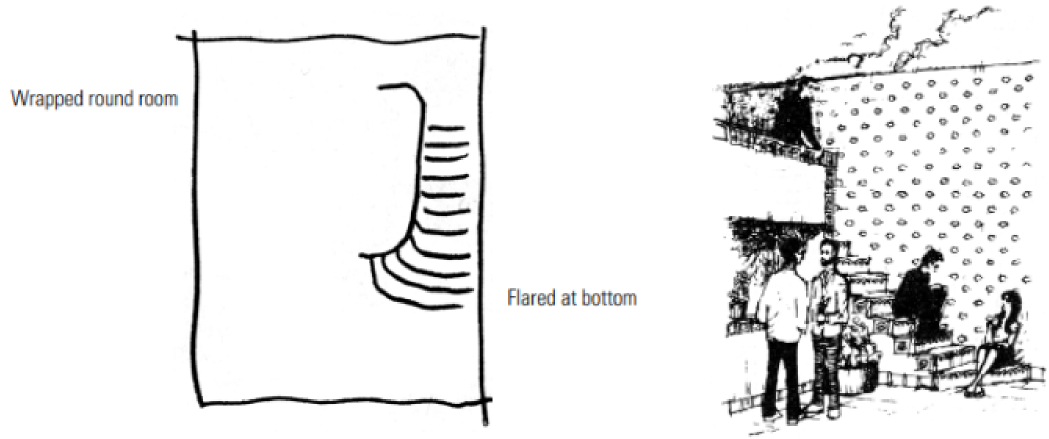


Figure 18. Pattern 133 - staircase as a stage, Christopher Alexander (1977)
Left: floor plan, Right: image of imagination

For instance, pattern 133 (Staircase as a stage) described that flared out staircase can be used as stair seats and it transforms the staircase into a social space where people would be naturally inclined to sit, chat or do something;

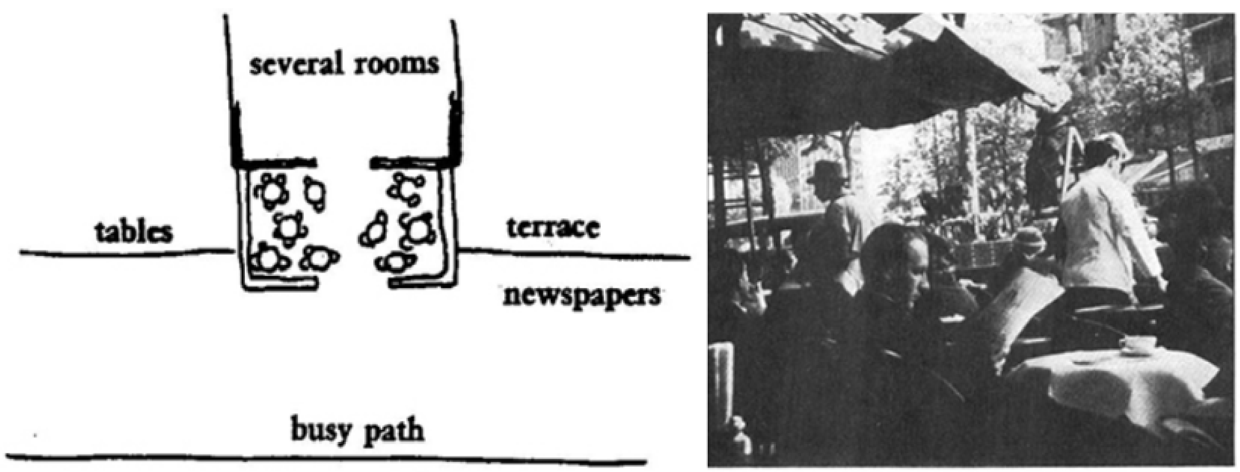


Figure 19. Pattern 88 - street cafe, Christopher Alexander (1977)
Left: floor plan, Right: image of imagination

Pattern 88 (Street cafe) mentioned that an intimate space which is open to a busy path would invite people sit lazily, legitimately, and watch the world go by.

256 patterns were proposed in order to provide a practical language that enables everyday users to become conscious of their living patterns. However, the solutions subjectively observed by Alexander were from a top-down perspective to describe user behaviors and space. The prescriptive approach has certain limitations in reality:

1. The spatial context described was too general. Not insightful in a real situation.
2. User behaviors differ from variations of spatial properties. The correlations between user and spatial settings were not verified.

In order to retrace or expand the patterns, data-driven techniques play a good role to provide bottom-up suggestions. Pattern 88 (Street cafe) is taken as an example for a public observation experiment.

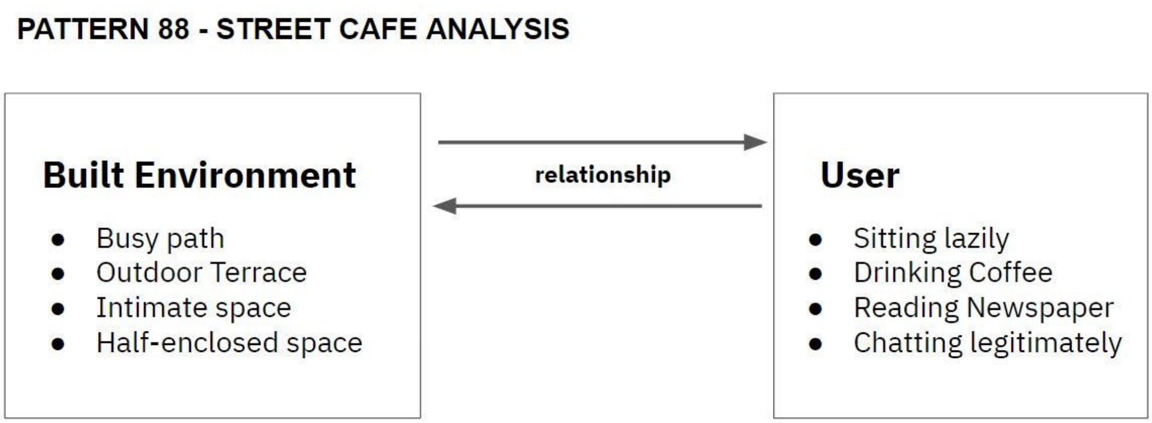


Figure 20. Breakdown of street cafe scenario into the built environment and the user, Charles Wu (2022)

3.3 Data Collection / Experiment Setup

In order to investigate the pattern 88 (Street cafe), 23 street cafes were visited. (See Apendix B)

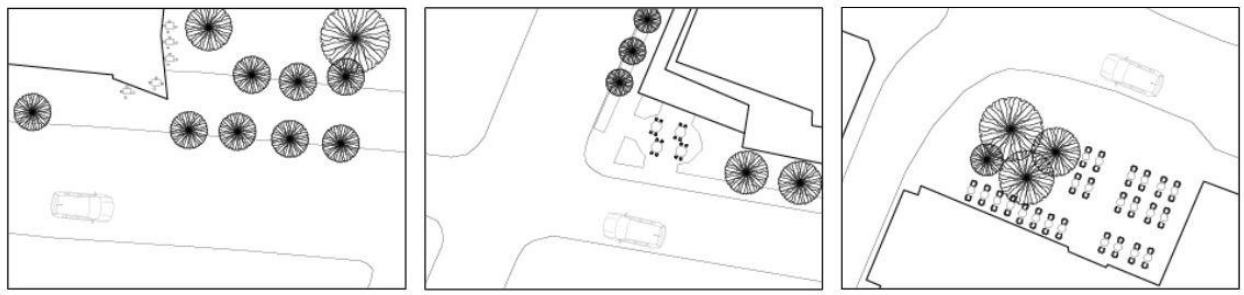


Figure 21. Street cafes with different spatial settings for investigation, Charles Wu (2022)

Methods

Firstly, in order to collect data of user behaviors, videos were recorded as close as possible. In terms of public observation, subjects should not be disturbed during investigation. Hence, the camera was set next to the subjects' table for close-up video recording (Figure 18). A roughly 10-second video was obtained for one user action.

Secondly, in order to collect data of spatial settings, videos were recorded at different positions for a more comprehensive vision. One camera was placed at the entrance of the street cafe to obtain a complete view of the spatial settings, whereas another camera was placed besides the subject to record what the subject sees as the context of the cafe while sitting in it.

Different camera position settings mimic human observers walking around to study the space.

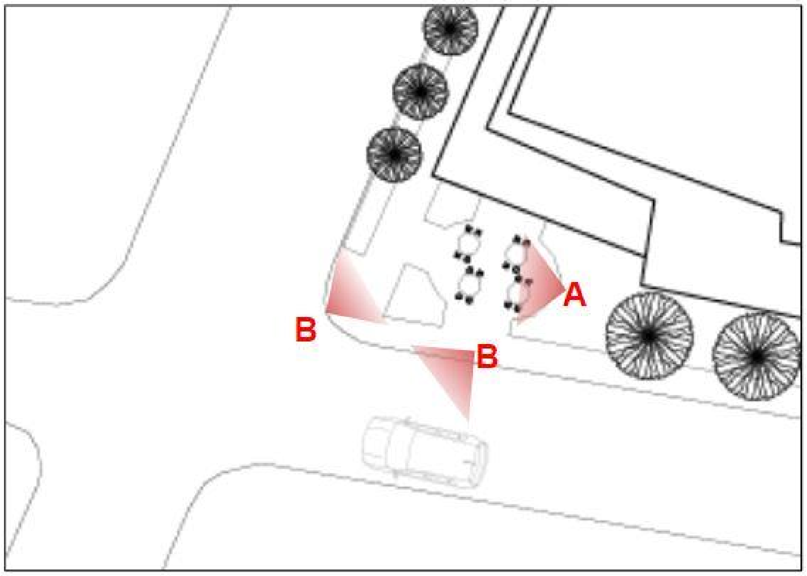


Figure 22. A represents the camera position to record close-up videos of user behaviors. Bs represent the camera positions to record spatial settings of street cafes, Charles Wu (2022)

Constants

In order to maintain consistency of the experiment setup, there are several constant variables of video recordings:

- Studied cafes are around Cambridge, Massachusetts
- All videos are recorded within the time from 1 pm to 4 pm in March with similar weather condition.

3.4 User Behavior Analysis

The discussion of human-centered design continues by establishing understanding of human behaviors in space. Users usually stay in a space for a period of time (it can be 5 mins, or an hour). For street cafes, users usually walk into cafes, order food or coffee, grab a newspaper and sit down, drink the coffee…etc. Many moments happened in different places, and involved people, animals, objects and natural phenomena. They are unfolded at time scales from a second to minutes, and the best way to record it as video data.

3.4.1 Video-based action classification

The Moments in Time Dataset, proposed by Monfort et al in 2019, collected one million short videos each with a label corresponding to an event unfolding in 3 seconds.[16] The events of humans are categorized into low-level actions and labeled from 330 different classes, such as eating, running, picking and sitting. It is important to start with low-level actions as activities that occur at longer time scales can be represented by sequences of three second actions. For example, speaking and throwing an object could be interpreted as the compound actions “fighting”, “playing basketball” or “juggling” depending on the context of the activity (e.g. agent and scene). Hypothetically, when describing such a “fighting” event, one can decompose it into the details of the movement of each joint and limb of the persons involved. Hence, the ability to automatically recognize these short actions is a core step for automatic video comprehension. By implementing the dynamic-temporal models from the Moments in Time Dataset, we can develop a computational perception of understanding the moments happened in different space such as cafes. The research question becomes: how would people behave differently in different street cafe settings?

After collecting video data of 23 different street cafes, Moment In Time dataset was employed to recognize users’ actions. An experiment was carried out to study what kind of actions may occur

[16] M Monfort, A Andonian, K Ramakrishnan, T Yan, A Oliva are with Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139 USA.

at different street cafes with different spatial settings and if the pre-trained model can recognize it precisely. There are a few interesting examples as follow:

In Cafe 1, there were two persons sitting at the entrance of it. It was a squeezed recess cafe terrace with shade. The prediction results were “dining” with 0.724 accuracy, “serving” with 0.144, and “discussing” with 0.067 accuracy. “Socializing” and “speaking” were also detected, but the low accuracies of prediction should be neglected. The prediction results were promising, except “serving”. It also indicated the ability of recognizing multi-person events as general action classification models usually have low accuracy of prediction when there are more than two people in the scene.

Prediction	Accuracy
dining	0.724
serving	0.144
discussing	0.067
socializing	0.017
speaking	0.017

Figure 23. Two persons were discussing in cafe 1, Charles Wu (2022)

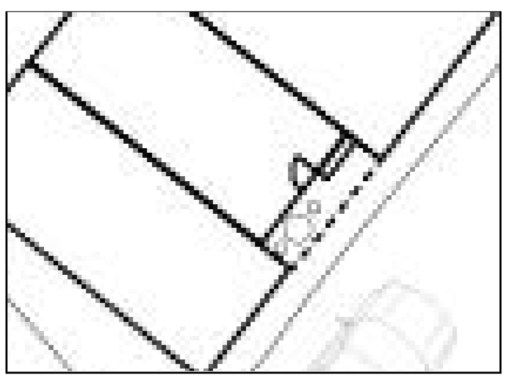


Figure 24. Floor plan of cafe 1, Charles Wu (2022)

Cafe 8 was located next to a large open green area. People are more relaxed while enjoying their coffee under sunshine. “Laying” with 0.432 accuracy was detected from the video, along with “sitting”, “speaking”, “yarning”, and “kicking”. The results were convincing even the background of the video was complex.

Prediction	Accuracy
laying	0.432
sitting	0.241
speaking	0.061
yarning	0.021
kicking	0.021

Figure 25. People were relaxing in Cafe 8, Charles Wu (2022)

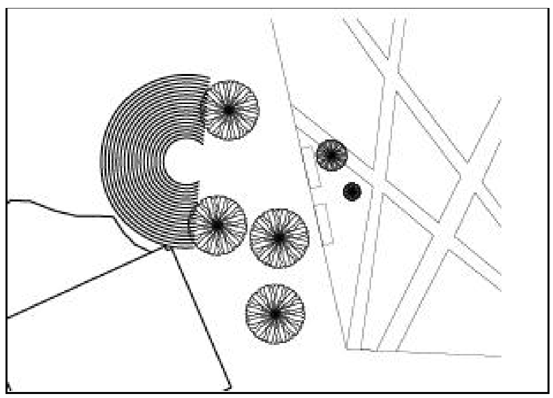


Figure 26. Floor plan of cafe 8, Charles Wu (2022)

Cafe 3 has a huge shaded area at the entrance with hard pavement. Unlike cafe 8, user tends to playing with their phone instead of laying down. The model successfully recognized “dialing” with 0.432 which is relatively acceptable as a low-level actions. However, “filming” was incorrect with 0.241 accuracy. Comparing cafe 3 with cafe 8, they both consist of large open area, but one is shaded while another is not. It is interesting to study if the shade contributes to the difference of user behaviors.

Prediction	Accuracy
dialing	0.432
filming	0.241
taping	0.061
bending	0.021
rubbing	0.021

Figure 27. User was playing with his phone in cafe 3, Charles Wu (2022)

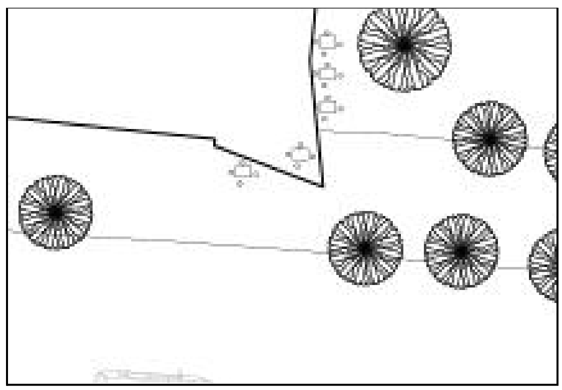
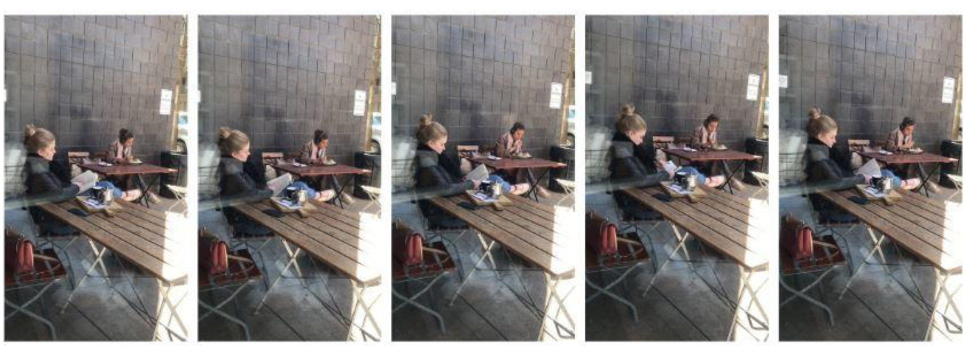


Figure 28. Floor plan of cafe 3, Charles Wu (2022)

Cafe 16 is similar to cafe 3. The floor is covered with hard pavements but without shade. In addition, it is open to a more busy street than cafe 10. The user was “reading”, “writing”, and “autographing”. One interesting observation is that the motion of users in cafe 16 was very slow (almost static while reading the book). The speed of motions seems to be considered during the process of classification.



Prediction	Accuracy
reading	0.890
writing	0.665
autograph	0.632
snapping	0.034
painting	0.039

Figure 29. User was reading slowly in cafe 16, Charles Wu (2022)

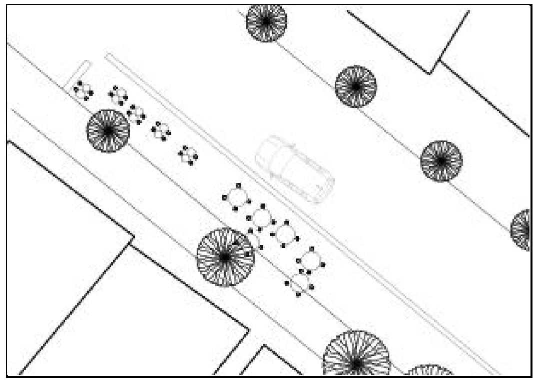


Figure 30. Floor plan of cafe 16, Charles Wu (2022)

Cafe 13 is located at an urban-scaled open plaza with 20 sets of tables. Two-third of the tables are completely open while one-third of them are aside the building with shade. Two sets of videos were recorded in two different settings for comparison. In the open area, a group of people were playing chess and discussing. The prediction results were disappointing that “rocking” with 0.072 accuracy was detected which did not make any sense. The reason for the low accuracy is believed to be that there were more than two people in the scene. The complicated motions of multiple people confused the model.

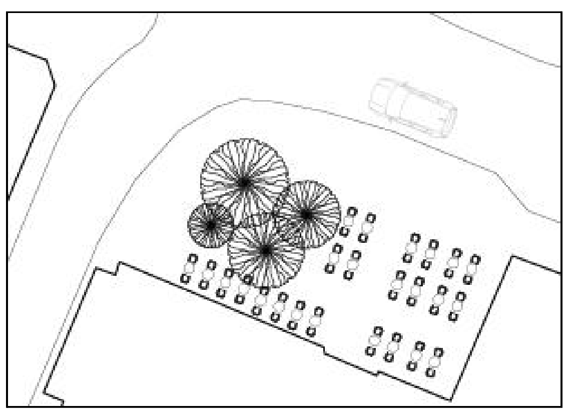


Figure 31. Floor plan of cafe 13, Charles Wu (2022)

Prediction	Accuracy
rocking	0.072
juggling	0.034
shouting	0.031
speaking	0.031
starting	0.028

Figure 32. A group of people were playing chess in cafe 13, Charles Wu (2022)

The shaded area at cafe 18 invited some people to work with their laptops. However, the results were "opening" with 0.235 accuracy, and "closing" with 0.197 accuracy, which were inaccurate. There were two main reasons for this inaccuracy. First, the recording distance was too far away from the subject. Second, there were people passing by that interrupted the data collection process.

Despite of the imprecision of classification, different types of actions were observed in different spatial settings. It likely suggests that users tend to work in shaded areas and relax in open areas.

Prediction	Accuracy
opening	0.235
closing	0.197
entering	0.103
exiting	0.028
barking	0.018

Figure 33. User was working under shade area in cafe 13, Charles Wu (2022)

After the analysis of 23 cafe video recordings, the results are summed up in the following table (table 1). Certain popular cafe-related actions are listed for comparison. It is noted that some results are not available as the video recordings did not record every single user in the cafes.

Hence, it is possible that users were drinking in the cafes but they were not captured. It suggests the use of 360 camera for data collection process may solve the problem in the future studies.

	sitting	drinking	eating	reading	speaking	typing	interviewing	grabbing	…	writing
Cafe 1	✔	✔	✔	✔	✖	✖	✔	✖		✔
Cafe 2	✔	✔	✔	✔	✖	✔	✖	✔		✖
Cafe 3	✔	✖	✔	✖	✔	✔	✔	✖		✔
Cafe 4	✔	✔	✖	✔	✖	✖	✔	✖		✖
Cafe 5	✔	✔	✔	✖	✔	✖	✖	✖		✔
Cafe 6	✔	✖	✔	✖	✖	✔	✖	✔		✖
Cafe 7	✖	✔	✔	✔	✔	✔	✔	✔		✔
…										
Cafe 23	✔	✖	✔	✖	✔	✖	✔	✔		✔

Table 1. Certain populate cafe-related actions are listed for comparison, Charles Wu (2022)

Categorization of Lexicon to Different Events

While the recognition of low-level actions is promising, a deeper understanding of higher-level events could be possibly developed. For instance, dining includes chewing, biting, dipping, drinking, speaking…etc. However, speaking can also happen during working and socializing; drinking can also happen during exercising and partying. It is difficult to understand the context of events from a single action recognition. Hence, further categorization of action tags forms a network which connects all the actions together in different event situations.

Due to the pandemic, people use space very differently from the past. For example, working in a living space, exercising in a garage, eating on the street…etc. It is interesting to study how it is changed in terms of spatial setting when multi-used space is encouraged and promoted.

Categorization of lexicon to different events constructs a higher-level understanding of events by gathering low-level actions. For example

1. Dining includes: sitting, eating, speaking, drinking, grabbing…etc
2. Working includes: sitting, typing, speaking, interviewing, writing…etc
3. Exercising includes: running, lifting, squatting, throwing, drinking…etc

It is worth mentioning that the lexicon of the Moments In Time dataset was mixing the low-level actions with the high-level events, for example, socializing and drinking were both included. It affects the accuracy of predictions with unclear categorization.

HIGH-LEVEL EVENTS	Working	Socializing	Exercising
LOW-LEVEL ACTIONS	Sitting Typing Speaking Interviewing Writing....	Eating Cooking Speaking Drinking....	Grabbing Squatting Lifting Drinking....

Figure 34. Categorization of low-level actions to high-level events, Charles Wu (2022)

Sequential Analysis of Actions

Last but not least, the sequential analysis of actions is essential for user behavior study. An experiment was carried out to record a subject sitting at cafe 21 for half an hour. The video was segmentized into many 4-second clips for action classification by Moments In Time pre-trained model.

Figure 35. Screenshots of the half-hour recording of a subject in cafe 21, Charles Wu (2022)

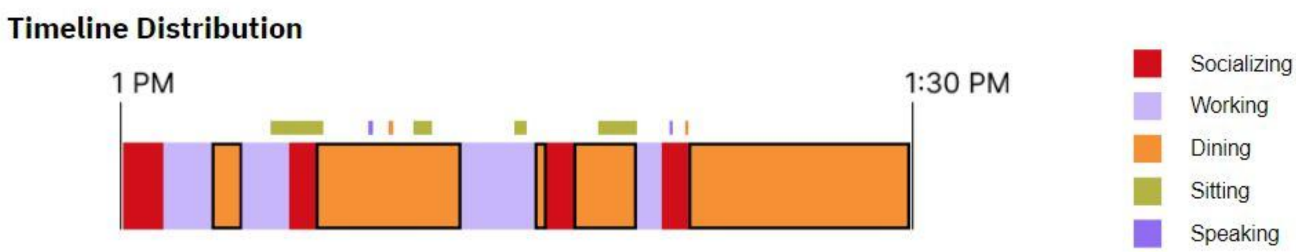


Figure 36. The timeline distribution of actions within the half hour in cafe 21, Charles Wu (2022)

The timeline distribution showed the subject spent half of the time dining. In the meantime, they switched to work periodically. They were also socializing some of the time.

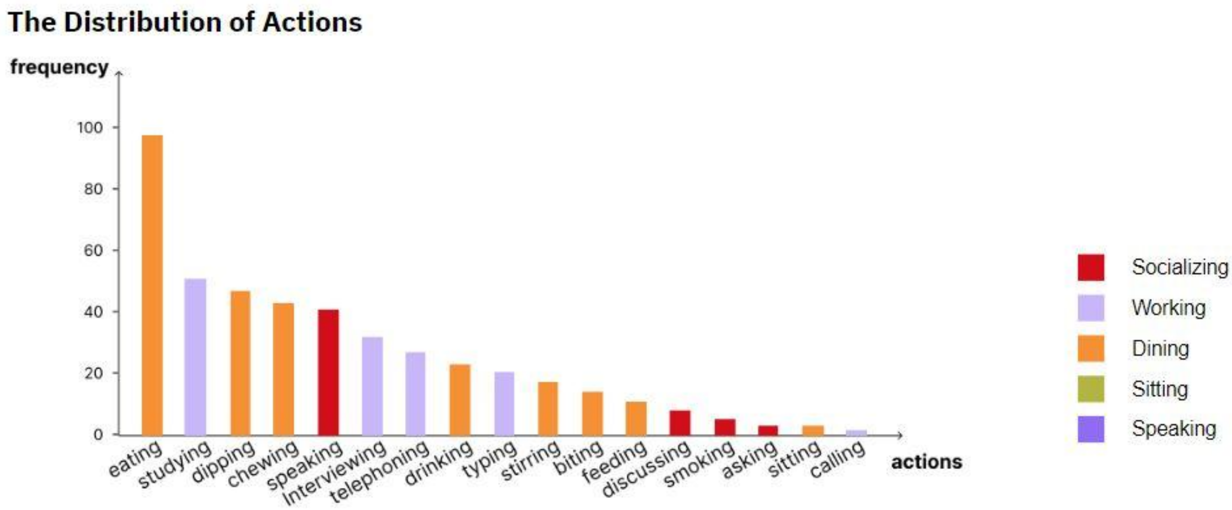


Figure 37. The frequency of different actions happened in cafe 21, Charles Wu (2022)

The frequency plot also demonstrated that actions related to dining dominate the half-hour, associated with work-related actions.

The data visualization of user behavior suggests living patterns in space and can be used as a reflection of daily lives. The data-driven methods enable everyday users to become conscious of their living patterns.

3.4.2 Sentimental Analysis of Actions

After a deep analysis of moments in space, I realized it is a purely objective observation without any sentimental interpretation. However, human perception is capable to understand how people feel by observing their actions and body language. For example, one is stressed while running fast whereas relaxed while walking slowly; one is stressed while sitting tight whereas relaxed while laying down.

Taking sitting disposition as an example, humans could easily recognize the level of relaxedness by observing the images of sitting dispositions below. According to Alexander's description, people 'sit lazily' in street cafes. When Alexander, as a human, is able to recognize people 'sitting lazily', machines can possibly recognize it as well.

Figure 38. The level of relaxedness of different sitting dispositions, Charles Wu (2022)

A custom-trained machine learning model was proposed to predict the relaxedness of users by analyzing their sitting disposition recorded in videos. 864 frames of recorded videos were extracted as a dataset for a CNN prediction model. The images were labeled manually on the crowd-sourcing platform Amazon Mechanical Turk (AMT). AMT workers are expected to determine the relaxedness of sitting dispositions. Label '1' as stressed, label '3' as neutral, and label '5' as relaxed.

After the data collection process, the model was trained under the CNN architecture. The sequential model consists of three convolution blocks, with a max pooling layer in each of them. There is a fully-connected layer with 128 units on top of it that is activated by a ReLU activation function.[17]

```
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 sequential_1 (Sequential)   (None, 180, 180, 3)       0

 rescaling_2 (Rescaling)     (None, 180, 180, 3)       0

 conv2d_3 (Conv2D)           (None, 180, 180, 16)      448

 max_pooling2d_3 (MaxPooling  (None, 90, 90, 16)       0
 2D)

 conv2d_4 (Conv2D)           (None, 90, 90, 32)        4640

 max_pooling2d_4 (MaxPooling  (None, 45, 45, 32)       0
 2D)

 conv2d_5 (Conv2D)           (None, 45, 45, 64)        18496

 max_pooling2d_5 (MaxPooling  (None, 22, 22, 64)       0
 2D)

 dropout (Dropout)           (None, 22, 22, 64)        0

 flatten_1 (Flatten)         (None, 30976)             0

 dense_2 (Dense)             (None, 128)               3965056

 dense_3 (Dense)             (None, 5)                 645

=================================================================
Total params: 3,989,285
Trainable params: 3,989,285
Non-trainable params: 0
_________________________________________________________________
```

Figure 39. All the layers of the network used in the CNN model, Charles Wu (2022)

[17] "Image Classification" TensorFlow, last modified Jan 26, 2022, https://www.tensorflow.org/tutorials/images/classification

Overfitting Results

In figure 39, the training accuracy is increasing linearly over time, whereas validation accuracy stalls at around 60% in the training process. Also, the difference in accuracy between training and validation accuracy is noticeable—a sign of overfitting.

When there are a small number of training examples, the model sometimes learns from noises or unwanted details from training examples—to an extent that it negatively impacts the performance of the model on new examples. This phenomenon is known as overfitting. It means that the model will have a difficult time generalizing on a new dataset.

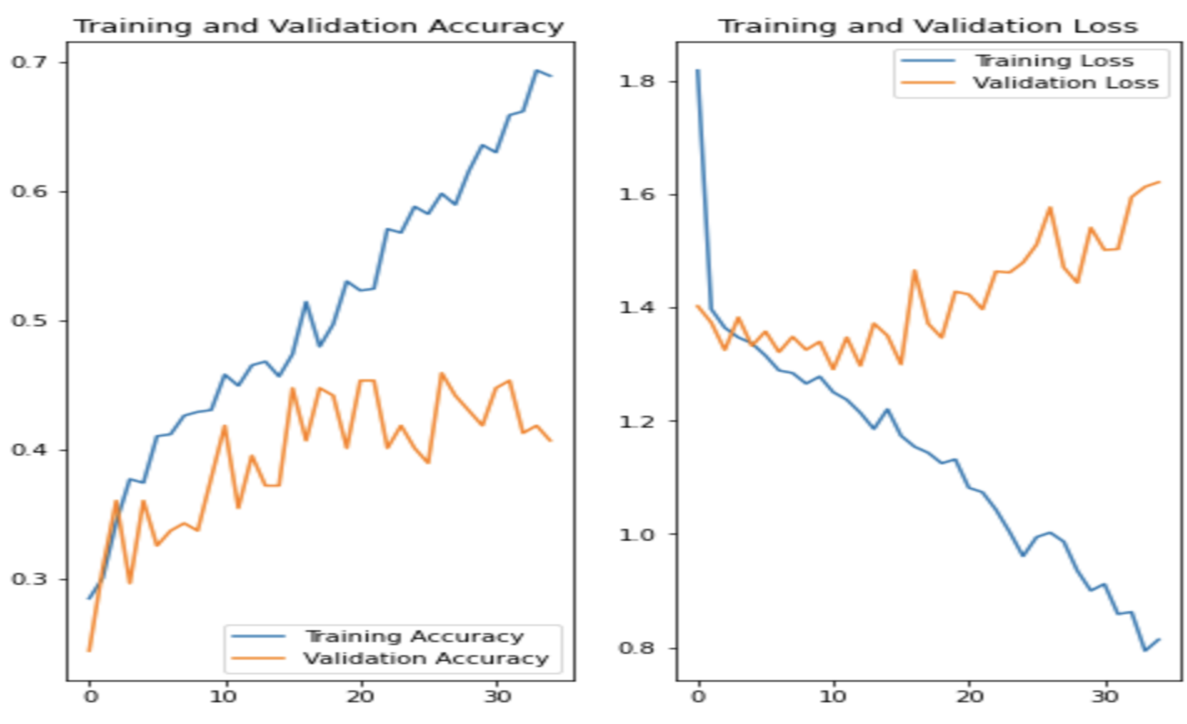


Figure 40. Standard approach with low validation accuracy, Charles Wu (2022)

There are multiple ways to fight overfitting in the training process. In this tutorial, you'll use data augmentation and add Dropout to your model.

Data augmentation

Overfitting generally occurs when there are a small number of training examples. Data augmentation takes the approach of generating additional training data from your existing examples by augmenting them using random transformations that yield believable-looking images. This helps expose the model to more aspects of the data and generalize better.

Figure 41. Data augmentation to increase the size of dataset, Charles Wu (2022)

Dropout

Another technique to reduce overfitting is to introduce dropout regularization to the network. When you apply dropout to a layer, it randomly drops out (by setting the activation to zero) a number of output units from the layer during the training process. Dropout takes a fractional number as its input value, in forms such as 0.1, 0.2, 0.4, etc. This means dropping out 10%, 20%, or 40% of the output units randomly from the applied layer.

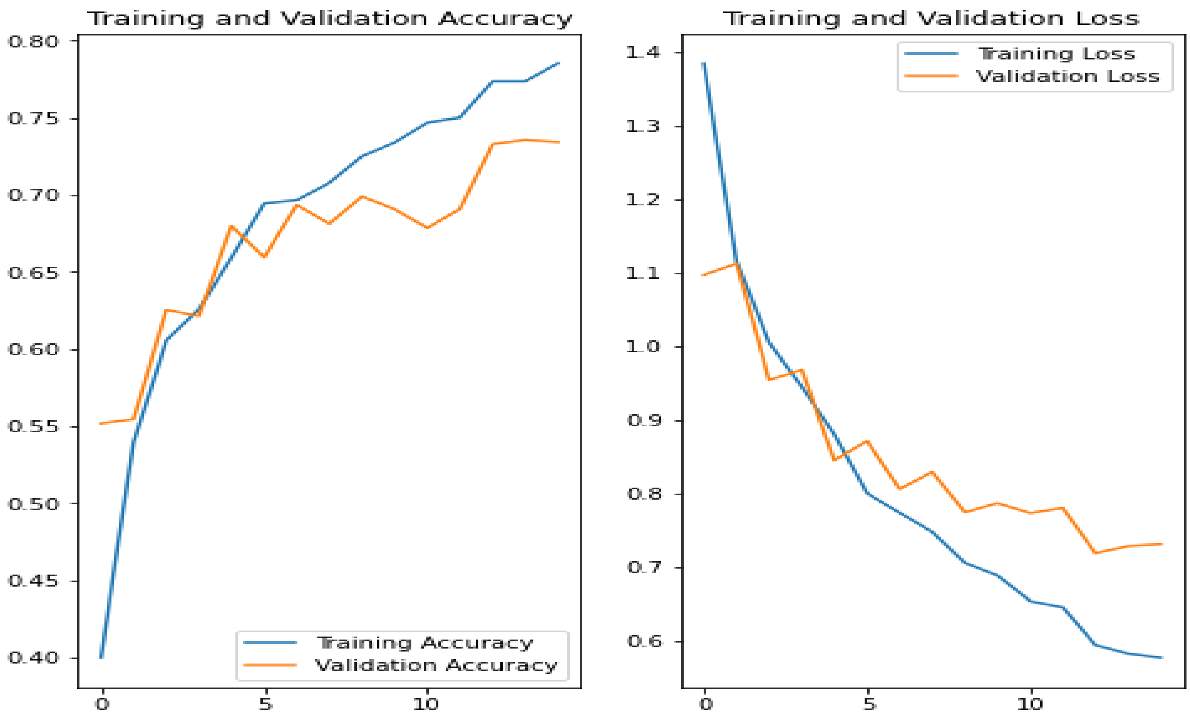


Figure 42. Fine-tuned result with promising validation accuracy, Charles Wu (2022)

3.5 Built Environment Analysis

In Pattern 88 Street Cafe, Alexander mentioned,

“The street cafe provides a unique setting, special to cities: a place where people can sit lazily, legitimately, be on view, and watch the world go by.”

It is such an attractive and relaxing experience to “sit lazily” and enjoy a cup of coffee at a street cafe. Trying to understand why it is so attractive, Alexander explained that people enjoyed mixing in public, in parks, and squares, along promenades and avenues, and in street cafes. The setting implied the precondition that gives you the right to be there, and people feel safe enough to relax, nod at each other, and perhaps even meet. There are a few conditions, as Alexander described, to fulfill a good cafe terrace that allows a person to sit there for hours in public. First, it anchors in the local neighborhood. Second, it is open to the street. Third, it contains several other spaces such as soft chairs, fire, and newspapers. Last but not least, it is a half-public, half-private space. The hand-sketched floor plan Alexander showed confirmed those criteria of a good street cafe terrace.

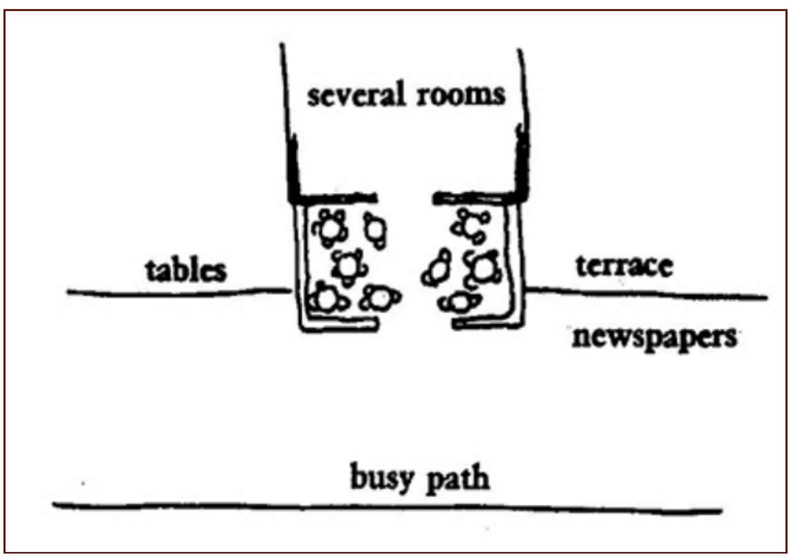


Figure 43. Floor plan of street cafe pattern, Christopher Alexander (1977)

It is clear that several spatial conditions are essential to allow certain user behaviors. Before investigating the user behaviors in space, it is essential to systematically measure the space to test which properties trigger different user behaviors. Space could be broken down into quantifiable factors such as length, width, geometry, and scale. According to Alexander's description, I summed up six main spatial properties to measure street cafes: (1) Greenery, (2) Density, (3) Enclosure, (4) Traffic, (5) Coverage, and (6) Materials.

3.5.1 Understanding the Properties of the Built Environment

In order to collect data on the six spatial properties and allow the machine system to understand space from similar attributes, a methodology is proposed to develop a computational perception of space. AI techniques are explored and employed to analyze visual recordings (e.g. single image and video) such as object recognition, scene attributes extraction, and color vividness extraction. These techniques help machine systems understand low-level features of space and could be further studied as properties of the built environment.

Object recognition

CNN has been widely used for object recognition based on single images. Numerous pre-trained machine learning models are open-sourced for the public to study. I conducted an experiment to explore if it would be able to provide insight to understand the spatial settings of street cafes. I used a pre-trained model from the OpenCV (Open Source Computer Vision Library) to analyze photos of street cafes.[18] The model is capable of recognizing 91 different categories of objects, such as buses, birds, elephants, shoes, and hats.

41 images of street cafes were studied and it is capable to recognize a good number of low-level components such as potted plants, rocks, bins, and people. Although the accuracy is satisfying, some objects cannot be recognized when it is overlapped in the photos due to the angle of taking photos. However, it has an outstanding performance for recognizing persons. Even if the body is blocked by some other objects in the photo, the head can still be precisely recognized. Hence, it could be used to analyze the density of street cafes by counting the number of users.

2 persons
5 potted plants
1 rock
1 bin

Figure 44. Object recognition by YOLOv2, Charles Wu (2022)

[18] "OpenCV", OpenCV Team, last modified 2022, https://opencv.org/

Left: the image of scene Right: Recognition result

The scene attributes

Although the object recognition algorithm is useful, there are several disadvantages of the conventional categorical label approach:

1. The experiment by Biederman suggested that people did not recognize space through object information and details but through the encoding of the global configuration.[19] We understand space as a whole through assembling low-level features we see.
2. Space can never be described by just one or two labels, most of the time-space could fall into more than one category. For instance, one space can be large and quiet, while another can be large and busy. The multidimensionality of space should be addressed by different levels of scene attributes.
3. Categorical recognition technique can be easily “fooled” or confused by adding some details to images. As the figure is shown below, it is difficult for a machine learning engine to distinguish the differences between dogs and mops, while it is easy for humans to do that. The categorical algorithm is not robust in certain conditions.

Figure 45. (source: https://dev.to/swyx/serverless-machine-learning-at-google-cp9)

[19] Biederman, Irving. "Recognition-by-components: a theory of human image understanding." *Psychological review* 94, no. 2 (1987): 115.

To overcome the shortcomings of conventional categorical classification for space, Zhou et al. developed in-depth the Places Database and evaluated the place recognition using Convolutional Neural Networks (CNNs).[20] The Places database aims to create a human-like performance for scene recognition, which requires a higher level of abstraction and understanding of space compared to conventional object recognition. The Places database consists of 10 million scene photographs, labeled with 434 scene semantic categories.

The model was trained based on three CNN architectures, **VGG 16 convolutional-layer CNN**[21], **AlexNet**[22], and **GoogLeNet**[23], then the baseline models are tested on **Places205**[24] and **Places365-Standard**[25].

I conducted an experiment that employs Places to analyze images of street cafes. Taking the same image for object recognition as an example, several scene attributes are predicted, including man-made, no horizon, open area, enclosed area, horizontal components, sunny, wood, and metal. The results are promising in terms of providing insights to understand the spatial setting of street cafes.

---

[20] Zhou, Bolei, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. "Places: A 10 million image database for scene recognition." *IEEE transactions on pattern analysis and machine intelligence* 40, no. 6 (2017): 1452-1464.

[21] Qassim, Hussam, Abhishek Verma, and David Feinzimer. "Compressed residual-VGG16 CNN model for big data places image recognition." In *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 169-175. IEEE, 2018.

[22] Iandola, Forrest N., Song Han, Matthew W. Moskewicz, Khalid Ashraf, William J. Dally, and Kurt Keutzer. "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size." *arXiv preprint arXiv:1602.07360* (2016).

[23] Zhong, Zhuoyao, Lianwen Jin, and Zecheng Xie. "High performance offline handwritten chinese character recognition using googlenet and directional feature maps." In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 846-850. IEEE, 2015.

[24] Wang, Limin, Sheng Guo, Weilin Huang, and Yu Qiao. "Places205-vggnet models for scene recognition." *arXiv preprint arXiv:1508.01667* (2015).

[25] Li, Chau Yi, Ali Shahin Shamsabadi, Ricardo Sanchez-Matilla, Riccardo Mazzon, and Andrea Cavallaro. "Scene privacy protection." In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2502-2506. IEEE, 2019.

Type of environment: **outdoor**
Scene Attributes:

- **man-made**
- **no horizon**
- **natural light**
- **open area**
- **enclosed area**
- **horizontal components**
- **sunny**
- **wood**
- **metal**

Figure 46. Scene attribute prediction using Places dataset, Charles Wu (2022)

Time-related attribute (e.g. traffic)

Apart from understanding the spatial setup of street cafes, it is also important to recognize their context of them. From the floor plan, Alexander drew, the “busy path” is clearly labeled as one of the key elements of street cafes. The research question becomes: How can we create a computational perception to tell if the street is busy or not?

As a human perception, traffic conditions could be addressed mainly by two factors: (1) the number of cars (2) the speed of cars

The number of cars

The number of cars can be detected through object recognition as mentioned above. The results are robust and promising as shown below.

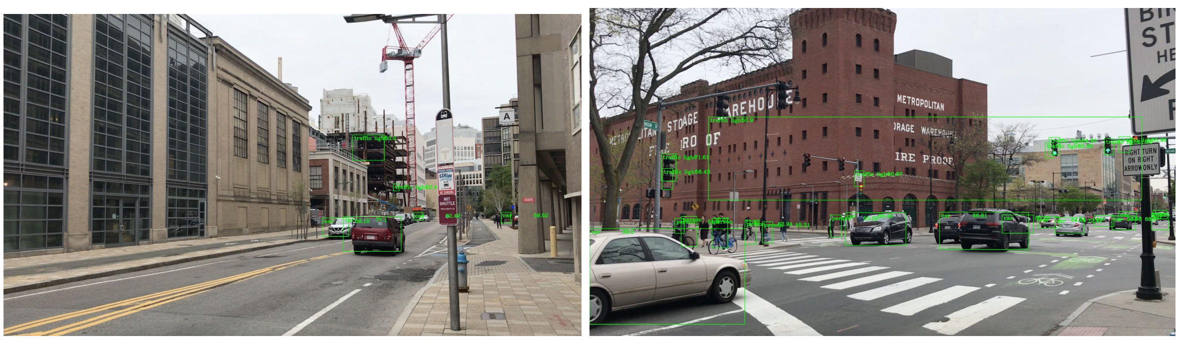


Figure 48. Left: Vassar St with fewer number of cars Right: Mass Ave. with higher number of cars, Charles Wu (2022)

The speed of cars

Speed and motion as time-related attributes are recorded as videos. To analyze video data using CNN architecture, it is necessary to consider how video information should be presented as input. Bilen et al proposed the concept of ‘dynamic image’ in which the video content is compacted as a single still image which then could be processed by a standard CNN architecture such as GoogLeNet.[26]

---

[26] Bilen, Hakan, Basura Fernando, Efstratios Gavves, Andrea Vedaldi, and Stephen Gould. "Dynamic image networks for action recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3034-3042. 2016.

It is observed that dynamic images behave differently for actions at different speeds. For slow actions, like “blowing hair dry”, the motion seems to be dragged over many frames. For faster actions, such as “golf swing”, the dynamic image reflects key steps in the action such as preparing to swing and stopping after swinging. For longer-term actions such as “horse-riding”, the dynamic image reflects different parts of the video; for instance, the rails that appear as a secondary motion contributor are superimposed on top of the horses and the jockeys who are the main actors.

An experiment is conducted to detect the speed of traffic from video information. Dynamic images from traffic videos are generated and they are capable of distinguishing the differences in motion speed. The heat maps highlighted the higher speed area.

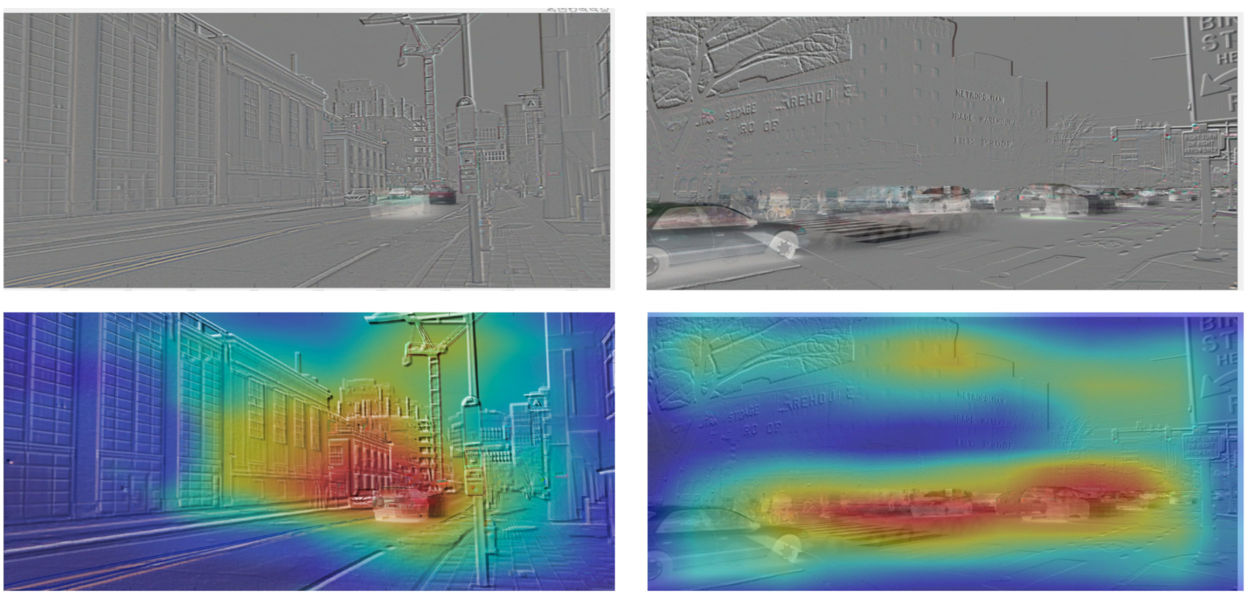


Figure 49. Top-left: dynamic image of Vassar St’s video; Bottom-left: heat map of Vassar St’s traffic; Top-right: dynamic image of Mass Ave’s video; Bottom-right: heat map of Mass Ave’s traffic, Charles Wu (2022)

3.6 Correlations between the Built Context and the User Behavior

Video data was useful to quantify the built environments and their users by using state-of-the-art data-driven machine learning models in order to develop a computational understanding of space and user. The last part of the analysis, hence, is to discover any correlation between the properties of the built environments and their user behavior.

The spatial properties and user data of 23 street cafes were quantified and plotted on scatter graphs. Certain spatial properties were found impactful to user behaviors, whereas some were not.

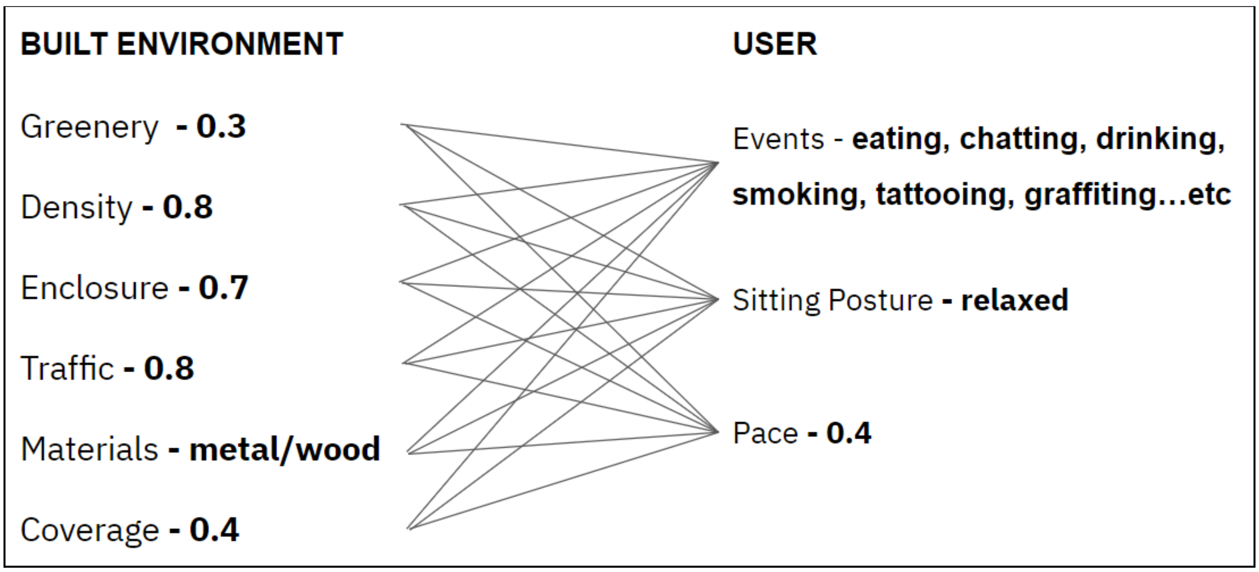


Figure 50. Examples of correlation between the built environments and their user behaviors, Charles Wu (2022)

The correlation between the traffic and the relaxedness of sitting disposition is visualized in the graph below.(Fig. 51) It is shown that heavy traffic decreases the level of relaxedness of sitting disposition, and vice versa. The traffic and the relaxedness of sitting disposition are inversely proportional.

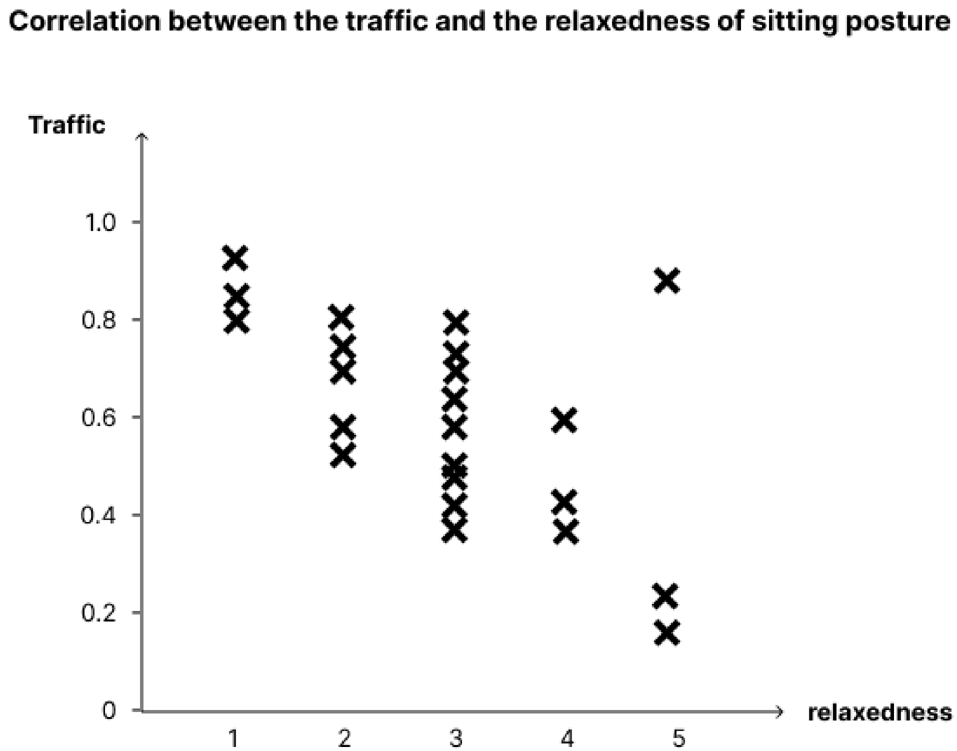


Figure 51. Correlation between the traffic and the relaxedness of the sitting disposition, Charles Wu (2022)

The correlation between the greenery and the relaxedness of the sitting disposition is visualized in the graph below.(Fig. 52) It is shown that higher greenery invites a higher level of relaxedness in sitting disposition, and vice versa. The greenery and the relaxedness of the sitting disposition are directly proportional.

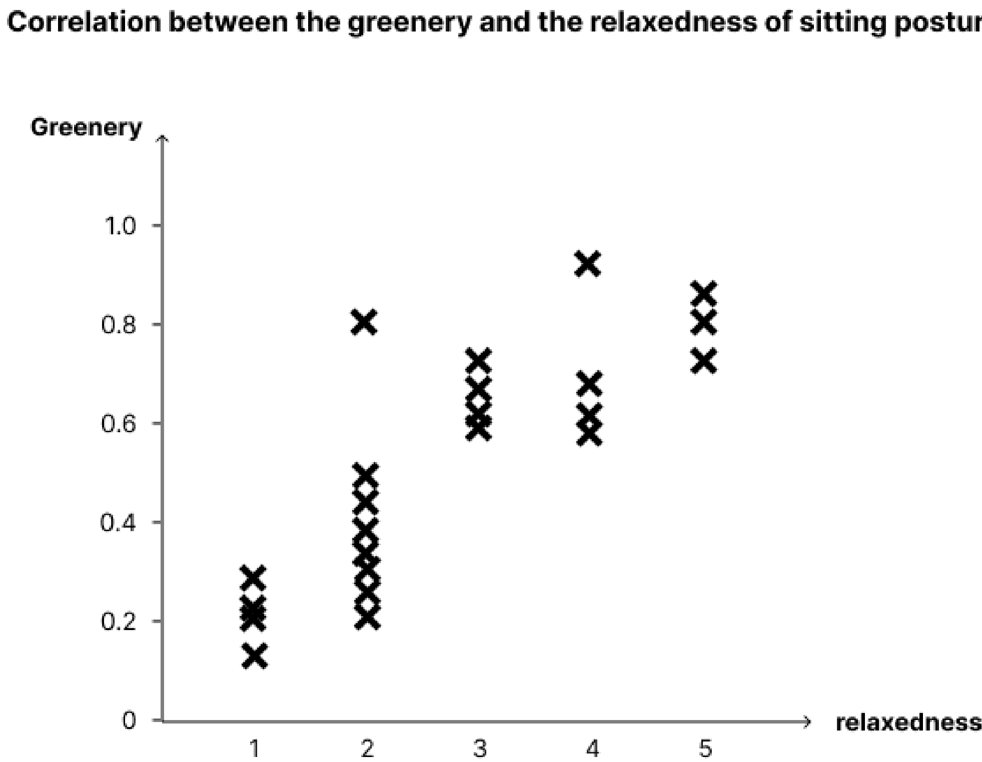


Figure 52. Correlation between the greenery and the relaxedness of the sitting disposition, Charles Wu (2022)

The correlation between the density and the relaxedness of the sitting disposition is visualized in the graph below. (Fig. 53) It is shown that dense space decreases the level of relaxedness of sitting disposition and vice versa. The density and the relaxedness of the sitting disposition are inversely proportional.

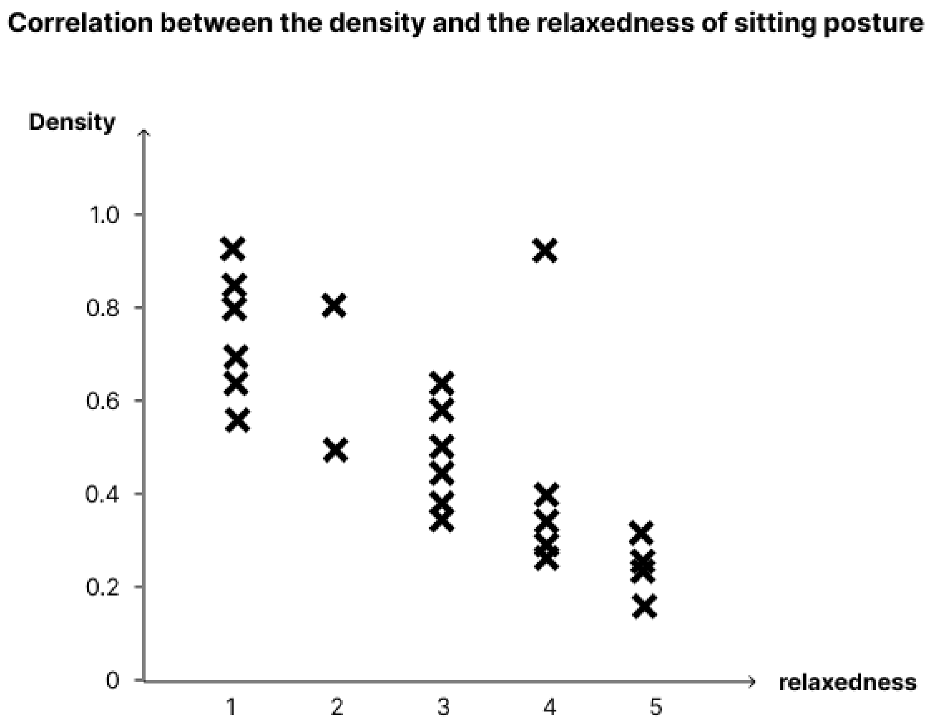


Figure 53. Correlation between the density and the relaxedness of sitting disposition, Charles Wu (2022)

The correlation between the traffic and the coverage on typing is visualized in the graph below.(Fig. 54) It indicates that the typing action usually happens in space with higher coverage or shade. Users tend to work under shelters rather than a completely open area. However, the typing action occurs no matter how heavy the traffic is. Users can work in street cafes either with busy street or with quiet path.

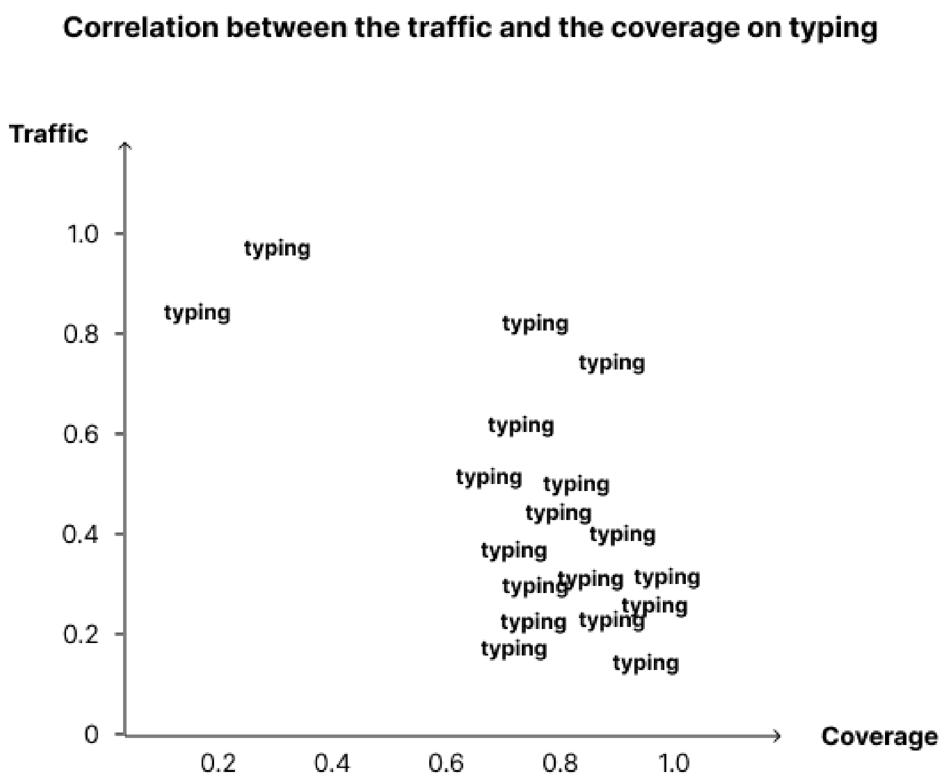


Figure 54. Correlation between the traffic and the coverage on typing, Charles Wu (2022)

The correlation between the density and the greenery on typing is visualized in the graph below.(Fig. 55) The typing actions happen at different levels of greenery and density. The scattered points proved no correlation between the greenery and the density on typing.

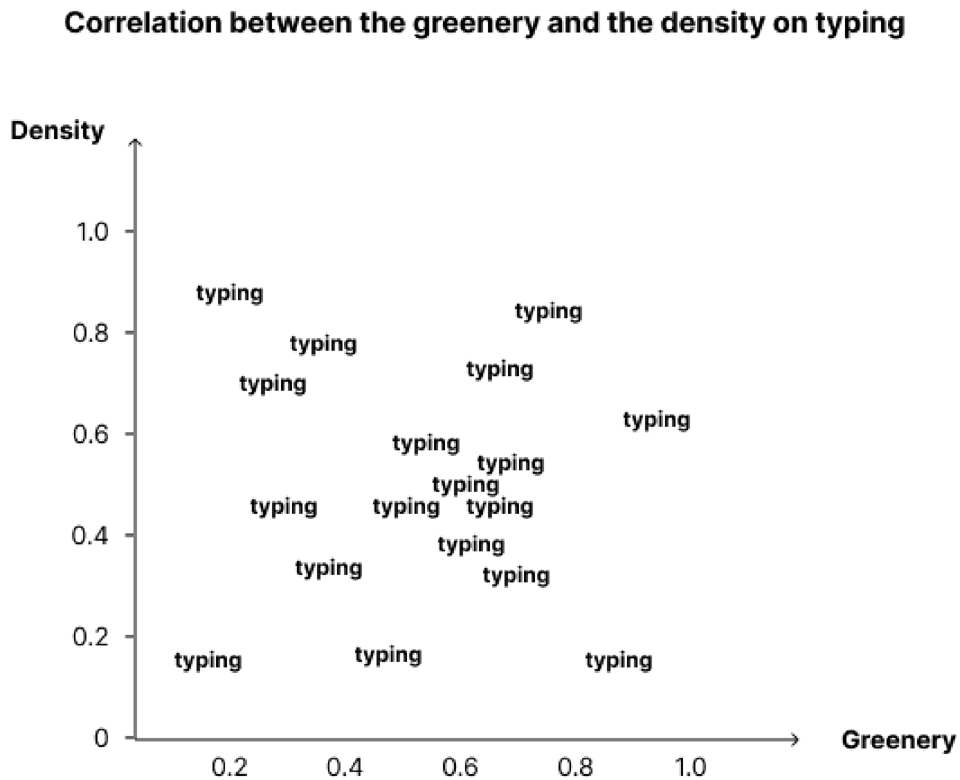


Figure 55. Correlation between the greenery and the density on typing, Charles Wu (2022)

In conclusion, there are a few interesting correlations discovered:

1. More greenery invites a higher level of relaxedness in sitting disposition
2. Users tend to work with their laptops in outdoor spaces with more shade
3. Less busy traffic invites a higher level of relaxedness of sitting disposition
4. Dense space invites a lower level of relaxedness of sitting disposition

4 | CONCLUSION

4.1 Summary and Results

Understanding the relationship between the built environments and their use is essential to architectural design. This urgent task encouraged architects and researchers, for instance, Christopher Alexander, and Herman Hertzberger, to investigate the relationships between space and users in the late 70s. Hertzberger believed that the core function of a building does not provide the total solution to space usage: it is a framework that should enable its users to interpret and define how they inhabit it. Lin Mij Textile Workshop, designed by Hertzberger, offers flexible 'in between' spaces that encourage our deeper human needs of dwelling and social activity. Alexander published *A Pattern Language* which concluded 256 design patterns for everyday users to become designers for their own living patterns. He argued that architecture should be called a 'living structure' as humans are the most essential element of architecture. However, both Herzberger and Alexander's approaches are relatively top-down interpretations based on their own empirical experience of lives. It inspired the research question of this thesis - Can machines provide a relatively bottom-up approach to understanding the relationship between the built environments and their users? A systemic computational framework to study space and its users was developed.

An experiment was carried out to investigate pattern 88 - street cafe from one of Alexander's patterns. It inquired if bottom-up observations can provide meaningful insights into spatial design. It also argued that video-based data facilitates the analysis of user behaviors in the built environments in order to provide insights into correlation.

By visiting 23 cafes around Cambridge, Massachusetts, both the spatial settings and their user activities were recorded as video data for analysis. The analysis was mainly divided into three parts: (1) user behavior, (2) spatial setting, and (3) correlation between space and user.

User Behavior

To explore the possible use of video data for user behavior analysis, the Moments In Time dataset, by Aude Oliva, was employed to recognize user actions in videos recorded. A variety of actions were successfully recognized such as chewing, stirring, biting, dipping, typing, speaking…etc. These low-level actions can be further categorized into higher-level events. For instance, dining includes chewing, stirring, dipping, and speaking; working includes typing, discussing, interviewing, and writing. Apart from the categorization of actions, the sequential analysis of action is also important to develop a computational understanding of user behaviors. However, recognizing the actions was far from satisfying. According to Alexander's description, people 'sit lazily' in street cafes. When Alexander, as a human, is able to recognize people 'sitting lazily', machines can possibly recognize it as well. A custom-trained machine learning model was proposed to predict the relaxedness of users by analyzing their sitting disposition recorded in videos. 864 frames of recorded videos were extracted as a dataset for a CNN prediction model. The final prediction accuracy was around 0.73 which was promising.

Built Environment

Based on what Alexander described as a street cafe, it should "spring up in each neighborhood", an "intimate space", and "open to a busy path". For the built environment analysis of video data, the Places pre-trained model, by Bolei Zhou, was employed to recognize the scene attributes of space from video recorded. Attributes such as outdoor, natural light, enclosed area, metal, and no horizon were useful for spatial setting analysis. In addition, object recognition using YOLOv2, by Joseph Redmon, was used to recognize the number of people in street cafe scenes in order to identify the density of space. In terms of traffic, the technique of dynamic image, by Hakan Bilen, was used for analyzing the speed of cars on streets.

Correlation between Space and User

Video data was useful to quantify the built environments and their users by using state-of-the-art data-driven machine learning models in order to develop a computational understanding of space and user. The third part of the analysis, hence, is to discover any correlation between the properties of the built environments and their user behavior. There are a few interesting correlations discovered:

1. More greenery invites a higher level of relaxedness in sitting disposition
2. Users tend to work with their laptops in outdoor spaces with more shade
3. Less busy traffic invites a higher level of relaxedness of sitting disposition
4. Dense space invites a lower level of relaxedness of sitting disposition

4.2 Conclusion

A series of tests and experiments proved that video-based on-location data facilitates the analysis of user behaviors in built environments and provides valuable insights into correlation between the behaviors and spatial design. A framework of a relatively bottom-up approach was proposed to discover unrecognized patterns using data-driven methods.

Revisiting the research question mentioned in the introduction, “Can machines provide a relatively bottom-up approach to understanding the relationship between the built environments and their users?” The result of the investigation suggests that it is possible to allow machines to ‘see’ the space as well as the users of it, like the way human architects do. Machines are capable of “recognizing” certain situations in architecture by learning from a relatively large dataset. But the difficult part still is allowing them further to “recognize” the correlations between data, much like the process of thinking as human brains do. Machines can help to provide insights into design using bottom-up data-driven methods, but its role is yet far from replacing architects or humans in the architectural design process.

4.3 Contributions

The contributions of this thesis are:

- A computational methodology of recognizing the built environment using video data. Architectural designers care about certain factors of the built environment such as openness, light, material, and form. However, many of them are perceived visually when we visited the space. They usually are difficult to be recorded through conventional architectural drawings such as plans, sections, and elevations. Hence, the computational recognition of space is essential for machines to perceive as humans do.
- A computational methodology of recognizing user behavior using video data. User behavior can hardly be recognized through images as actions are time-related. Video data demonstrated the advantage of the time dimension over image data when studying user behavior. The sequential analysis and the rate of recurrence are important quantitative measures of behavior in space.
- A methodology of video-driven computational analysis of architecture. After studying *A Pattern Language*, by Christopher Alexander, a framework that compiled several state-of-the-art machine learning models was proposed to discover new patterns in space and users. This relatively bottom-up approach provides deeper insights into users' spatial demands for designers' reference. Also, the pipeline of observation can potentially be automated to allow machines to 'see' the space and user.
- A custom-trained model for detecting the level of relaxedness of sitting disposition was created. The ultimate goal of computer vision is to allow machines 'see' as humans do. When humans can predict sentiment by observing body language, it is possible to allow machines to perceive the same.

	User Analysis	**Spatial Analysis**
Moments in Time - Aude Oliva	Action Recognition	-
Dynamic Images - Hakan Bilen	Pace Analysis	Traffic Speed Analysis
YOLOv2 - Joseph Redmon	-	Object Recognition
Places - Bolei Zhou	-	Scene Attribute Prediction
Sitting Disposition Dataset - Charles Wu	Relaxedness Analysis	-

Table 2. Existing Computer Vision Techniques Employed in this thesis, Charles Wu (2022)

4.4 Limitations

There are several limitations in this methodology:

- Low accuracy of prediction in certain situations:
    1. There are overlappings of subjects in recordings. It is difficult for the machines to recognize either objects or actions when overlappings happened. Hence, the camera positions were essential for collecting data. An alternative way to avoid overlapping is to walk around the space and record it in different angles.
    2. Cannot record every single person in the street cafe at the same time. The Moments In Time model can only process video data with no more than two persons in one scene. It increases the difficulty of collecting data as groups have to be filmed separately, otherwise, the complicated motions of a group of people messed up the predictions. To solve the problem, the 360 camera may do the job.
    3. Far away from the subjects being recorded. The closer distance between the subject and the camera, the higher accuracy of the prediction. It was tested the

optimal recording distance is around 10”. However, it is not an appropriate distance for public observation.

- Ethical issue in data collection
  In the close-up recording experiment, the subject was my girlfriend who consented to this social experiment. However, it is difficult to acquire consent from every single stranger in a public space for recording.

- Certain properties of the built environments cannot be quantified “yet”, for instance, enclosure. It is challenging to determine whether the space is enclosed by fences from video data, and quantifying the level of the enclosure is not yet developed.

- Bias in machine learning models. Every single machine learning model is biased due to the bias of humans. For instance, the sitting disposition classification dataset was labeled by humans manually. However, people might have different understandings of the relaxedness of body language. Subjective interpretation interrupted the precision of predictions. Hence, the prediction results can be considered as references, but not the ground truth.
- Complex real-life situations are difficult to be classified. In the experiment in this thesis, only six spatial properties were considered, in real-life situations, however, are more complex than just six properties. It could be ten, twenty, or even more. A complete computational understanding of space requires a more powerful computational cost as humans do.

4.5 Future Work

- Due to the pandemic, people use space very differently from the past. For example, working in a living space, exercising in a garage, eating on the street…etc. It is interesting to study how it is changed in terms of spatial setting when multi-used space is encouraged and promoted. The methodology studying correlations between the built environments and their users could be used to provide insights into adapting architecture to new spatial demands.

- While the spatial demands are dynamically ever-changing, there is an increasing trend in multi-use space design. However, conventional architectures are barely adaptive to fulfill multi-spatial needs. To deal with this inevitable architectural crisis, the methodology of studying correlations between the built environments and their users could be used to provide design insights into adapting architecture to dynamic spatial demands.
- The process of discovering correlations between data can theoretically be automated when the size of the dataset is large enough. A deep neural network (Fig. 56) to find correlations between the built environments and their users can potentially be developed to recognize new patterns from underlooked big data.

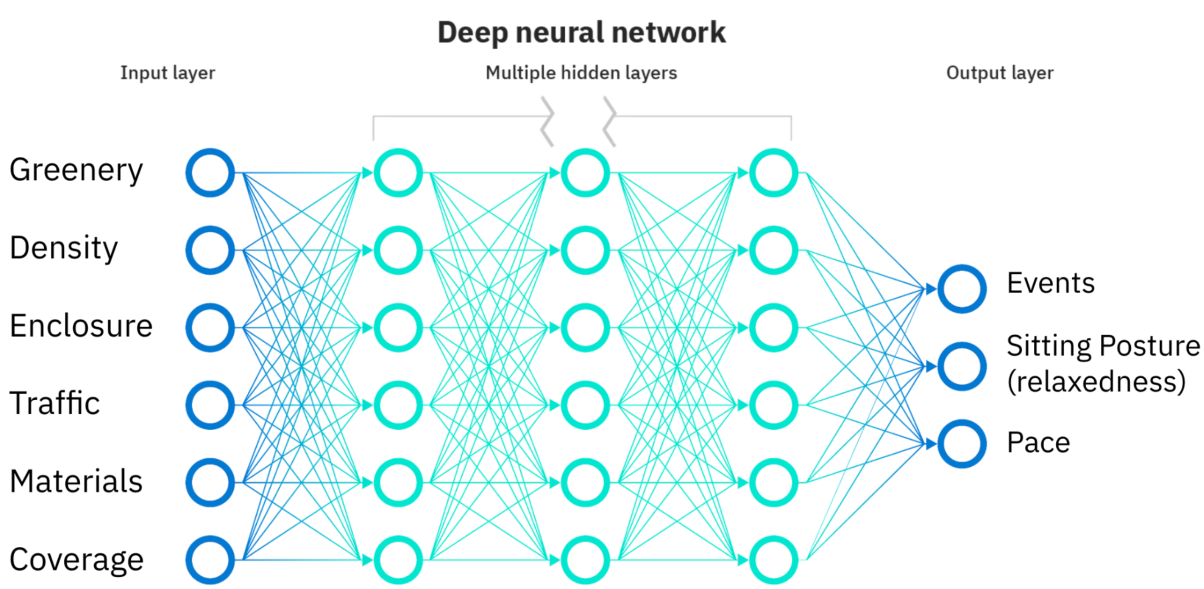


Figure 56. A deep neural network to automatically find correlations between the built environments and their users, Charles Wu (2022)

5 | BIBLIOGRAPHY

Forty, Adrian. *Words and buildings: A vocabulary of modern architecture*. Vol. 268. London: Thames & Hudson, 2000.

Hertzberger, H., ‘Flexibility and Polyvalency’, Forum, vol. 16, no. 2, February-March 1962, pp. 238-39; ‘Architecture for People’, A+U, March 1977, pp. 124-46

Swain, H., ‘Building for People’, *Journal of the Royal Institute of British Architects*, vol.68, Nov. 1961, pp. 508-10

Sullivan, Louis H. *The tall office building artistically considered*. 1922.

Alexander, Christopher. *A pattern language: towns, buildings, construction*. Oxford university press, 1977.

R. Doorley, L. Alonso, A. Grignard, N. Maciá and K. Larson, *"Travel Demand and Traffic Prediction with Cell Phone Data: Calibration by Mathematical Program with Equilibrium Constraints,"* 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), 2020, pp. 1-8, doi: 10.1109/ITSC45102.2020.9294614.

Owen, Matthew, Edwin R. Galea, and Peter J. Lawrence. “The Exodus Evacuation Model Applied To Building Evacuation Scenarios.” *Journal of Fire Protection Engineering 8*, no. 2 (May 1996): 65–84. https://doi.org/10.1177/104239159600800202.

Cheung, Kenneth Chun-Wai. *“Understanding behavior with ubiquitous computing for architectural design”*. MIT, 2007.

Komuro, N., Hashiguchi, T., Hirai, K. *et al.* Predicting individual emotion from perception-based non-contact sensor big data. *Sci Rep* 11, 2317 (2021). https://doi.org/10.1038/s41598-021-81958-2

Roy, Brandon Cain. "The birth of a word." PhD diss., Massachusetts Institute of Technology, 2013.

Monfort, Mathew, Alex Andonian, Bolei Zhou, Kandan Ramakrishnan, Sarah Adel Bargal, Tom Yan, Lisa Brown et al. "Moments in time dataset: one million videos for event understanding." *IEEE transactions on pattern analysis and machine intelligence* 42, no. 2 (2019): 502-508.

Bilen, Hakan, Basura Fernando, Efstratios Gavves, Andrea Vedaldi, and Stephen Gould. "Dynamic image networks for action recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3034-3042. 2016.

LeCun, Yann, Bernhard Boser, John S. Denker, Donnie Henderson, Richard E. Howard, Wayne Hubbard, and Lawrence D. Jackel. "Backpropagation applied to handwritten zip code recognition." *Neural computation* 1, no. 4 (1989): 541-551.

Zhou, Bolei, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. "Places: A 10 million image database for scene recognition." *IEEE transactions on pattern analysis and machine intelligence* 40, no. 6 (2017): 1452-1464.

“Image Classification” TensorFlow, last modified Jan 26, 2022, https://www.tensorflow.org/tutorials/images/classification

“OpenCV”, OpenCV Team, last modified 2022, https://opencv.org/

Biederman, Irving. "Recognition-by-components: a theory of human image understanding." *Psychological review* 94, no. 2 (1987): 115.

Qassim, Hussam, Abhishek Verma, and David Feinzimer. "Compressed residual-VGG16 CNN model for big data places image recognition." In *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 169-175. IEEE, 2018.

Iandola, Forrest N., Song Han, Matthew W. Moskewicz, Khalid Ashraf, William J. Dally, and Kurt Keutzer. "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size." *arXiv preprint arXiv:1602.07360* (2016).

Zhong, Zhuoyao, Lianwen Jin, and Zecheng Xie. "High performance offline handwritten chinese character recognition using googlenet and directional feature maps." In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 846-850. IEEE, 2015.

Wang, Limin, Sheng Guo, Weilin Huang, and Yu Qiao. "Places205-vggnet models for scene recognition." *arXiv preprint arXiv:1508.01667* (2015).

Li, Chau Yi, Ali Shahin Shamsabadi, Ricardo Sanchez-Matilla, Riccardo Mazzon, and Andrea Cavallaro. "Scene privacy protection." In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2502-2506. IEEE, 2019.

6 | APPENDIX

Appendix A - Prediction Results of User Behaviors at Street Cafes from Moments In Time

timestamp	prediction_1	acc_1	prediction_2	acc_2	prediction_3	acc_3	prediction_4	acc_4	prediction_5	acc_5
0	gripping	0.082	chewing	0.062	reading	0.049	adult+male+speaking	0.033	asking	0.032
4	chewing	0.346	eating	0.088	reading	0.059	asking	0.052	sitting	0.032
8	sneezing	0.166	coughing	0.144	reading	0.106	eating	0.102	chewing	0.051
12	studying	0.178	sneezing	0.113	reading	0.11	coughing	0.061	cramming	0.056
16	studying	0.175	asking	0.085	reading	0.067	cramming	0.054	discussing	0.053
20	reading	0.234	studying	0.118	cramming	0.084	asking	0.071	writing	0.037
24	studying	0.226	reading	0.149	cramming	0.078	writing	0.038	handwriting	0.032
28	studying	0.093	interviewing	0.063	reading	0.059	discussing	0.057	asking	0.055
32	studying	0.211	reading	0.09	drying	0.084	cramming	0.037	telephoning	0.032
36	drying	0.086	blowing	0.058	studying	0.048	discussing	0.031	reading	0.03
40	asking	0.177	studying	0.128	cramming	0.081	discussing	0.045	reading	0.042
44	asking	0.158	discussing	0.1	adult+male+speaking	0.081	sitting	0.051	interviewing	0.028
48	eating	0.096	chewing	0.079	studying	0.076	cramming	0.047	reading	0.041
52	reading	0.916	studying	0.023	tapping	0.011	cramming	0.003	asking	0.003
56	reading	0.744	studying	0.032	asking	0.013	adult+male+spe	0.01	tapping	0.009

							aking			
60	studying	0.283	cramming	0.148	asking	0.052	reading	0.026	packing	0.022
64	studying	0.161	cramming	0.111	reading	0.063	asking	0.059	adult+male+speaking	0.033
68	studying	0.06	adult+male+speaking	0.051	reading	0.048	asking	0.045	discussing	0.041
72	reading	0.125	discussing	0.079	asking	0.066	studying	0.055	sitting	0.054
76	interviewing	0.263	asking	0.103	telephoning	0.089	discussing	0.071	reading	0.07
80	interviewing	0.269	asking	0.119	reading	0.112	discussing	0.069	sitting	0.058
84	reading	0.08	interviewing	0.078	discussing	0.063	smoking	0.058	asking	0.047
88	whistling	0.06	adult+male+speaking	0.059	discussing	0.05	asking	0.045	smoking	0.03
92	adult+male+speaking	0.046	studying	0.04	drying	0.034	discussing	0.028	hanging	0.025
96	reading	0.052	discussing	0.048	adult+male+speaking	0.044	sitting	0.03	telephoning	0.027
100	discussing	0.051	adult+male+speaking	0.05	tapping	0.037	whistling	0.031	drying	0.031
104	adult+male+speaking	0.056	studying	0.044	discussing	0.04	sitting	0.038	asking	0.037
108	gripping	0.231	reading	0.16	studying	0.035	adult+male+speaking	0.027	inflating	0.015
112	reading	0.216	studying	0.177	cramming	0.046	asking	0.028	writing	0.027
116	readin	0.40	studying	0.12	crammin	0.04	discussi	0.02	asking	0.019

	g	9		6	g	8	ng			
120	reading	0.353	studying	0.284	cramming	0.061	writing	0.032	telephoning	0.02
124	asking	0.116	discussing	0.082	adult+male+speaking	0.081	interviewing	0.063	adult+female+speaking	0.038
128	adult+male+speaking	0.042	assembling	0.036	discussing	0.036	building	0.033	inflating	0.03
132	reading	0.059	discussing	0.058	working	0.055	lecturing	0.041	sitting	0.035
136	typing	0.278	writing	0.093	reading	0.067	telephoning	0.061	studying	0.061
140	reading	0.11	asking	0.08	telephoning	0.058	adult+female+speaking	0.053	studying	0.053
144	frying	0.073	stirring	0.065	flipping	0.042	fishing	0.034	barbecuing	0.025
148	chewing	0.139	smoking	0.112	licking	0.062	interviewing	0.05	eating	0.045
152	chewing	0.073	brushing	0.039	gripping	0.032	discussing	0.026	working	0.02
156	drying	0.093	adult+female+speaking	0.048	chewing	0.041	asking	0.029	rinsing	0.028
160	chewing	0.115	combing	0.052	brushing	0.046	discussing	0.036	drying	0.028
164	unpacking	0.11	packing	0.09	gripping	0.03	reading	0.025	working	0.024
168	smoking	0.08	chewing	0.054	gripping	0.05	working	0.041	building	0.039
172	drying	0.078	building	0.074	discussing	0.041	working	0.031	smoking	0.029
176	discussing	0.082	camping	0.031	smoking	0.024	working	0.02	adult+male+speaking	0.018
180	brushing	0.043	cleaning	0.043	chewing	0.038	tying	0.036	vacuuming	0.034

184	working	0.038	flooding	0.037	discussing	0.035	drying	0.031	smoking	0.023
188	working	0.064	autographing	0.046	discussing	0.039	chewing	0.037	interviewing	0.036
192	asking	0.131	interviewing	0.103	discussing	0.089	lecturing	0.053	reading	0.053
196	discussing	0.092	chewing	0.082	adult+male+speaking	0.058	interviewing	0.05	asking	0.046
200	discussing	0.134	interviewing	0.122	adult+female+speaking	0.092	asking	0.071	chewing	0.052
204	discussing	0.051	telephoning	0.05	smoking	0.034	calling	0.033	typing	0.032
208	smoking	0.167	telephoning	0.041	chewing	0.04	discussing	0.029	tapping	0.024
212	smoking	0.262	tapping	0.057	chewing	0.031	injecting	0.028	telephoning	0.025
216	tapping	0.097	smoking	0.07	drying	0.052	telephoning	0.047	brushing	0.045
220	smoking	0.467	telephoning	0.058	tapping	0.056	calling	0.035	injecting	0.031
224	smoking	0.155	telephoning	0.127	discussing	0.072	chewing	0.062	asking	0.052
228	smoking	0.1	discussing	0.093	asking	0.075	adult+male+speaking	0.07	interviewing	0.061
232	smoking	0.236	telephoning	0.075	chewing	0.063	discussing	0.046	interviewing	0.042
236	discussing	0.123	telephoning	0.09	interviewing	0.087	adult+female+speaking	0.085	smoking	0.059
240	telephoning	0.104	smoking	0.099	discussing	0.086	interviewing	0.071	tapping	0.048
244	discussing	0.158	asking	0.138	interviewing	0.123	telephoning	0.042	smoking	0.041
248	interviewing	0.247	discussing	0.194	asking	0.103	smoking	0.061	adult+female+speaking	0.04

252	tapping	0.195	interviewing	0.083	discussing	0.072	smoking	0.046	typing	0.045
256	smoking	0.104	interviewing	0.1	discussing	0.095	chewing	0.095	adult+female+speaking	0.048
260	discussing	0.116	interviewing	0.088	smoking	0.072	asking	0.071	adult+female+speaking	0.067
264	discussing	0.16	interviewing	0.159	asking	0.072	telephoning	0.062	smoking	0.048
268	discussing	0.064	smoking	0.057	telephoning	0.055	tapping	0.031	asking	0.03
272	smoking	0.124	discussing	0.106	interviewing	0.101	adult+female+speaking	0.044	chewing	0.043
276	chewing	0.088	discussing	0.088	interviewing	0.031	adult+female+speaking	0.029	adult+male+speaking	0.029
280	smoking	0.164	chewing	0.122	discussing	0.115	interviewing	0.062	adult+female+speaking	0.047
284	smoking	0.251	drying	0.046	interviewing	0.022	chewing	0.018	telephoning	0.016
288	smoking	0.233	discussing	0.048	interviewing	0.044	telephoning	0.032	adult+female+speaking	0.031
292	smoking	0.081	interviewing	0.063	discussing	0.052	asking	0.031	telephoning	0.029
296	smoking	0.214	interviewing	0.113	telephoning	0.056	discussing	0.044	asking	0.041
300	interviewing	0.119	discussing	0.091	asking	0.055	telephoning	0.05	adult+female+speaking	0.046
304	smoking	0.1	interviewing	0.07	discussing	0.055	telephoning	0.047	asking	0.042
308	smoking	0.103	drying	0.066	interviewing	0.056	discussing	0.055	adult+female+speaking	0.039
312	smoking	0.319	interviewing	0.05	asking	0.041	adult+female+sp	0.04	discussing	0.036

							eaking			
316	smoking	0.107	interviewing	0.087	discussing	0.057	telephoning	0.049	asking	0.033
320	smoking	0.133	interviewing	0.082	adult+female+speaking	0.054	discussing	0.049	flooding	0.036
324	chewing	0.137	coughing	0.091	sneezing	0.084	eating	0.068	discussing	0.062
328	smoking	0.249	chewing	0.102	discussing	0.076	interviewing	0.053	eating	0.048
332	discussing	0.133	telephoning	0.096	interviewing	0.085	adult+female+speaking	0.07	asking	0.045
336	discussing	0.144	interviewing	0.079	telephoning	0.061	adult+female+speaking	0.057	asking	0.032
340	telephoning	0.12	discussing	0.078	smoking	0.077	interviewing	0.069	adult+female+speaking	0.047
344	chewing	0.133	discussing	0.117	adult+female+speaking	0.109	asking	0.061	telephoning	0.057
348	smoking	0.166	discussing	0.087	telephoning	0.069	interviewing	0.06	sitting	0.049
352	interviewing	0.132	discussing	0.124	asking	0.079	sitting	0.05	smoking	0.045
356	smoking	0.118	chewing	0.073	discussing	0.063	asking	0.059	telephoning	0.043
360	discussing	0.086	asking	0.053	interviewing	0.047	adult+male+speaking	0.043	chewing	0.039
364	asking	0.096	signing	0.086	interviewing	0.071	smoking	0.061	chewing	0.052
368	discussing	0.215	interviewing	0.145	adult+female+speaking	0.085	sitting	0.082	asking	0.047
372	discussing	0.102	asking	0.089	telephoning	0.087	interviewing	0.062	typing	0.062
376	teleph	0.17	typing	0.15	calling	0.07	discussi	0.05	paying	0.026

	oning	8		7		1	ng	8		
380	telephoning	0.152	typing	0.114	calling	0.068	discussing	0.035	paying	0.031
384	telephoning	0.096	discussing	0.095	sitting	0.08	adult+female+speaking	0.06	asking	0.058
388	telephoning	0.187	interviewing	0.124	typing	0.113	discussing	0.1	calling	0.05
392	typing	0.306	telephoning	0.166	calling	0.05	discussing	0.044	tapping	0.031
396	interviewing	0.092	telephoning	0.09	discussing	0.079	sitting	0.078	adult+female+speaking	0.057
400	telephoning	0.077	discussing	0.073	interviewing	0.067	smoking	0.055	typing	0.046
404	telephoning	0.129	discussing	0.092	interviewing	0.059	typing	0.047	calling	0.047
408	discussing	0.145	telephoning	0.119	interviewing	0.087	asking	0.08	sitting	0.058
412	discussing	0.078	asking	0.068	adult+male+speaking	0.06	adult+female+speaking	0.051	sitting	0.05
416	stretching	0.218	balancing	0.18	gripping	0.033	chewing	0.026	kneeling	0.021
420	drinking	0.747	eating	0.059	chewing	0.059	licking	0.007	sniffing	0.006
424	discussing	0.107	chewing	0.056	interviewing	0.046	adult+male+speaking	0.041	tapping	0.038
428	discussing	0.099	interviewing	0.08	sitting	0.03	adult+female+speaking	0.027	adult+male+speaking	0.026
432	drinking	0.155	asking	0.07	discussing	0.067	eating	0.06	interviewing	0.054
436	eating	0.085	drinking	0.081	discussing	0.079	chewing	0.05	smoking	0.044
440	drying	0.075	discussing	0.075	chewing	0.057	interviewing	0.044	combing	0.028

444	discussing	0.126	interviewing	0.092	adult+male+speaking	0.03	socializing	0.029	inflating	0.026
448	chewing	0.114	eating	0.056	biting	0.053	discussing	0.05	inflating	0.036
452	chewing	0.212	tapping	0.078	biting	0.031	licking	0.027	drinking	0.026
456	eating	0.09	chewing	0.087	tattooing	0.044	biting	0.042	feeding	0.038
460	tattooing	0.047	inflating	0.044	chewing	0.042	drinking	0.037	fishing	0.031
464	tattooing	0.144	inflating	0.037	chewing	0.035	drying	0.035	discussing	0.026
468	tattooing	0.129	inflating	0.047	chewing	0.046	feeding	0.027	drying	0.026
472	discussing	0.05	eating	0.046	chewing	0.042	drinking	0.033	interviewing	0.025
476	chewing	0.097	discussing	0.083	adult+female+speaking	0.041	interviewing	0.034	sitting	0.031
480	chewing	0.087	discussing	0.064	adult+female+speaking	0.03	adult+male+speaking	0.028	gripping	0.025
484	chewing	0.29	discussing	0.095	interviewing	0.054	eating	0.047	adult+female+speaking	0.044
488	discussing	0.099	chewing	0.043	adult+female+speaking	0.029	interviewing	0.029	adult+male+speaking	0.024
492	chewing	0.069	discussing	0.065	interviewing	0.039	drying	0.027	biting	0.027
496	discussing	0.202	interviewing	0.104	drying	0.06	adult+female+speaking	0.04	chewing	0.035
500	discussing	0.132	interviewing	0.103	adult+female+speaking	0.08	asking	0.044	sitting	0.032
504	discussing	0.238	interviewing	0.17	adult+female+speaking	0.082	asking	0.046	socializing	0.023

508	chewing	0.141	eating	0.072	biting	0.049	smoking	0.036	coughing	0.033
512	drinking	0.182	pouring	0.065	chewing	0.044	dipping	0.043	eating	0.04
516	pouring	0.072	emptying	0.064	filling	0.052	drinking	0.049	drying	0.036
520	drying	0.13	chewing	0.086	tapping	0.038	drinking	0.035	pouring	0.035
524	camping	0.117	discussing	0.029	drinking	0.026	chewing	0.024	gripping	0.019
528	chewing	0.063	hunting	0.031	smoking	0.026	gripping	0.026	playing+videogames	0.025
532	chewing	0.029	eating	0.026	gripping	0.026	adult+male+speaking	0.024	discussing	0.022
536	discussing	0.044	balancing	0.038	adult+male+speaking	0.036	whistling	0.03	asking	0.029
540	adult+female+speaking	0.238	asking	0.094	discussing	0.078	interviewing	0.049	adult+male+speaking	0.035
544	discussing	0.126	asking	0.114	adult+male+speaking	0.11	interviewing	0.074	smoking	0.053
548	chewing	0.114	smoking	0.046	discussing	0.041	drinking	0.037	adult+male+speaking	0.036
552	chewing	0.076	frowning	0.061	drinking	0.051	dining	0.044	blowing	0.038
556	chewing	0.065	adult+male+speaking	0.041	discussing	0.04	frowning	0.038	drinking	0.034
560	drinking	0.193	dining	0.117	eating	0.088	chewing	0.059	discussing	0.056
564	drying	0.161	discussing	0.035	chewing	0.029	reading	0.027	adult+female+speaking	0.023
568	drying	0.06	chewing	0.03	discussi	0.03	adult+fe	0.02	reading	0.019

		1		7	ng	2	male+speaking			
572	discussing	0.072	drying	0.06	chewing	0.037	interviewing	0.034	adult+female+speaking	0.032
576	drying	0.088	chewing	0.081	discussing	0.053	adult+female+speaking	0.026	adult+male+speaking	0.026
580	chewing	0.06	discussing	0.058	adult+female+speaking	0.037	asking	0.031	adult+male+speaking	0.03
584	discussing	0.06	sitting	0.046	chewing	0.042	drying	0.04	adult+female+speaking	0.035
588	eating	0.092	chewing	0.087	dining	0.052	asking	0.029	sitting	0.029
592	eating	0.133	discussing	0.075	drinking	0.068	chewing	0.058	sitting	0.057
596	drinking	0.3	chewing	0.128	eating	0.084	dining	0.052	gripping	0.02
600	discussing	0.097	asking	0.084	interviewing	0.065	adult+female+speaking	0.06	smoking	0.051
604	discussing	0.109	adult+female+speaking	0.048	drying	0.041	interviewing	0.034	asking	0.034
608	discussing	0.091	drying	0.06	adult+female+speaking	0.044	gripping	0.039	interviewing	0.036
612	discussing	0.068	drying	0.048	chewing	0.041	gripping	0.03	adult+female+speaking	0.029
616	discussing	0.138	sitting	0.043	adult+female+speaking	0.037	adult+male+speaking	0.035	interviewing	0.033
620	discussing	0.088	adult+male+speaking	0.046	chewing	0.042	inflating	0.03	adult+female+speaking	0.03
624	gripping	0.089	discussing	0.062	inflating	0.032	drying	0.032	adult+male+speaking	0.028

628	discussing	0.079	sitting	0.048	dining	0.046	gripping	0.046	chewing	0.037
632	discussing	0.077	gripping	0.041	dining	0.039	smoking	0.034	sitting	0.034
636	gripping	0.071	discussing	0.057	chewing	0.041	smoking	0.035	drying	0.025
640	inflating	0.054	gripping	0.051	discussing	0.051	chewing	0.037	dining	0.033
644	discussing	0.077	inflating	0.046	drying	0.041	chewing	0.037	adult+male+speaking	0.033
648	eating	0.924	chewing	0.037	dining	0.032	biting	0.004	feeding	0.003
652	eating	0.45	dining	0.44	chewing	0.048	serving	0.014	feeding	0.008
656	eating	0.695	chewing	0.25	dining	0.039	biting	0.008	feeding	0.004
660	eating	0.785	dining	0.137	chewing	0.047	feeding	0.015	biting	0.005
664	eating	0.45	dining	0.396	chewing	0.047	feeding	0.016	serving	0.013
668	eating	0.677	dining	0.152	chewing	0.079	feeding	0.019	biting	0.013
672	eating	0.968	dining	0.01	biting	0.008	chewing	0.007	feeding	0.005
676	eating	0.944	chewing	0.026	dining	0.013	feeding	0.008	biting	0.007
680	eating	0.909	dining	0.045	feeding	0.022	chewing	0.012	biting	0.007
684	eating	0.972	dining	0.012	chewing	0.01	feeding	0.004	biting	0.003
688	eating	0.82	dining	0.121	chewing	0.026	feeding	0.01	biting	0.009
692	eating	0.736	chewing	0.105	dining	0.067	feeding	0.031	biting	0.016
696	eating	0.819	dining	0.089	biting	0.036	chewing	0.031	feeding	0.013

700	eating	0.779	dining	0.128	chewing	0.061	biting	0.023	feeding	0.003
704	eating	0.582	dining	0.285	biting	0.051	chewing	0.043	feeding	0.01
708	eating	0.921	dining	0.035	chewing	0.024	biting	0.014	feeding	0.006
712	eating	0.863	chewing	0.099	biting	0.024	dining	0.007	feeding	0.005
716	eating	0.949	dining	0.025	chewing	0.016	biting	0.007	feeding	0.002
720	eating	0.557	dining	0.185	chewing	0.131	biting	0.038	feeding	0.011
724	eating	0.844	dining	0.096	chewing	0.039	biting	0.011	feeding	0.006
728	eating	0.52	dining	0.312	chewing	0.1	biting	0.023	feeding	0.01
732	eating	0.742	dining	0.215	chewing	0.027	biting	0.007	feeding	0.004
736	eating	0.905	dining	0.044	chewing	0.02	feeding	0.012	biting	0.01
740	eating	0.792	dining	0.113	chewing	0.058	biting	0.016	feeding	0.003
744	eating	0.559	dining	0.32	chewing	0.077	biting	0.027	feeding	0.004
748	eating	0.709	dining	0.269	chewing	0.012	feeding	0.004	biting	0.002
752	eating	0.448	dining	0.446	chewing	0.063	biting	0.011	feeding	0.004
756	eating	0.932	chewing	0.033	dining	0.029	feeding	0.003	biting	0.002
760	eating	0.911	dining	0.049	chewing	0.027	biting	0.006	feeding	0.004
764	eating	0.697	dining	0.126	chewing	0.1	biting	0.038	feeding	0.006
768	eating	0.773	dining	0.121	chewing	0.069	biting	0.017	feeding	0.005
772	eating	0.84	dining	0.09	chewing	0.03	biting	0.02	feeding	0.005

		6		2		1				
776	eating	0.706	dining	0.182	chewing	0.035	biting	0.033	feeding	0.007
780	eating	0.895	chewing	0.072	dining	0.018	biting	0.009	feeding	0.006
784	eating	0.75	dining	0.094	biting	0.07	chewing	0.062	feeding	0.006
788	eating	0.905	dining	0.046	chewing	0.04	biting	0.005	feeding	0.003
792	eating	0.817	chewing	0.147	dining	0.018	biting	0.013	feeding	0.002
796	interviewing	0.237	discussing	0.113	chewing	0.066	eating	0.053	sitting	0.051
800	eating	0.335	dining	0.213	chewing	0.148	biting	0.079	feeding	0.017
804	eating	0.417	chewing	0.256	dining	0.235	biting	0.028	feeding	0.011
808	eating	0.737	dining	0.15	chewing	0.069	biting	0.005	feeding	0.004
812	eating	0.725	dining	0.126	chewing	0.067	biting	0.016	feeding	0.014
816	eating	0.537	chewing	0.192	dining	0.107	stirring	0.026	cooking	0.02
820	eating	0.798	chewing	0.091	dining	0.044	feeding	0.017	biting	0.012
824	eating	0.579	dining	0.314	chewing	0.043	feeding	0.008	biting	0.006
828	eating	0.626	dining	0.169	chewing	0.106	stirring	0.014	feeding	0.011
832	eating	0.743	chewing	0.104	dining	0.102	biting	0.014	feeding	0.011
836	chewing	0.356	eating	0.3	dining	0.11	biting	0.051	discussing	0.032
840	discussing	0.052	sitting	0.043	chewing	0.04	interviewing	0.036	dining	0.03
844	typing	0.19	tapping	0.073	playing+music	0.051	working	0.034	gripping	0.031

848	chewing	0.074	dining	0.072	eating	0.056	discussing	0.044	interviewing	0.039
852	chewing	0.054	eating	0.046	interviewing	0.045	discussing	0.042	dining	0.036
856	typing	0.202	interviewing	0.054	discussing	0.045	sitting	0.039	tapping	0.029
860	interviewing	0.057	discussing	0.055	typing	0.053	sitting	0.052	tapping	0.033
864	interviewing	0.259	discussing	0.116	dining	0.055	sitting	0.033	chewing	0.029
868	eating	0.141	dining	0.133	biting	0.063	chewing	0.054	working	0.017
872	interviewing	0.307	discussing	0.164	sitting	0.053	asking	0.024	socializing	0.023
876	tapping	0.537	typing	0.18	playing+ music	0.022	interviewing	0.016	scratching	0.016
880	tapping	0.337	typing	0.331	discussing	0.031	interviewing	0.031	sitting	0.026
884	dining	0.425	eating	0.089	chewing	0.035	discussing	0.027	interviewing	0.026
888	dining	0.151	discussing	0.089	interviewing	0.077	eating	0.065	chewing	0.054
892	interviewing	0.096	discussing	0.081	sitting	0.07	chewing	0.046	dining	0.043
896	typing	0.382	discussing	0.051	tapping	0.045	interviewing	0.038	manicuring	0.029
900	interviewing	0.266	discussing	0.151	dining	0.091	sitting	0.054	eating	0.053
904	interviewing	0.121	discussing	0.1	sitting	0.041	chewing	0.037	dining	0.023
908	dining	0.051	camping	0.046	discussing	0.039	chewing	0.038	eating	0.027
912	tapping	0.153	typing	0.056	chewing	0.028	dining	0.027	discussing	0.024
916	dining	0.218	eating	0.202	chewing	0.131	discussing	0.038	biting	0.031
920	eating	0.63	chewing	0.20	dining	0.09	biting	0.02	feeding	0.01

		1		3				8		
924	eating	0.572	chewing	0.173	dining	0.074	biting	0.037	feeding	0.031
928	eating	0.87	dining	0.067	chewing	0.026	feeding	0.015	biting	0.012
932	eating	0.559	dining	0.213	chewing	0.076	biting	0.019	cooking	0.018
936	eating	0.68	dining	0.117	chewing	0.07	biting	0.024	feeding	0.019
940	eating	0.498	dining	0.371	chewing	0.072	feeding	0.013	biting	0.008
944	eating	0.681	chewing	0.166	dining	0.122	biting	0.017	feeding	0.006
948	eating	0.446	dining	0.323	chewing	0.066	carving	0.025	dipping	0.011
952	eating	0.901	dining	0.055	chewing	0.03	feeding	0.006	biting	0.002
956	eating	0.764	feeding	0.074	dining	0.05	chewing	0.047	biting	0.013
960	eating	0.831	chewing	0.094	dining	0.046	feeding	0.008	biting	0.007
964	eating	0.307	stirring	0.11	dining	0.083	chewing	0.081	dipping	0.05
968	eating	0.695	dining	0.104	chewing	0.057	stirring	0.051	frying	0.017
972	eating	0.562	chewing	0.063	dining	0.057	stirring	0.04	dipping	0.032
976	eating	0.963	chewing	0.012	dining	0.011	feeding	0.01	biting	0.003
980	eating	0.712	chewing	0.221	dining	0.039	biting	0.017	feeding	0.005
984	eating	0.826	chewing	0.111	dining	0.026	biting	0.016	feeding	0.009
988	eating	0.673	chewing	0.23	biting	0.06	dining	0.012	feeding	0.012
992	eating	0.82	chewing	0.131	biting	0.027	dining	0.01	feeding	0.006

996	eating	0.936	chewing	0.046	biting	0.008	dining	0.007	feeding	0.004
1000	eating	0.519	dining	0.164	chewing	0.16	biting	0.028	peeling	0.022
1004	eating	0.555	dining	0.135	chewing	0.122	biting	0.024	peeling	0.023
1008	eating	0.742	chewing	0.199	biting	0.038	dining	0.008	feeding	0.007
1012	eating	0.34	chewing	0.221	dining	0.072	biting	0.051	discussing	0.02
1016	eating	0.677	chewing	0.144	dining	0.046	biting	0.026	feeding	0.015
1020	eating	0.788	chewing	0.149	dining	0.014	feeding	0.011	whistling	0.005
1024	eating	0.664	chewing	0.282	biting	0.018	dining	0.013	feeding	0.005
1028	eating	0.858	chewing	0.077	biting	0.049	feeding	0.009	dining	0.006
1032	eating	0.71	chewing	0.235	biting	0.02	dining	0.017	feeding	0.004
1036	eating	0.827	chewing	0.11	biting	0.019	dining	0.011	feeding	0.011
1040	eating	0.889	chewing	0.066	dining	0.025	biting	0.011	feeding	0.007
1044	eating	0.609	chewing	0.262	dining	0.058	biting	0.035	feeding	0.008
1048	eating	0.826	chewing	0.106	biting	0.024	dining	0.018	feeding	0.013
1052	eating	0.696	chewing	0.239	biting	0.053	feeding	0.005	dining	0.004
1056	chewing	0.395	eating	0.286	dining	0.137	biting	0.056	discussing	0.011
1060	chewing	0.612	eating	0.279	biting	0.035	dining	0.013	smoking	0.008
1064	eating	0.156	chewing	0.095	biting	0.09	adult+male+speaking	0.073	dining	0.058

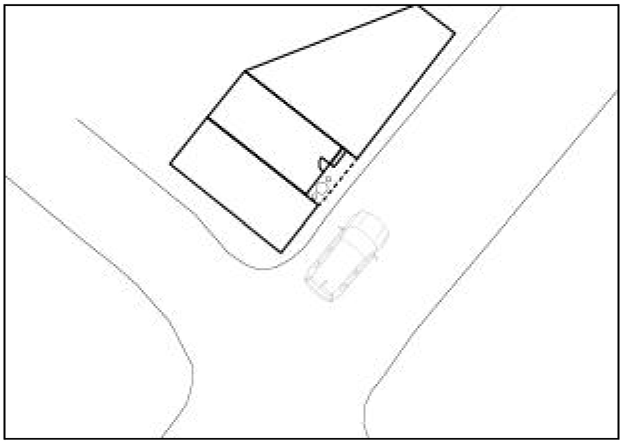
1068	eating	0.425	chewing	0.314	biting	0.067	dining	0.044	feeding	0.017
1072	chewing	0.626	eating	0.219	biting	0.088	dining	0.019	discussing	0.005
1076	chewing	0.405	biting	0.195	eating	0.13	dining	0.028	drying	0.019
1080	chewing	0.115	eating	0.072	biting	0.069	dining	0.065	interviewing	0.054
1084	eating	0.284	chewing	0.195	dining	0.135	biting	0.096	feeding	0.013
1088	eating	0.275	chewing	0.103	dining	0.09	adult+male+speaking	0.058	biting	0.052
1092	dining	0.107	eating	0.104	drinking	0.073	sitting	0.036	chewing	0.034
1096	eating	0.191	dining	0.117	chewing	0.09	biting	0.089	inflating	0.059
1100	inflating	0.05	interviewing	0.043	drying	0.042	discussing	0.04	assembling	0.033
1104	interviewing	0.168	discussing	0.068	autographing	0.055	asking	0.041	signing	0.022
1108	eating	0.525	chewing	0.214	dining	0.071	feeding	0.027	dipping	0.017
1112	eating	0.55	dining	0.186	chewing	0.1	biting	0.018	feeding	0.014
1116	eating	0.17	chewing	0.143	dining	0.062	peeling	0.055	gripping	0.038
1120	eating	0.312	chewing	0.19	dining	0.11	biting	0.039	feeding	0.016
1124	eating	0.394	chewing	0.145	peeling	0.067	dining	0.045	dipping	0.036
1128	eating	0.534	chewing	0.111	dining	0.1	feeding	0.038	dipping	0.03
1132	eating	0.563	chewing	0.125	dining	0.063	stirring	0.034	feeding	0.027
1136	eating	0.835	chewing	0.072	dining	0.03	feeding	0.028	biting	0.013

1140	eating	0.833	chewing	0.068	dining	0.035	feeding	0.03	biting	0.015
1144	eating	0.791	chewing	0.112	dining	0.036	biting	0.025	feeding	0.011
1148	eating	0.535	chewing	0.396	dining	0.031	biting	0.009	feeding	0.007
1152	eating	0.611	chewing	0.21	dining	0.089	biting	0.012	feeding	0.007
1156	eating	0.687	chewing	0.233	dining	0.024	biting	0.023	feeding	0.005
1160	eating	0.323	dining	0.2	chewing	0.109	sitting	0.041	discussing	0.029
1164	dining	0.192	eating	0.097	sitting	0.068	chewing	0.059	discussing	0.036
1168	eating	0.4	chewing	0.109	dining	0.084	biting	0.04	discussing	0.036
1172	eating	0.226	dining	0.163	asking	0.068	adult+female+speaking	0.064	chewing	0.052
1176	drinking	0.56	dining	0.106	eating	0.066	chewing	0.017	laughing	0.013
1180	eating	0.654	dining	0.096	chewing	0.064	biting	0.03	feeding	0.012
1184	chewing	0.334	eating	0.249	dining	0.075	biting	0.067	inflating	0.026
1188	carving	0.115	eating	0.092	dining	0.083	chewing	0.057	cutting	0.03
1192	inflating	0.168	camping	0.074	fishing	0.072	repairing	0.067	dining	0.049
1196	eating	0.147	dining	0.11	chewing	0.068	discussing	0.039	camping	0.035
1200	eating	0.239	chewing	0.194	biting	0.09	dining	0.085	gripping	0.02
1204	eating	0.54	chewing	0.194	dining	0.041	gripping	0.03	biting	0.018
1208	dining	0.242	eating	0.095	gripping	0.063	chewing	0.039	serving	0.026

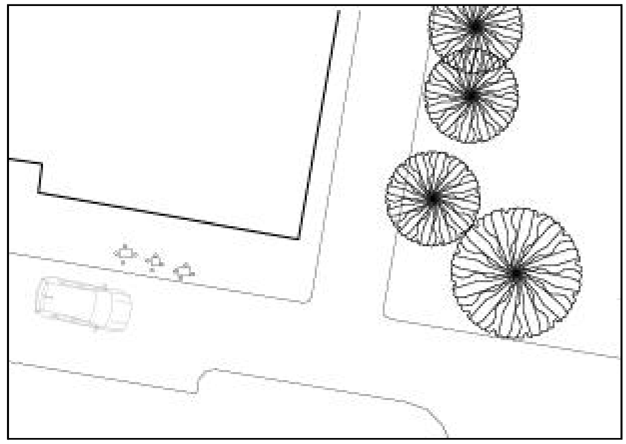
1212	eating	0.401	dining	0.089	dipping	0.062	cooking	0.058	drinking	0.047
1216	drinking	0.173	pouring	0.142	eating	0.121	chewing	0.056	emptying	0.037
1220	eating	0.223	dining	0.11	drinking	0.104	chewing	0.052	asking	0.025
1224	drinking	0.339	eating	0.247	dining	0.1	chewing	0.054	smelling	0.015
1228	dining	0.199	eating	0.074	pouring	0.049	serving	0.033	dipping	0.026
1232	eating	0.087	dining	0.052	drying	0.047	blowing	0.037	chewing	0.035
1236	eating	0.329	dining	0.275	chewing	0.128	biting	0.018	feeding	0.012
1240	eating	0.504	dining	0.099	chewing	0.076	biting	0.026	gripping	0.019
1244	eating	0.287	dining	0.229	chewing	0.047	biting	0.028	cooking	0.016
1248	dining	0.28	eating	0.249	cooking	0.056	feeding	0.025	barbecuing	0.02
1252	eating	0.764	dining	0.079	chewing	0.035	feeding	0.027	cooking	0.009
1256	eating	0.404	chewing	0.22	carving	0.047	dining	0.034	biting	0.033
1260	eating	0.811	chewing	0.092	dining	0.023	feeding	0.012	biting	0.008
1264	eating	0.347	dining	0.288	chewing	0.105	sitting	0.017	feeding	0.015
1268	eating	0.309	dining	0.15	chewing	0.118	peeling	0.051	feeding	0.024
1272	chewing	0.203	carving	0.203	eating	0.173	peeling	0.069	dining	0.057
1276	eating	0.462	chewing	0.13	peeling	0.07	dining	0.066	cooking	0.039
1280	eating	0.783	chewing	0.08	dining	0.041	feeding	0.022	biting	0.014
1284	eating	0.61	dining	0.08	carving	0.08	chewing	0.05	dipping	0.022

		8		8		4		7		
1288	eating	0.57	chewing	0.142	dining	0.045	feeding	0.032	biting	0.016
1292	eating	0.666	chewing	0.247	biting	0.019	dining	0.012	feeding	0.011

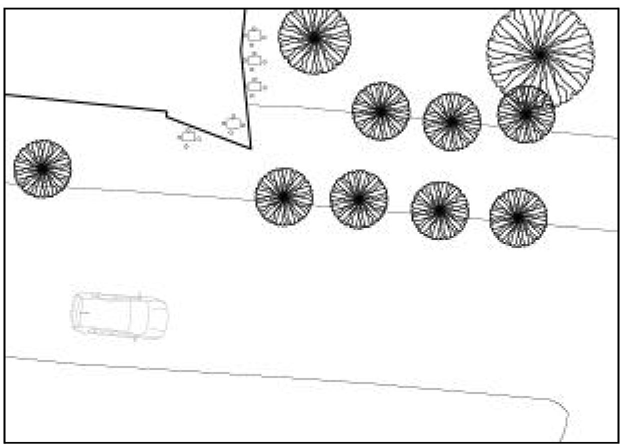
Appendix B - Floor Plan of 23 Street Cafes



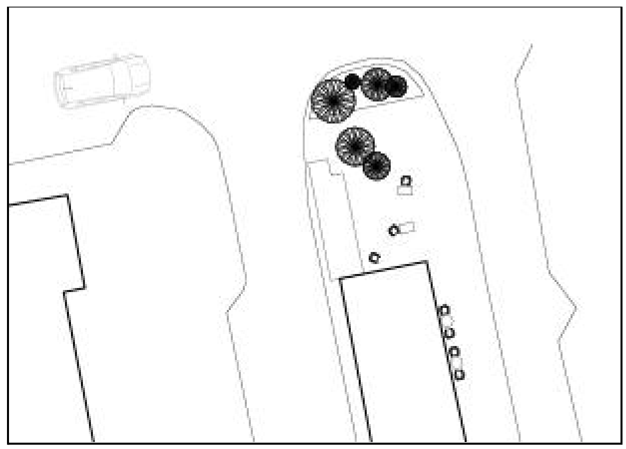
Cafe 1



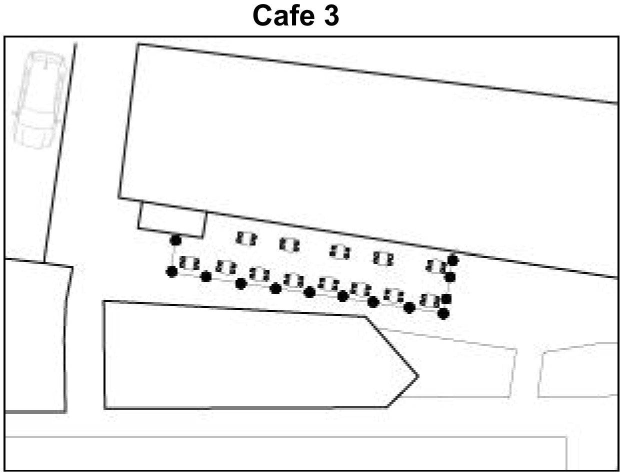
Cafe 2



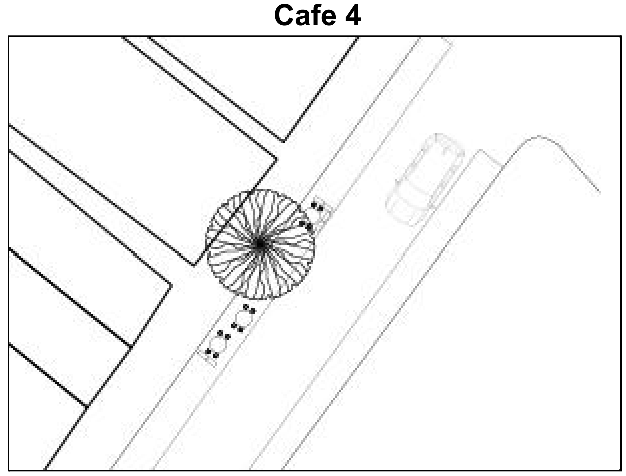
Cafe 3



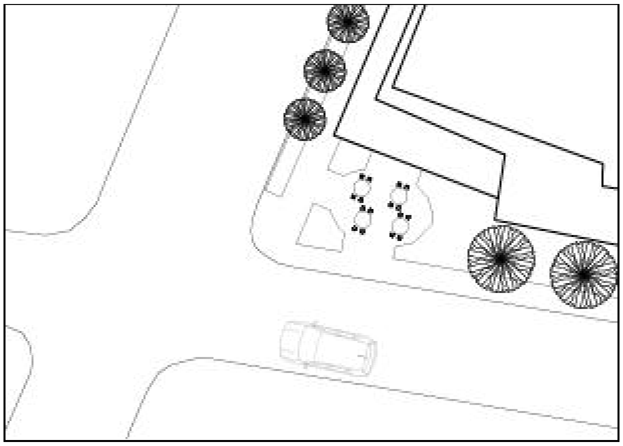
Cafe 4



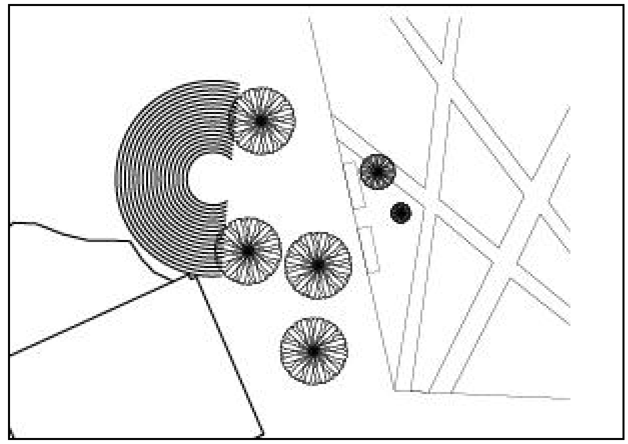
Cafe 5



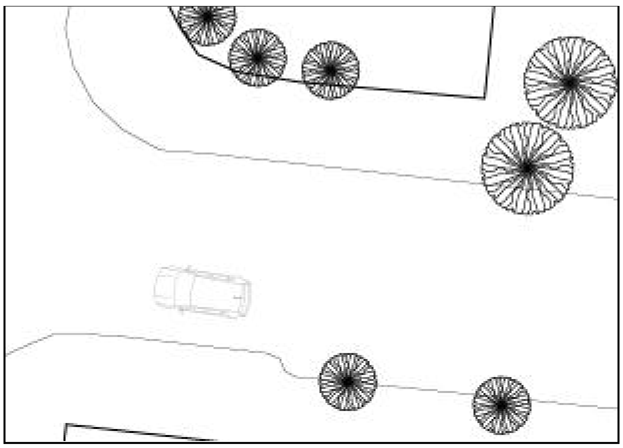
Cafe 6



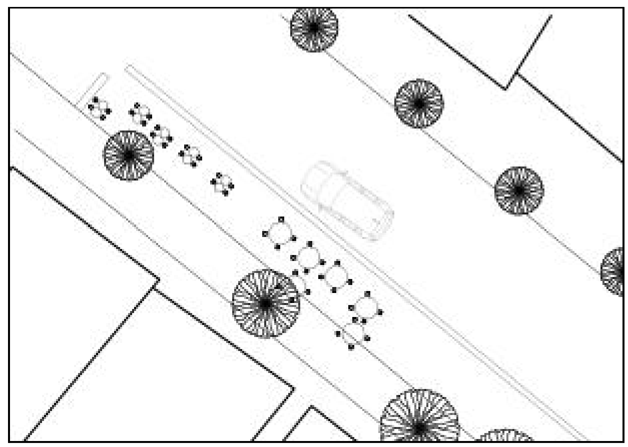
Cafe 7



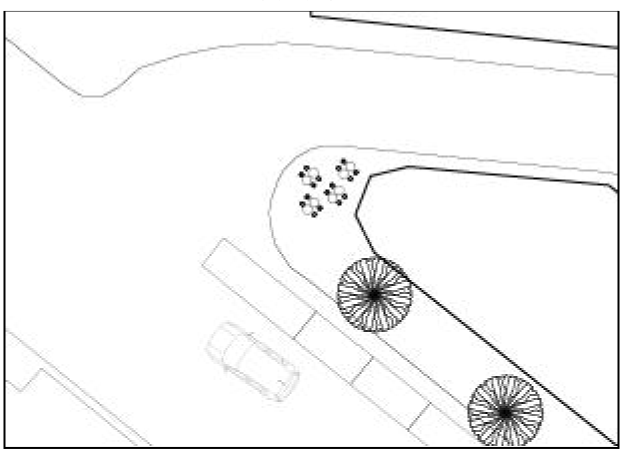
Cafe 8



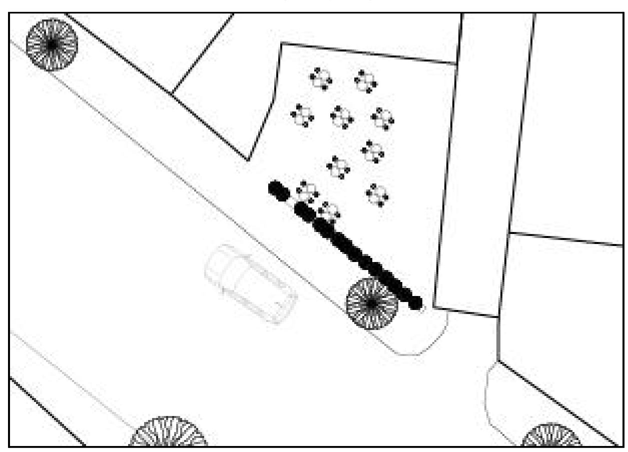
Cafe 9



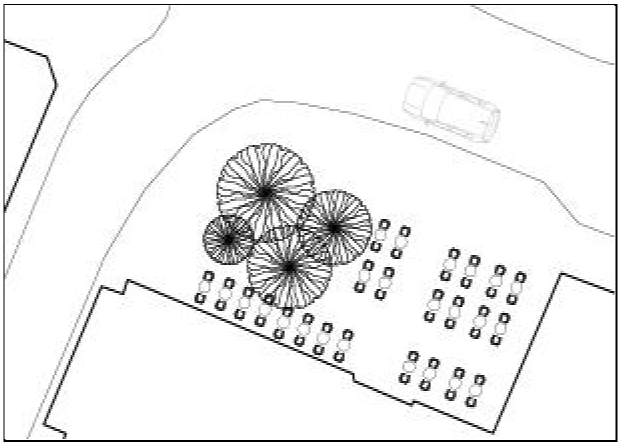
Cafe 10



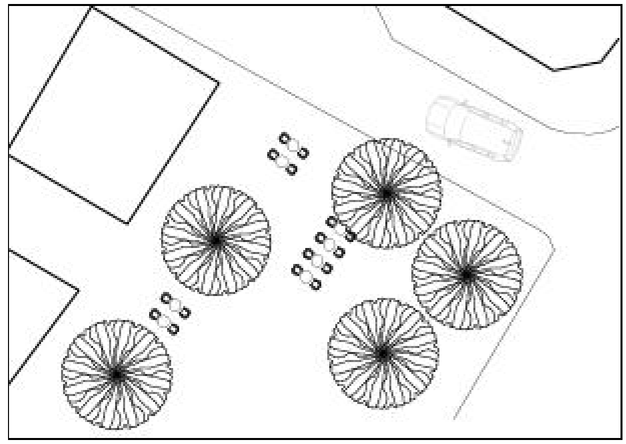
Cafe 11



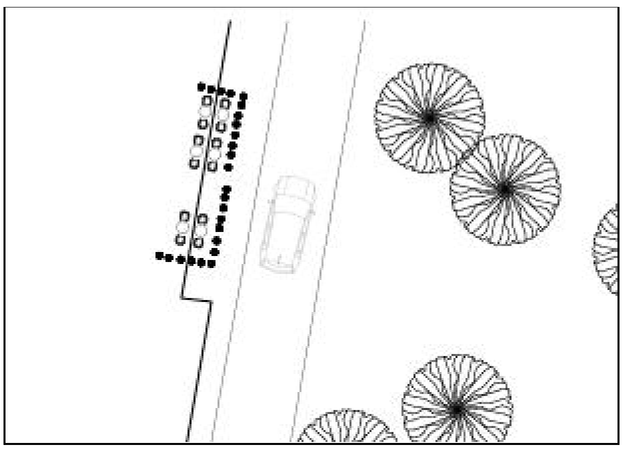
Cafe 12



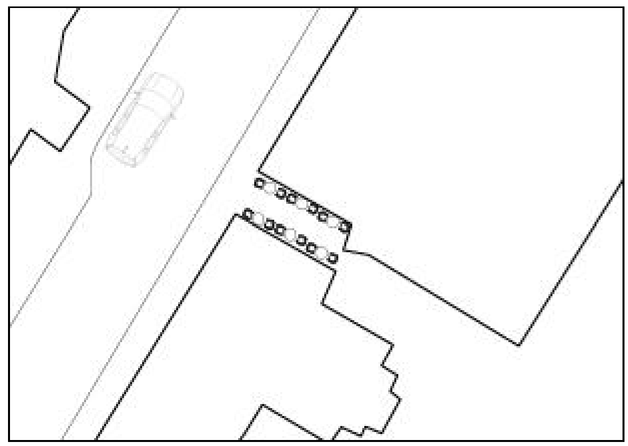
Cafe 13



Cafe 14



Cafe 15



Cafe 16