UrSA

Implementation of digital image processing technology and application for urban design

Supervisor

Prof. Eugenio Morello

Author

Gao Lei	**850161**
Ma Mengsha	**851018**

Politecnico di Milano

School of Architecture Urban Planning Construction Engineering

Master of Science Degree in Urban Planning and Policy Design

Dicember 2017

Acknowledgement

Thanks to Professor Eugenio Morello in Politecnico di Milano, for his dedicated instruction, the valuable time, academic spirits and profound knowledge in Urban Planning. During the academic year, he enlightened us not only in this research but also in other planning orientations, such as energy planning. The treasurable experience would not be forgotten.

Moreover, thanks to SegNet's designers - Alex Kendall, Vijay Badrinarayanan, and Roberto Cipolla. Meanwhile, thanks to the Visual Geometry Group, the designer of VGG model. We appreciate for all those who have shared their experience and knowledge and who had helped us in inspiring the work or critics it, for you have pushed us to examine the research from different perspectives.

Finally, we are grateful for our families, with the endless love and support even though thousands of miles away from China.

Abstract

No matter in which period, urban street as an important urban factor to urban space construction, on one hand it is the physical framework that carries transportation and culture; on the other hand it is also the space of urban life and the main public space of human perception to the city; thus, the urban street analysis is a part that can not be ignored in urban planning.

This work is main researching on how to apply the deep learning technology into urban street analysis, and we propose a method: Urban Street Analysis (UrSA), which studies on urban street through applying Google Street View (GSV) image API, convolutional neural network (SegNet) model, as well as data quantification, to obtain an analysis of urban street on the human visual perception level.

UrSA technology works on the street analysis in different ranges to achieve the target of urban research in various level; from the view of spatial features, which range from global cities' character, functional zones' character and to the specific urban streets' features. The result from UrSA also can relate with social issues, such as housing price and crime rate; at the same time, we proposed a standard that in based on the human sense of visual closure to make a classification on urban streets' open degree, and also to seek the relationship between the urban streets and social issues.

In the end, the practical ability of UrSA is still limited and exists some shortages, but we expect to help urban planners to realize more about the cities where we live in, thereby it can be as a reference and guidance to urban planning.

Key words: *UrSA; Computer Vision; urban street; social issues; deep learning; CNN; SegNet; Google Street View; urban space*

Riassunto

Non importa in quale periodo, la strada urbana è un importante fattore urbano per la costruzione dello spazio urbano, da un lato è la struttura fisica che trasporta i trasporti e la cultura; d'altra parte è anche lo spazio della vita urbana e il principale spazio pubblico della percezione umana per la città; quindi, l'analisi urbana delle strade è una parte che non può essere ignorata nella pianificazione urbana.

Questo articolo è la ricerca principale su come applicare la tecnologia di deep learning all'analisi delle strade urbane e proponiamo un metodo: Urban Street Analysis (UrSA), che studia su street urbani applicando l'API di immagine di Google Street View (GSV), la rete neurale convoluzionale (SegNet), così come la quantificazione dei dati, per ottenere un'analisi della via urbana sul livello di percezione visiva umana.

La tecnologia UrSA lavora sull'analisi stradale in diverse gamme per raggiungere l'obiettivo della ricerca urbana a vari livelli; dal punto di vista delle caratteristiche spaziali, che vanno dal carattere delle città globali, al carattere delle zone funzionali e alle caratteristiche specifiche delle strade urbane. Il risultato di UrSA può anche riguardare questioni sociali, come il prezzo delle abitazioni e il tasso di criminalità; allo stesso tempo, abbiamo proposto uno standard basato sul senso umano della chiusura visiva per fare una classificazione sul grado aperto delle strade urbane e anche per cercare la relazione tra le strade urbane e le questioni sociali.

Alla fine, l'abilità pratica di UrSA è ancora limitata ed esistono alcune carenze, ma ci aspettiamo di aiutare l'urbanista a realizzare di più sulle città in cui viviamo, quindi può essere un riferimento e una guida per la pianificazione urbana.

Parole chiave: *UrSA; Visione computerizzata; strada urbana; le questioni sociali; apprendimento approfondito; CNN; SegNet; Google Street View; spazio urbano*

TABLE OF CONTENTS

Table of Figures

Table of Tables

Table of Bar Charts

Chapter Content

1 Introduction

This chapter as a beginning to the dissertation provides a general framework and demonstrates the main work of our research, and explains what are the UrSA and VCI and these applications in the urban planning field.

1.1 Aim of Research

With the development of society, the research content of urban planning has been wide in corresponding. The various concepts of urban planning have appeared in the different periods. Today, urban planning is no longer a single discipline, which has intersected with many other disciplines, such as sociology, economy, environmental studies and so forth. Therefore, urban planning research need to consider much more than before, especially today, social issues are frequently showing in people's vision, such as housing issue, education, the safety issue and so forth, which is intensive with the lifestyle and also the quality of life. However, people are easy to think from the existed object and then to infer the possible consequence, but what we do is to reverse this process, which means from the aspect of the social issue to think and to seek reason in urban form.

We proposed Urban Street Analysis (UrSA), which is a method, and also it includes technology. The essential and hardest part of technology is image processing, which is also an important part of Computer Vision. Currently, Computer Vision is widely used in many fields, but not used that much in urban planning. This technique is a tool to help urban planners to know about the city in a more effective way, and it is based on Google Street View (GSV) images and Convolution Neural Network (CNN)[1]; as for the approach, though the image recognition to obtain the quantification result of urban street facts that includes 11 categories, such as the tree, sky, and building etc., and then compare and analyze the result to discover the features and characters of the street, in addition, the research scale depends on the data size.

The goals and aims of UrSA are to provide a more effective way to help urban research compare to the former method of urban data acquisition, and also it can be used for mapping a city in human visual perception, which we can discover some relationships between a city and people.

In a more specific scale, according to the urban street factors, we can map human perception in visual green, visual sky and visual building, which is regarded as references and guidance for urban afforestation planning, urban landscape design, urban space design, public space design and so forth.

[1] *In machine learning, a convolutional neural network (CNN, or ConvNet) is a class of deep, feed-forward artificial neural networks that has successfully been applied to analyzing visual imagery (Source from Wikipedia).*

Based on UrSA, we proposed a method for urban street study. The standard is the human sense of visual closure, which is visual closure index (VCI) that we proposed a new index and that can be seen as an abstract urban street form. We found that VCI depends on three main factors: building view ratio (BVR), tree view ratio (TVR) and sky view ratio (SVR). The VCI is a bridge for us to explore the relationship between the social issues and urban street form. As we know there are many reasons lead to the social issues, urban street form maybe as one of these reasons.

1.2 Main Works

The main work that we did in this process, firstly we took a long time in technical exploration, such as finding a fitting technical path and database; secondly, in the practice process, because of some limitation we needed to modify and adjust; and the third, proposed our idea, searched and constructed the theoretical foundation.

Chapter 1 is an introduction to our research, the goals, and application of UrSA. Chapter 2 is mainly about the two subjects' background, from development to the status quo, and also the cooperation of them. Chapter 3 presents two case studies, which are ImageNet and ILSVRC, Treepedia. The content of Chapter 4 and 5 are the implementation and application of UrSA, as well as the VCI; chapter 4 shows the technical exploration: semantic segmentation technology, technology practice framework, image database acquisition, image annotation and classification, semantic Segmentation Using SegNet, model assessment and perfection and statistics and visualization; in addition, it also displays the application of UrSA from the global level to local level: global city level (15 cities), functional zone level (8 districts), urban street level (36 blocks), which reflects the wide degree and deep degree of UrSA's application. Differently from the last chapter, chapter 5 shows the methodology of research, besides we invented a classification method of urban street and its application with two social issues, housing price and crime rate, in 4 cities: London, New York, Detroit, and Johannesburg. Chapter 6 is the conclusion, which described the main contribution and limitation in each part, as well as the further work.

Chapter Content

2 Research Context

2 Research Context

The proposal is based on the two fields, urban planning, and computer vision, which play different roles in our research. Urban planning is the destination to our research, while computer vision plays a role as assistant, to help us study in a more effective way. Thus, in chapter 2, we introduced the development and cooperation of these two fields in general level, which is as a scientific background framework, in order to realize clearly what is the role of these two fields and how they cooperate together in a research. The final aim is to explore the potential of Computer Vision for urban planning application.

2.1 Urban Planning Background

According to the discovery by archaeologists, the earliest urban planning appeared in Mesopotamian, Indus Valley, Minoan, and Egyptian civilizations in the third millennium BCE, and the street pattern in there was in the grid. (Davreu, R. (1978). Cities of Mystery: The Lost Empire of the Indus Valley. The World's Last Mysteries, 121-129.)

> *"Cities existed for various reasons, and the diversity of urban forms can be traced to the complex functions that cities perform. "(Ellis, C. (2011). History of cities and city planning. Recuperado de http:// www. art. net/~ hopkins/Don/simcity/manual/history. html.)*

With society's development and continuing improvement, urban planning has also been advanced with the times. Various street patterns, dimensions, as well as the aesthetic value, and the relationship with the whole society and so forth, urban planning in each period reflected the specific characteristics of that era, such as the residents' lifestyle, thinking of objects, the internal relationship within the social factors and so forth, of course, the urban form and appearance and functions.

Greeks

> *"In the ancient Greek urban planning layout, except for the humanism and naturalism, with the development of ancient Greek aesthetics and natural science, rational thinking, it appeared another urban planning form with apparently artificial effects-Hippodamus, and Aristotle called him 'the father of city planning'."(Aristotle, Politica II)*
>
> *"With the straight ruler," says Meton, "I set to work to inscribe a square within this circle; in its center will be the market-place, into which all straight streets will lead, converging to this center like a star, which . . . sends forth its rays in a straight line from all sides." (Lewis, M. (1961). The city in history. San Diego, New York, London: A Harvest Book Harcourt, 172)*
>
> *"More likely, as Lavedan suggests, Hippodamus may have introduced the formal, enclosed agora in planning the Piraeus. His true innovation consisted in realizing that the form of the city was the form of its social order, and that to remold one it is necessary to introduce appropriate changes in the other. He seems, too, to have realized that town planning should have not merely an immediate practical aim, but an ideal goal of larger dimensions; and he thought of his art as a means of formally embodying and clarifying a more rational social order. "(Lewis, M. (1961). The city in history. San Diego, New York, London: A Harvest Book Harcourt,172)*

Hippodamos system, in which the street network performs in gridiron way regarded as the urban frame, which shapes the orthonormal public center in the city, in order to maintain the city order and as well as the urban aesthetics. The famous and typical city of Hippodamos is Miletus, also it played an important role in the modern history of the colonial urban form.

> *"Aristotle put into words the nature of this transition from the preparatory urban processes and functions to emergent human purposes, in terms it would be hard to improve: 'Men come together in the city to live; they remain there in order to live the good life.' To define the nature of the city in any particular cultural setting would be in part to define both the local and the more universal qualities of the good life. "(Lewis, M. (1961). The city in history. San Diego, New York, London: A Harvest Book Harcourt, 111)*

On the one hand, Hippodamus, which is a system integrated with geometrization and mechanization, was satisfied for efficient and simplified urban construction after Greco-Persian Wars and in ancient Rome. At the same time, it established a new urban order and ideal, that followed the principle of Ancient Greek Mathematics and aesthetics, as well as the persuasion of elegant living in the middle or upper class.

In terms of urban planning, the appearance of Hippodamus changed the former flexible and organic way into formal aesthetics, which had an impact on activity and development of the city. (Hong Liangping, 2002)

Ancient Rome

Owens remarks that "new towns and cities were founded for the purpose of military security, administrative efficiency and economic exploitation, and of course to assist the process of 'Romanisation'."(Owens, E. J. (1989). Roman town planning. Roman Public Buildings, Exeter, 7)

> *"The regular checkerboard layout within a rectangular boundary, the arcaded walks, the forum, the theater, the arena, the baths, the public lavatories…Except for the elaborateness of the public baths and the over-sized arena (which even in a small town might hold twenty thousand people), none of these facilities was new. What Rome did was to universalize them: making them, as we would say today in somewhat Roman terms, 'standard equipment'."(Lewis, M. (1961). The city in history. San Diego, New York, London: A Harvest Book Harcourt, 208)*

During the period of colonization, there were many regions that were planning by a similar order in the landscape, arranging the roads and making the regions into rectangular parcels, which still can be seen from the air and respected in daily use. At that time, the system of 'centuriation' has been widely used in many places, such as Italy, Africa and also Dalmatia. (Lewis, M. (1961). The city in history. San Diego, New York, London: A Harvest Book Harcourt, 209)

However, the orthogonal patterns are the most obvious feature whenever and it served for primarily organizational function, but some scholars have disputed with this.

> *“William MacDonald believed that the main orthogonal streets, rather than merely organizing the area of the town, served the function of displaying principle buildings of the town in question. The effect for someone passing through the city would be a “sequential viewing of public monuments”. This notion is discussed in The City in the Roman West in which Laurence, Cleary and Sears remark: The grid of streets was an easy way in which to provide a shape to the space within the walls of a city, but it was not intended as a planning tool in the manner of the layout of Manhattan in 1807 with a view to further expansion of the city over the 2,028 blocks of Manhattan Island. “(Laurence, R., Cleary, S. E., & Sears, G. (2011). The City in the Roman West, c. 250 BC–c. AD 250. Cambridge University Press, 116.)*

The first century BC author and architect Vitruvius writes in his “de Architectura” of the importance of “the general plan of the walls of a city and its public buildings”(Vitruvius de Architectura 1.3.1.), but he did not mention that the orthogonal grid street plan or the city more generally. From this, it seems to support the argument of Roman planners, which means the designed orthogonal streets were much more concentrating on the linear experience of passing through a space of urban streets, rather than the layout of urban space.

Indeed, Simon Keay and Martin Millett (two prominent archaeological surveyors) believe that the two main roads, the monumental architecture and the walls of the town were created first and the central grid was only expanded to fit the walled enclosure later on. (Laurence, R., Cleary, S. E., & Sears, G. (2011). The City in the Roman West, c. 250 BC–c. AD 250. Cambridge University Press, 119-121.) The feature of the city in ancient Roma, the one is secularization and also reflects monarchical power at that time, in the other hand, it is the intensive sense of order through the creation of public space in big size, as well as the axis line system.

To compare with the ancient Greek urban planning, whose manifestations is more concerned and basic on human dimension and human life, the ancient Roman urban planning and architecture design are making the urban, architecture and any parts of the city in coordinated and harmony, the whole city was regarded an individual system that parted from the people.

Middle Ages

> *In the majority of medieval town these powers, the spiritual and the temporal, with their vocational orders, the warrior, the trader, the priest, the monastery, the bard, the scholar, and the craftsman, achieved something like equilibrium. That balance remained delicate and uncertain; but the effort to maintain it was constant and the effect real, because each social component was weighted, each duly represented. (Lewis, M. (1961). The city in history. San Diego, New York, London: A Harvest Book Harcourt, 252)*
>
> *It is important to mention that religious elements have been extremely important throughout urban history. Ancient peoples had sacred places, often associated with cemeteries or shrines, around which cities grew. Ancient cities usually had large temple precincts with monumental religious buildings. Many medieval cities were built near monasteries and cathedrals. (Ellis, C. (2011). History of cities and city planning. Recuperado de http://www. art. net/~ hopkins/Don/simcity/manual/history. html.)*

At that time the religion main focused on human psychology and mentality, also concentrated on a set of creeds and ceremonies, instead of the exploration of social reform and social justice.

Besides, the urban streets form in the most of the early medieval cities in Europe were still the same as before, which means it was in irregular lines, therefore apparently there was no change and effort to make in their road network in geometrical form. (Adams, T. (1935). Outline of town and city planning.) Among all the different scales of a city, in general, it was the same pattern: before the church, there was a strong sense of visual closure by a square in the arch or irregular shape, those square with church formed the central public space.

> *"There are short approaches to the great buildings and the blocked street views, which increase the effect of verticality: one looks not to right or left over a wide panorama, but skyward. This ambulatory enclosure was an organic part of the processional movement and also of the relation between the structures to each other. The usual movement for eye is up and down, and the movement direction of the walkers was always changing, which could help to create dynamic and three-dimensional spatial forms through each deep passage, that gave walkers the feeling of constriction in the narrow streets as well as of release when they meet a public space, such like parvis and market place. "(Lewis, M. (1961). The city in history. San Diego, New York, London: A Harvest Book Harcourt, 278)*

Firstly, the idea of urban planning in Middle Ages, apparently affected by the power of religion that is why churches always located in the city center and is involved in the urban plan, and it also constructed the main public space with the square that is usually in front of the cathedral, which made an order and beautiful skyline. The second is the document we hardly found some famous and outstanding planner, it seems that there is no unified and integrated plan intention. From the result, the urban plan and landscape are preferred to in a spontaneous approach, which the urban construction took full advantages of topography and natural landscape, thus it appeared the diversity in urban features in Middle Age. (Gibberd, F. (1967). TOWn Design, New York: The Frederick/L Praeger.)

Renascence

> *"When come to the real renascence of European culture, it represents the great age of city and building, as well as the intellectual triumph, which began in the 12th century and had an achievement of a symbolic apotheosis in the work of an Aquinas, a Dante, an Albertus Magnus and a Giotto. In addition, the Renascence is regarded as a movement toward freendom and the re-establishment of the dignity of man. "(Lewis, M. (1961). The city in history. San Diego, New York, London: A Harvest Book Harcourt, 345) There is no doubt that the Humanism is the core and foundation of that period.*

The "narrow" and "small" urban structure in middle age could not satisfy the need of new life, some cities in early renascence period, such as Milan, Bologna, and Ferrara, were transformed on transportation and infrastructure for improving the quality of life, standard sanitation and increasing the defense etc. People believed that there was an existence of "ideal form" of urban and it could be controlled. Thus, the "ideal cities" appeared. However, most of these plans were not come true in the society than in which the economy and politics did not create the opportunities for it, but it still had a deep influence in followed classicism and baroque style. In that situation, the planners had to give up a set of plan about "ideal cities", and focused on a part of urban transformation then expanded to the surroundings, such as plaza, villa, and courtyard etc. The typical transformation was the square of Roma city hall by Michelangelo.

From the urban transformation of Ferrara by Biagio Rossetti, who has been seen as one of the earliest modern urban planners and design the unique vision system, which connected all roads with vision points, we can see in that period, the vision system had been highly thinking for the urban plan.

> *"It is generally agreed that the starting point of baroque architecture lies in the re-construction of pontifical Rome subsequent to the Council of Trent (1545-1563). Driven by an exceptional sense of the future of the city, Pope Sixtus V (1520-1590) had the scale of the city dilated and regulated by a network of tensions around seven poles, seven ancient churches scattered over a space that looked outwards towards the countryside: this project, the plans of which are conserved within the Vatican library, where the tensions between poles are clearly marked by vectors, introduces a method of governing space based on the existence of strong points - monuments, squares, etc. - acting at a distance both on the future development of the city and on the rearrangement- requalification of what already exists." (Le Dantec, J. P. (1991). For a Baroque Approach to Cities and Architecture. Architecture and Behaviour/Architecture et Comportement, 7(4), 473-478)*
>
> *But for those new planners, they had no attempt to build a harmony relationship between their design and old medieval patterns. The old was still standing, the new buildings create a rich, complex order, and more satisfying esthetically than the uniform, simple composition of a later period. The classic example of this visual achievement is the Uffizi whose narrow and strait street enclosed by the building in renascence Florence. (Lewis, M. (1961). The city in history. San Diego, New York, London: A Harvest Book Harcourt, 349)*

In the late 16th century, the eastern architecture design and urban planning in which appeared two different art forms, baroque and classicism.

> *"True to the popular Baroque traditions of the time, the Europeans designed streets, squares, and markets in an elegant geometry: eastern cities like Halifax, Charlottetown ... and the French fortress at Louisburg reveal such influences. This approach to planning reflected the triumph of authority over landscape; despite the grade of the hill or the presence of waterways, the formal pattern laid out by military engineers dominated. Legal systems provided for private property ownership, imposing an economic order that would continue to influence the shape and development of communities for centuries to come."(Grant, J. (2000). Planning Canadian cities: context, continuity and change. Canadian cities in transition: The twenty-first century, 444.)*

The concept of the baroque was obviously forming itself in the 17th century, and it held itself in the two contradictory elements of that age. "First, the abstract mathematical and methodical side expressed to perfection in its rigorous street plans, its formal city layouts, and in its geometrically ordered gardens and landscape designs. And at the same time, in the painting and sculpture of the period, it embraces the sensuous, rebellious, extravagant, anti-classical, anti-mechanical side, expressed in its clothes and its sexual life and its religious fanaticism and its crazy statecraft. Between the sixteenth and the nineteenth century, these two elements existed together: sometimes acting separately sometimes held intension within a larger whole.” (Lewis, M. (1961). The city in history. San Diego, New York, London: A Harvest Book Harcourt, 351)

> *The neat, orderly arrangement of new towns and grand public spaces built as the European nation-state emerged helped Descartes imagine new ways to organize human inquiry. “It was the geometrical, mechanistic clarity emanating” from these planned urban landscapes “That was also perceived as reflecting the order of the universe, and through which Descartes himself helped usher a new era of confidence in the intellectual faculty of the individual.” (Akkerman, A. (2001). Urban planning in the founding of Cartesian thought. Philosophy & Geography, 4(2), 156.)*

> *“Finally, the baroque is not a style but an “approach”, an “attitude to architecture and space” as Hans Hollein so rightly puts it. The expression of an era that has lost all hope of a “radiant future” where artists, thinkers, scientists and politicians are faced with the disorder of a disenchanted world, unlike the trendy hedonism that supposedly represents “age of emptiness”, the baroque attempts to invent a new illusion-free spatial togetherness something like Andre Glucksmann’s humanism conscious of being habited by the inhuman. More than an aesthetic trend, therefore, the baroque is an ethic based on recognition of otherness.” (Le Dantec, J. P. (1991). For a Baroque Approach to Cities and Architecture. Architecture and Behaviour/Architecture et Comportement, 7(4), 473-478)*

In summary, in the Renascence, especially in the later, urban planning had become the proprietary right of a dignitary, and it had one-sidedness in practice, and also regarded as a tool for political power. At the same time, it ignored the public well-being and their quality of life.

19th century

In the 19th century, many countries in Western Europe and the United States have entered the stage of the rapid development of capitalist economy. The new production factors, social structures, lifestyle and social needs of industrial production have never been experienced in human history. Also, the industrial revolution has led to a ‘mushrooming growth’ of new industrial cities in a wide area. (Hall, P., & Tewdwr-Jones, M. (2010). Urban and regional planning. Routledge.)

At that time, the urban form, basic infrastructures, and ecological environment have been all ‘infected’ by the features of the industrial age.

> *Trade, industrial production, mechanization, organization, capital accumulation-all these activities helped the building and extension of cities. But these institutions do not account for the feeding of the*

> *hungry mouths, nor yet for the high sense of physical vitality that accompanied this whole effort. People do not live on air, even though "city air makes people free," as the German saying went. The thriving life of these towns was rooted in the agricultural improvement of the countryside: it is nothing less than a cockney illusion to separate the town's prosperity from the land's. (Lewis, M. (1961). The city in history. San Diego, New York, London: A Harvest Book Harcourt, 260)*

Therefore, this period in the field of urban planning in the west, many social reformer, planners, architects, engineers, ecologists aimed at the problems of the big cities, through transforming the physical environment to solve the social problems in big cities, moderate sharp social contradictions, thus to build a harmonious, efficient and new type of society. This shows that a correct urban planning is necessary for the development of the city.

In 1889s, an Austrian architect Camillo Sitte published the famous book "The Art of Building Cities". He was in the era of the industrial development of the city and argued that the status of construction of ignoring the artistry of space-urban landscape was monotonous and extreme regular, there was no relationship in space, and designed to meet the symmetry, so he came up with "certain art principle" for urban construction, and emphasizes the human dimension, the scale of the environment, as well as the coordination between the people's activities and human perception. (Sitte, C., & Stewart, C. T. (1945). The art of building cities: city building according to its artistic fundamentals. New York: Reinhold Publishing Corporation.)

20th century

From the late 19th century to early 20th century, the City Beautiful Movement appeared in Europe and the United States and its main purpose was to recover the good environment and attraction through the landscape created in the city, in order to remit suburbanization. However, the limitation is obviously and it was less help for solving the urban issues, the decorative urban planning seemed to satisfy the urban vanity instead of thinking in essence.

Before the World War II, the main concept in urban planning was always struggling between the humanity and mechanization. Until to the early post-war period, it was the summit of function rationalism, and later in the 1950s, the idea of the urban ecological environment has risen in the air. This is the general situation in modernism urban planning. Los Angeles, as a typical representative city of modernism, collected most urban issues, such as the disparity between the rich and the poor, high crime rate; interpersonal relationship became alienated and indifferent and so forth.

After the 1960s, with the rising of introspection and criticizing to the modernism, urban planning has turned to the exploration of urban social culture, rather than the construction of pure material space; from the aesthetic values of landscape turned to the creation of public space and public life in sociological meanings; from the layout of baroque style to the mental research in the environment. In summary, it was a beginning that from the different views, such as social, cultural, environmental and ecological aspect, to analysis and study in urban planning.

From the 1990s, the concept of urban planning has to focus on the ecological city and humanism. The related research results are “A decision-Centres View of Environmental Planning”, which was published in 1987s by A.Faludi; “Planning: Clearer Strategies and Environmental Controls, which was published in 1988s by R.Erhman; and “Planning for a Sustainable Environment”, which was written by A.Blowers in 1993 and so forth.

In the 1993s, Kunstler published “geography of nowhere” and pointed out that since World War II, the patter of urban development in the United States was incompact and unbridled, which leaded the sprawl of high way, as well as the social issues. Thus he thought this pattern should be changed and sought the reasons from former urban planning, transformed that situation caused by industrialization and modernization, which is the idea of New Urbanism.

The essence of New Urbanism is to satisfy the human needs by transform the old city center but to keep the surface and dimension, as for the suburban, it prefered to the compact development approach. After 1990s, Compact City has been regarded as a sustainable urban developed form, also Andres Duany and Elizabeth Zyberk came up with Traditional Neighborhood Development (TND), which focus on small dimension; and Peter Calthorpe proposed Transit Oriented Development (TOD), which consider the whole city and regional scale. Both of them reflect that the basic character of New Urbanism planning, compact, walkable, complex function, human dimension and environmental.

After World War II, the Western Europe highlighted the idea of control the urban development; meanwhile in the United States and Canada, it appeared Urban Sprawl, which had an impact on ecology and society. After 1990s, the scholars of North America began to inspect this urban developmental approach and came up with the idea of managing the development of land to improve the integrated benefit of space increase, which is the Growth Management. With the influence of ecological ideology and New Urbanism, in the 1997s, the governor of Maryland, P. N. G. Lendening firstly raised the concept of Smart Growth, later it was as an important content of New Livability Agenda for the 21st Century. In a simple word, to achieve the Smart Growth is the goal, employing the Growth Management is a method.

The Canada scholar, Mark Roseland proposed 10 principles (Urban Ecology,1996) for creating the ecological cities:

> *"(1) revise land-use priorities to create compact, diverse, green, safe, pleasant and vital mixed- use communities near transit nodes and other transportation facilities;*
>
> *(2) revise transportation priorities to favor foot, bicycle, cart, and transit over autos, and to emphasize 'access by proximity;*
>
> *(3) restore damaged urban environments, especially creeks, shore lines, ridgelines and wetlands;*
>
> *(4) create decent, affordable, safe, convenient, and racially and economically mixed housing;*
>
> *(5) nurture social justice and create improved opportunities for women, people of color and the disabled;*
>
> *(6) support local agriculture, urban greening projects and community gardening;*
>
> *(7) promote recycling, innovative appropriate technology, and resource conservation while reducing pollution and hazardous wastes;*
>
> *(8) work with businesses to support ecologically sound economic activity while discouraging pollution, waste, and the use and production of hazardous materials;*
>
> *(9) promote voluntary simplicity and discourage excessive consumption of material goods;*
>
> *(10) increase awareness of the local environment and bioregion through activist and educational projects that increase public awareness of ecological sustainability issues."*
>
> *(Roseland, M. (1997). Dimensions of the eco-city. Cities, 14(4), 197-202.)*

As for the humanism, Lewis Mumford wrote in "The City in History": "Significant improvements will come only through applying art and thought to the city's central human concerns, with a fresh dedication to the cosmic and ecological processes that enfold all being. We must restore to the city the maternal, life-nurturing functions, the autonomous activities, and the symbiotic associations that have long been neglected or suppressed. For the city should be an organ of love, and the best economy of cities is the care and culture of men." (Lewis, M. (1961). The city in history. San Diego, New York, London: A Harvest Book Harcourt, 575)

2.2 Computer Vision

> *"The goal of computer vision is to model and automate the process of visual recognition, a term we interpret broadly as 'perceiving distinctions between objects with important differences between them.' " (Forsyth, D., & Ponce, J. (2011). Computer vision: a modern approach. Upper Saddle River, NJ; London: Prentice Hall.)*

2.2.1 Development of Computer Vision

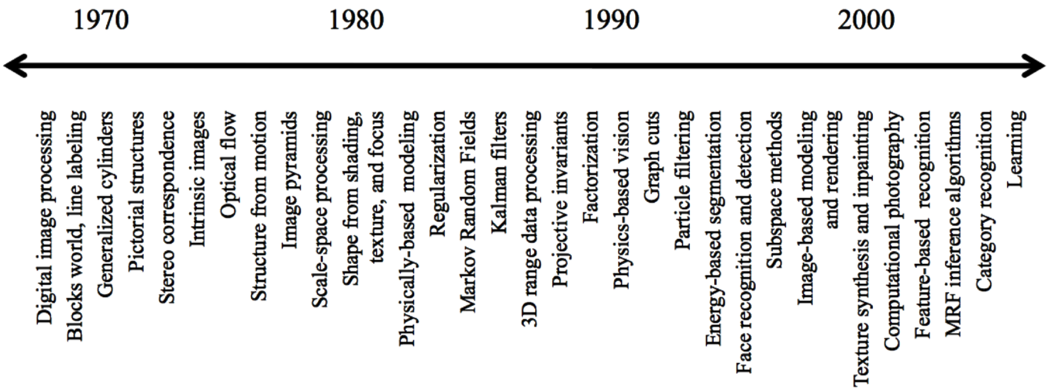


Figure 2.1 A rough timeline of some of the most active topics of research in computer vision

Source: Szeliski, R. (2010). Computer vision: algorithms and applications. Springer Science & Business Media

Computer vision first appeared in the early 1970s, at that time, it was regarded as a part of the visual perception of a huge agenda that is to imitate human intelligence and to give robots intelligent behavior. There were some pioneers in the early phase of artificial intelligence and robotics pointed out that it is easier to solve the "visual input" problem instead of solving other problems such as higher-level reasoning and planning. A well-known story, in 1966, Marvin Minsky at MIT asked his undergraduate student Gerald Jay Sussman to "spend the summer linking a camera to a computer and getting the computer to describe what it saw" (Boden 2006). But now, we know that achieving this goal is much more difficult than what we imagine.

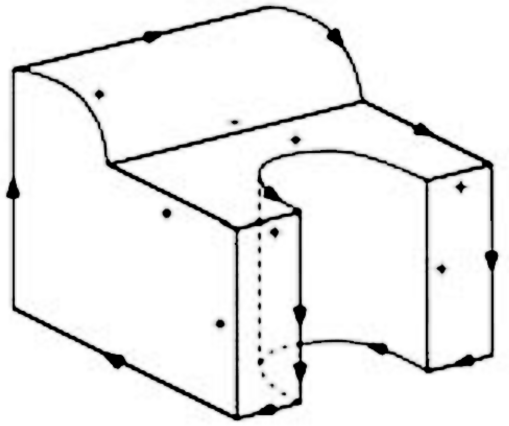


Figure 2.2 Line Labeling
Source: Szeliski, R. (2010). Computer vision: algorithms and applications. Springer Science & Business Media

> *"What distinguished computer vision from the already existing field of digital image processing was a desire to recover the three-dimensional structure of the world from images and to use this as a stepping stone towards full scene understanding. Winston (1975) and Hanson and Riseman (1978) provide two nice collections of classic papers from this early period. Early attempts at scene understanding involved extracting edges and then inferring the 3D structure of an object or a 'blocks world' from the topological structure of the 2D lines. Several line labeling algorithms were developed at that time gives a nice review of this area. The topic of edge detection was also an active area of research; a nice survey of contemporaneous work can be found in Davis 1975." (Szeliski, R. (2010). Computer vision: algorithms and applications. Springer Science & Business Media, 11-12)*

Meanwhile, three-dimensional modeling of non-polyhedral objects was also being developed and researched. A relatively popular method, which is used generalized cylinders, i.e., solids of revolution and swept closed curves, often arranged into parts relationships.

More quantitative approaches to computer vision were also exploited and studied at that time, which is containing many feature-based stereo correspondence algorithms and intensity-based optical flow algorithms. In addition, the early work in simultaneously recovering 3D structure and camera motion also in the starting stage around that time.

According to the Richard Szeliski's interpretation of the concept that is written by David Marr, there are three levels of the description of a visual information processing system:

- *Computational theory: What is the goal of the computation task and what are the constraints that are known or can be brought to bear on the problem?*

• Representations and algorithms: How are the input, output, and intermediate information represented and which algorithms are used to calculate the desired result?

• Hardware implementation: How are the representations and algorithms mapped onto actual hardware, e.g., a biological vision system or a specialized piece of silicon? Conversely, how can hardware constraints be used to guide the choice of representation and algorithm? With the increasing use of graphics chips (GPUs) and many-core architectures for computer vision, this question is again becoming quite relevant. (Szeliski, R. (2010). Computer vision: algorithms and applications. Springer Science & Business Media, 13)

In the 1980s, the main study direction was preferred to perform quantitative image and the analysis of scene through more complex and sophisticated mathematical techniques.

A popular approach, image pyramids, was used very well and widely to achieve goals, for instance, image blending (Figure 2.3) and coarse-to-fine correspondence search. Later the image pyramids had been developed as well, which was using the idea of scale-space processing. In the end of 1980s, the wavelets have started replaced or increasing regular image pyramids in some applications.

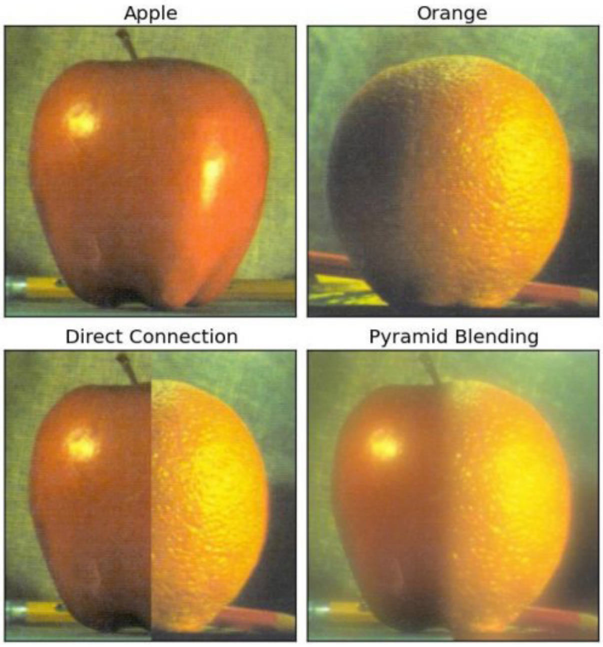


Figure 2.3 pyramid blending

Source: http://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_imgproc/py_pyramids/py_pyramids.html

As an important quantitative shape cue, the application of stereo was also developed by a set of shape-from-X techniques, such as photometric stereo, shape from shading, shape from focus and shape from texture.

It is worthy to mention here is that more deep research on edge and contour detection, which was active in that period, involved the adoption of dynamically evolving contour trackers, such as three-dimensional physically based models and snakes.

More and more researchers found that many of the stereo, shape-from-X, edge detection and flow algorithms, which could be using the same mathematical framework or unified when they were regarded as variational optimization issues and also made more well-posed while using regularization. Moreover, the three-dimensional range data processing, acquisition, merging, modeling, and recognition were being focused and developed in that period.

As mentioned before, there were a lot of topics developed and were also active during that period, but actually, just a few of them have become more active in the 1990s.

From here the fully automated sparse 3D modeling systems have been created, which is basic on a lot of techniques. At the first, the main task is for the recognition through a set of activities that using projective invariants, later it has become a coordinated effort in order to work on the structure from motion problem. The former activity was contributed to projective reconstructions that had no requirement to the knowledge of camera calibration. At the same time, factorization techniques have been studied as well and worked well on the problems of orthographic camera approximations, later it extended to the perspective condition. Finally, the result was that the field began to use the full global optimization[1] that was found to be the same with the bundle adjustment techniques, which is used in photogrammetry.

The multi-view stereo algorithms had a relatively high attention and were producing the complete 3D surface at that time, as well as today. Moreover, with some techniques that built on tracking and reconstructing smooth occluding contours continued to develop, the techniques for producing 3D volumetric descriptions from binary silhouettes have not been ignored. As for the tracking algorithms, which were improved apparently, especially the snakes, particle filters, level sets and intensity-based techniques, which were used frequently on tracking faces and whole bodies.

[1] Global optimization is a branch of applied mathematics and numerical analysis that deals with the global optimization of a function or a set of functions according to some criteria. Global optimization is distinguished from regular optimization by its focus on finding the maximum or minimum over all input values, as opposed to finding local minima or maxima (Source from Wikipedia).

Image segmentation, which is always been the active direction from the beginning of computer vision and it is producing techniques from the minimum energy and minimum description length, as well as the normalized cuts and mean shift, which we introduce in detail at the technical part.

During this period, the most remarkable improvement in the computer vision was the increase of interaction with computer graphics, particularly developed within the cross-disciplinary area between the image-based modeling, also rendering. The original idea of working on real-world imagery to produce new animations became highlighted with image morphing techniques, and then it was using in view interpolation and rendering of the full light-field scene. At the same time, in order to create the realistic 3D models through a lot of images, the image-based modeling techniques were also being introduced and then developed. Debevec, P. E., Taylor, C. J. and Malik, J. present a new method of modeling and rendering architecture photorealistically from a small number of photographs in 1996 (Figure 2.4).

> *The work consists of:*
>
> *Photogrammetric Modeling: A method for interactively recovering 3D models and camera positions from photographs*
>
> *View-Dependent Texture Mapping: A method for turning a 3D model and a set of photographs from known positions into renderings*
>
> *Model-Based Stereo: A method of refining an approximate geometric model of a scene to conform to its actual appearance in a set of photographs*

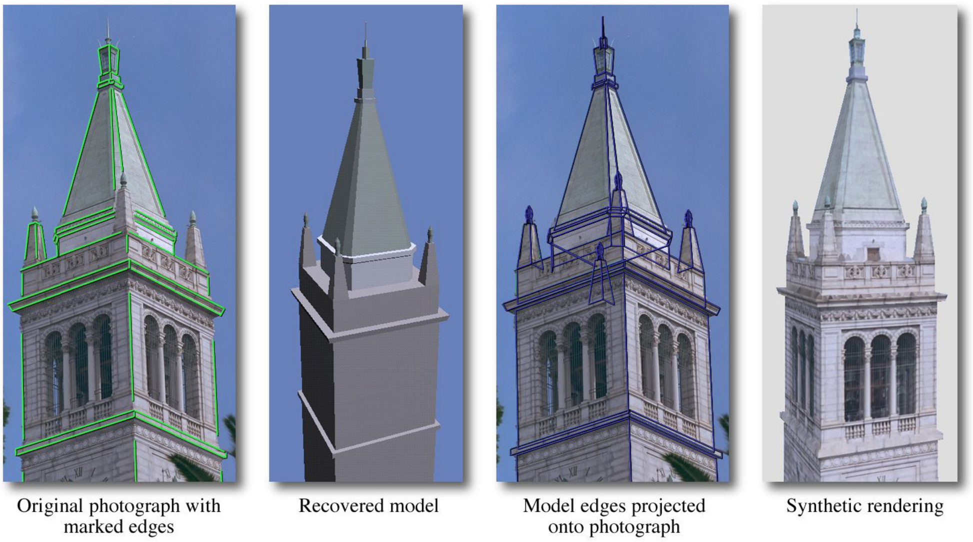

Original photograph with marked edges
Recovered model
Model edges projected onto photograph
Synthetic rendering

Figure 2.4 Photogrammetric Modeling

Source: Debevec, P. E., Taylor, C. J., & Malik, J. (1996, August). Modeling and rendering architecture from photographs: A hybrid geometry-and image-based approach. In Proceedings of the 23rd annual conference on Computer graphics and interactive techniques (pp. 11-20). ACM.

In the past decade, the interaction between the computer vision and graphics has been continued deepening. Especially many themes have been introduced under the rules of image-based rendering, for instance, the capture and rendering of light-filled, image stitching as well as the capture of high dynamic range (HDR) using exposure bracketing, all of these were untitled as computational photography, which was to prove the increased use in the everyday digital photography, such as the creation of HDR image through the rapid adoption of exposure bracketing needs to exploit the tone mapping algorithms, then to convert the image to the visible results. Except for merging multiple exposure images, it has been developed other techniques, not only to merge flash images and non-flash images but also to select different parts from the overlapping image in an interactive and automatic way.

In summary, there were three main trends in this period. The first prominent trend is for the object recognition, which are the feature-based techniques including learning that is also prominence in other recognition fields, for example, location recognition and scene recognition. But currently, the patch-based features are becoming to be the main trends; some groups are still working on the contours and regional segmentation.

The second main trend in the past decade was focusing on the complex global optimization problems with more efficient algorithms. However, firstly it started with graph cuts, but there were many obvious improvements in message passing algorithms.

The last trend, is the application of complex machine learning techniques in Computer Vision field, which is also the most popular theme in nowadays. It almost coincides with the increasing immense data on the Internet that is making easier for machine learning without the supervision.

2.2.2 Status Quo of Computer Vision

20 years ago, the emergence of computer revolutionized the development of workplace. Up to 2017, over 75% of the office in which work with the computer and the Internet. In 1990, about 15% of American households owned a computer and increased to 70% in 2017.

Computer vision is a new field in the artificial intelligence, which takes the theory of visual processing as a center and also an important branch of information processing that is based on image processing, pattern recognition, computer science and psychology. Computer vision technology with digital image processing, digital signal processing, optics, physics, applied mathematics, geometry, and pattern recognition as a whole, its application has already related with computational geometry, computer graphics, image processing, robotics, and other fields. Computer vision is not only an engineering field but also a challenging and important research field in science.

There are two aims of computer vision research: one is to develop an image understanding system that from the input of image's data to construct automatically the description of a scene, another one is to understand the human vision that in order to work replace people and solve the problem that people can not. Currently, computer vision is also an active and effective research topic in the field of artificial intelligence and robotics.

Vision understanding is an indispensable processing within the system of computer vision. At present, the robots that have the function of visual feedback are able to complete the complicated task instead of people, such as products of automatic assembly, welding and inspection, the biomedical automatic diagnosis and automatic interpretation of remote sensing images, a variety of vehicle automatic navigation, etc. The endowed vision processing capability of robots, which is like a human, and the expectation that they can serve for human are partly achieved in some certain condition. Today many applications have penetrated to all kinds of a research field that involved astronomy, geography, medicine, chemistry, and physics.

Some people maintain that computer vision is one of the keys of achieving intelligent robot and fifth-generation computer, and also consider that understanding and analysis of scene are the main points within the complicated processing process. The realization of scene understanding by machine builds on the interaction between input images, the pre-deposit related object structure and the knowledge of environmental constraint, and to create the meaningful and clear description. This kind of process can be divided into several steps, extracting object information of an image, completing some calculations, introducing the related knowledge that studied before during the different understanding phase, and then finishing the understanding processing. Actually, some similar work has started at the 50s, now the major work of three-dimensional objects and scene analysis is the recognition.

Other people think that the professor Marr's computer vision theory is the most outstanding contribution to this field. In the 90s, Rosenfeld pointed out that should pay attention to three aspects, the robustness of the calculation, the research of active vision and qualitative vision. The core of computer vision is information processing by making use of a set of measures and methods that offered by a computer, which are including vision information acquisition, image pre-processing, cut, description, and recognition. In the 1965s, L.Roberts came up with some basic methods about three-dimensional information acquisition that have been employed in the computer vision research field until today.

At present, the major methods for acquisition of visual information are divided into two main categories, which are the active and passive method. The active methods that need add the special artificial light to the test object, including the triangle light method, structured light method, and time-of-flight method.

The triangle light method is similar to the triangulation method, and it takes a lot of time to measure point by point. A structured light method is to put the projection of an image that is known structure to the object's surface because the object orientation of the surface is different, the standard image will produce distortion, using this distortion can calculate the 3D coordinates of the surface of the object. This method came up by a Japanese scholar in the earliest, which can be realized by laser scanning or projector.

The time-of-flight method is built on the principles of radar, which have the capability of measuring the distance of object surface indirectly way order to know 3D information and it does not refer to the issue of image processing. Employ this method by laser radar or ultrasonic radar that its shortcoming is hard to focus, but easy to processing.

The passive method is to obtaining 3d information under natural light conditions, which include the stereoscopic method, shadow restoration shape method, motion recovery shape method, and texture restoration shape and grayscale stereoscopic method.

The stereoscopic method is very similar to the vision theory of humans. Obtaining at least two images from the cameras that located in different position, then according to the triangulation measurement principle and using parallax of the corresponding point in the stereo image, finally calculating the 3D information of an object. Therefore, the key of the stereoscopic method is the matching within two or more images. At the early period, the matching was major depending on the region-based gray count, but modern methods are preferred to the feature matching.

The shape analysis method is based on the information analysis of the gray shading distribution, object movement and texture structure of the image. Motion sequence image analysis is used to obtain 3D information by means of the calculation of 3D motion parameters and it is also according to the object and camera motion order to obtain several sequential images. This method has a high attention in the computer vision research and has become a branch. In summary, the acquisition of 3D information is the base of computer vision research, as well as one of very active topic today, and plays an important role no matter in theory or in practice.

Computer visual information processing technology mainly depends on the image processing method, which includes image enhancement, data coding, and transmission, smooth, edge sharpening, segmentation, feature extraction, image recognition, and understanding, etc. The quality of the output image is improved considerably after these treatments, improving the visual effect of the image and it is also convenient for the computer to analyze, process and identify.

Here are some key techniques:

A. Data-driven segmentation:

The general data-driven cut includes cut based on edge detection, region segmentation, and integration of edge and region, etc. Figure 2.5 shows a sample of image edge detection. As for the cut based on edge detection, the conception is to detect edge point of an image at first, then connect to module resulting in constructing a split area. The cut based on region segmentation is according to features of image data and divide space of the image. There are some features used frequently, which are the level of gray of original image and color features.

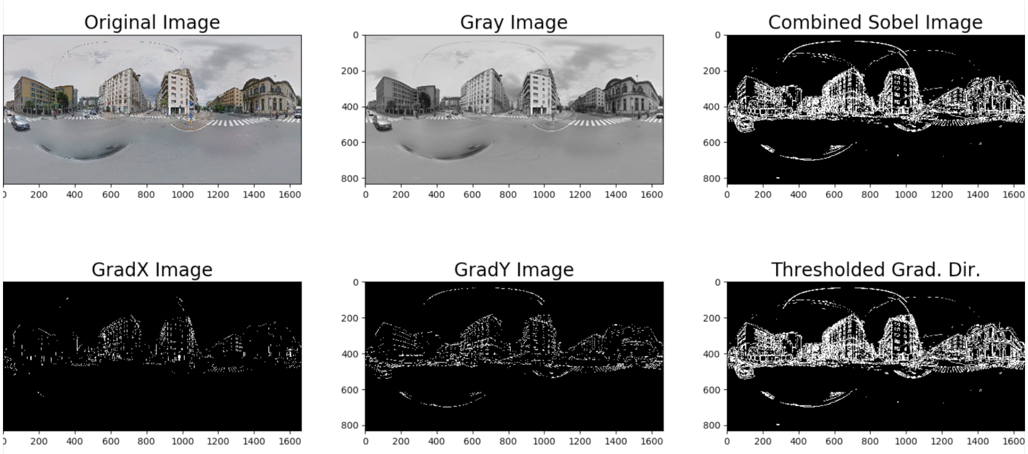


Figure 2.5 Image Edge Detection

Source: made by author

B. Model driven segmentation:

Common model-driven segmentation includes active contour model, combinatorial optimization model, objective geometry and statistical model. Active contour model, also called snakes, is a framework in computer vision for delineating an object outline from a possibly noisy 2D image. The Snakes model is used to describe the dynamic contours of segmentation targets, because of the integral operation of its energy function; it has better anti-noise and is not sensitive to the local vagueness of the target. Therefore, it has wide applicability, but this segmentation method is easy to local optimal, so the initial contour is as close to the actual contour as possible. The snakes model is greatly used in applications like object tracking, shape recognition, segmentation, edge detection and stereo matching. Figure 2.6 shows an result using an unconditionally stable numerical scheme to implement a fast version of the geodesic active contour model.

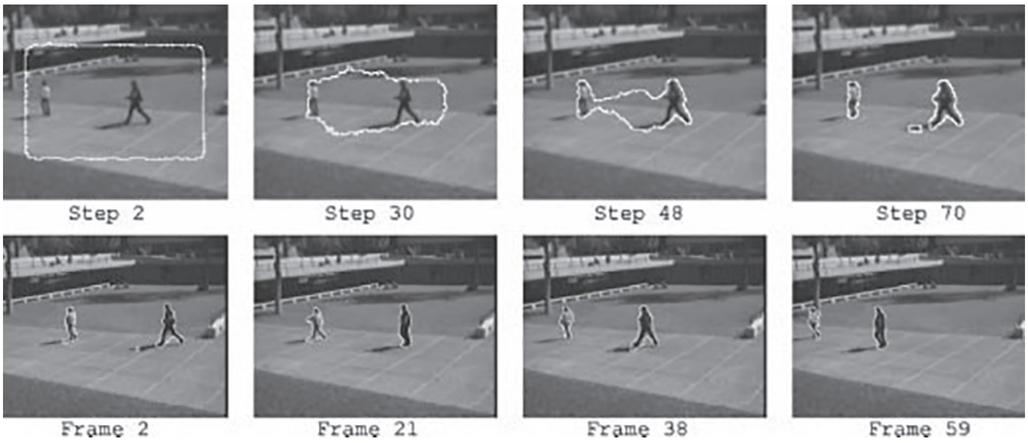


Figure 2.6 Tracking two people in a color movie

Top: curve evolution in a single frame
Bottom: tracking two walking people in a 60 frame movie

Source: Goldenberg, R., Kimmel, R., Rivlin, E., & Rudzsky, M. (2001). Fast geodesic active contours. IEEE Transactions on Image Processing, 10(10), 1467-1475.

C. Image enhancement:

Image enhancement is the process of adjusting digital images so that the results are easier to identify key features and more suitable for display or further image analysis. Such as remove noise, sharpen, or brighten an image. Figure 2.7 shows some useful enhancement methods in image processing. Image enhancement is used to adjust the contrast of images. Highlight important details in the image and improve visual quality. The gray histogram modification technique is usually used to complete image enhancement. The gray scale histogram of the image is a statistical property chart representing the distribution of gray scale in an image, which is closely connected with the contrast. If the histogram of an image is not satisfactory, it can be modified by the histogram equalization processing technique to make the image clearer.

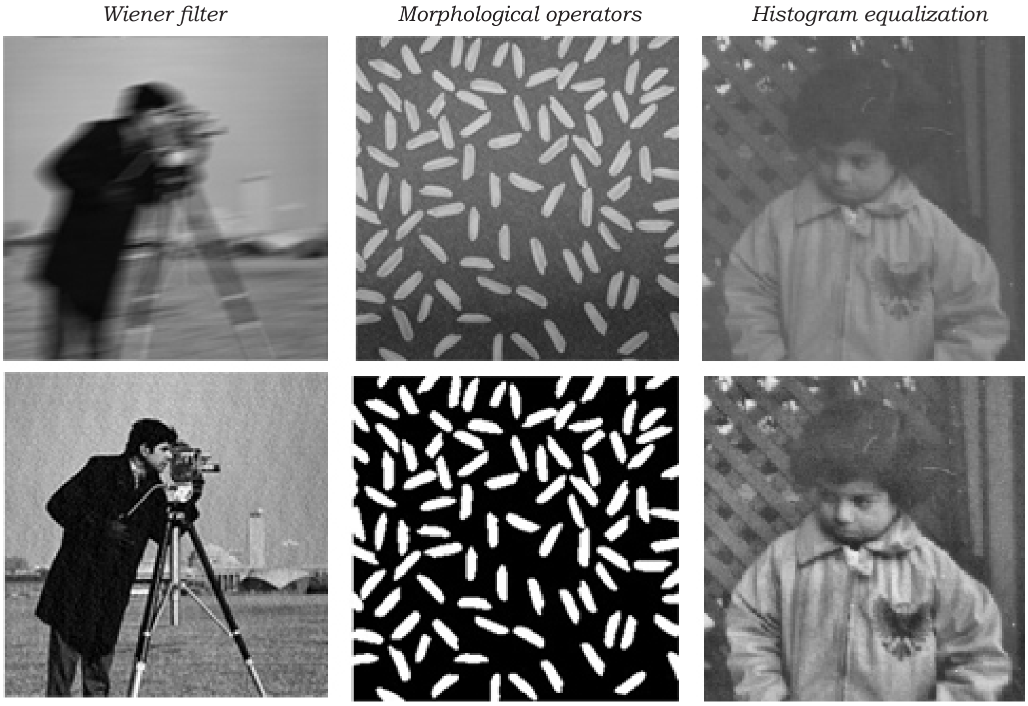


Figure 2.7 Three image enhancement methods

Source: https://cn.mathworks.com/discovery/image-enhancement.html

D. Image smoothing:

Image smoothing processing technology is the depressing noise process. The main purpose is to remove the image loss caused by imaging equipment and environment. Figure 2.8 shows how to apply different Gaussian smoothing filters to images using imgaussfilt function in Matlab. Gaussian smoothing filters are commonly used to reduce noise.

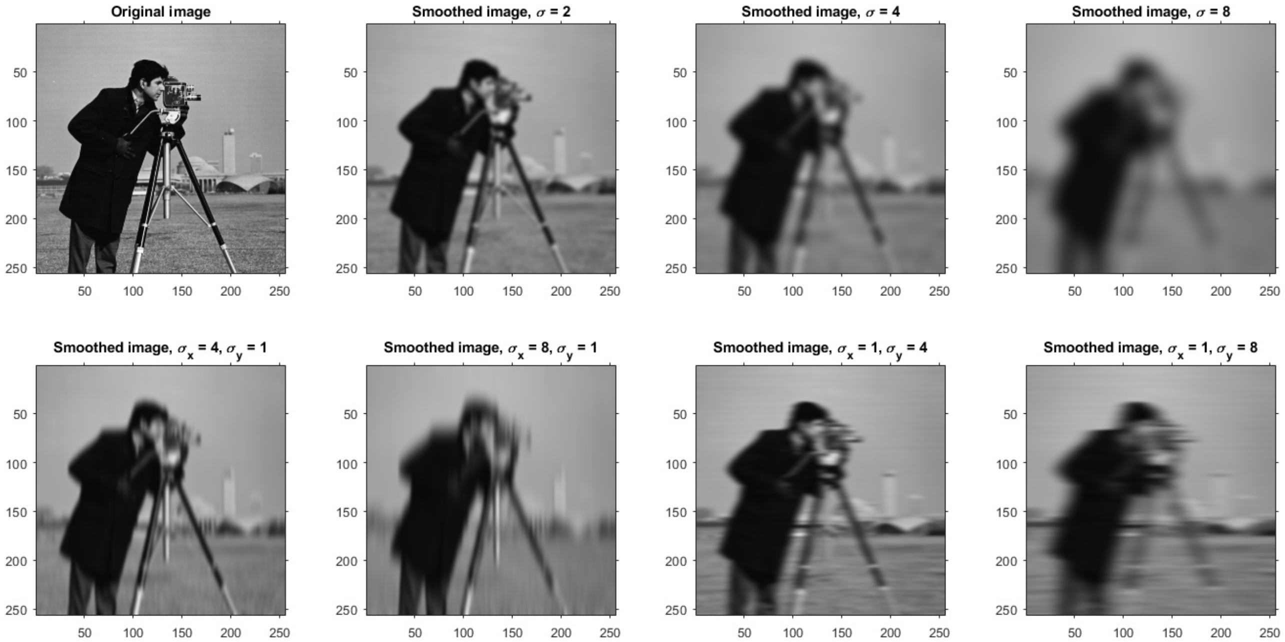

Original image
Smoothed image, σ = 2
Smoothed image, σ = 4
Smoothed image, σ = 8
Smoothed image, σ_x = 4, σ_y = 1
Smoothed image, σ_x = 8, σ_y = 1
Smoothed image, σ_x = 1, σ_y = 4
Smoothed image, σ_x = 1, σ_y = 8
50
100
150
200
250

Figure 2.8 Gaussian smoothing filters

Source: https://cn.mathworks.com/help/images/apply-gaussian-smoothing-filters-to-images.html

2.3 Incorporation of Urban Planning and Computer Vision

2.3.1 Overview of Interdisciplinary of Urban Planning and Computer Vision

As we know that Computer Vision is not a new concept, which we mentioned in chapter 2.2. Because of various limitation, such as incompletely theory, techniques and laggard hardware condition, the situation of Computer Vision field is still in the early development phase and in the process from laboratory to practice. With the great improvement of computer technique and artificial intelligence, also the parallel processing and neural network, all of these contributed to the utility progress of computer vision and other complicated visual research, thus made the computer vision entering to the booming development phase. Currently, computer vision technique is widely applied in computation geometry, computer graphics, image processing and robotology etc.

The research of computer vision is mainly focusing on the content of an image, and the information acquisition from the city images is the important part of the urban planning study. City images are the semantic carrier of criticism to social issues in modern cities, the directional context of specific background making researcher facing to the evolutional and complicated urban issues in a view of criticism to space, and realizing the cognition experience of reconstructing the urban space by imaging. However the traditional image analysis relies much on people, the staffs' professional experience and subjective judgment have a great influence on the result; and also the efficiency of collecting image data limit the depth of research: the small data capacity is easy to make misjudgments, the demand of big data leads to the collecting problem, in summary, the character of traditional city image research limits the depth and range of study. Therefore, computer vision offers a new technique and method for promoting the city image research and helps to solve the limits.

The systematic mechanism of urban planning is the bottom-up approach; it has an intensive attraction to the various information and elements in different fields. However, computer vision and Artificial Intelligence have a lot of intersections; the advanced Computer Vision is based on the support of Artificial Intelligence. The applied Computer Vision technique is processed by machine learning in the urban planning field.

2.3.2 Roles and Relationships

Computer vision has a wide application in city image, and it plays an important role in promoting the development of urban planning in an intellectualized way, strengthening the ability of data acquisition and analysis, which means Computer Vision technique as a new means of exploring the data, in particular, to complete information system in urban planning.

(1) Enrich the means of data acquisition and complete urban planning information system that includes urban form and features, environment, transport, building's quality, distribution of urban factor, changing of land use and border etc. Completing the urban planning information system needs to intensify the degree of fusion and sharing of big data in which from various fields and also pay much more attention to intelligence aggregation of multi-field, inter-discipline, software development and information technology in urban planning.

(2) Strengthen the capability of urban data analysis and management, reduce the labor wastage, looking through the targeted content of images in any time, in addition, using the pattern matching to compare related image data effectively.

(3) Intensify the urban sensibility; reveal the perceptual knowledge through a rational approach. With the development of technology, image processing has been working on the urban analysis, such as through the remote sensing image information acquisition to do the research on green planning, land use, land classification, fire prediction and so forth.

(4) Promote the integration process of data exploration and visualization. Computer Vision technology promotes the ability of visualization after data acquisition.

(5) Promotes the artificial intelligence process in urban planning. Computer Vision technology and Artificial Intelligence are interdepended and interactive, especially in urban planning; Computer Vision technology is one of the tools for data analysis, as well as an important promoter for making the urban planning into Artificial Intelligence phase. The widespread use of Computer Vision in urban planning has improved the work efficiency, and also offer an intellectual running in several urban planning fields, such as space form recognition, environment evaluation, land use monitoring and extracting features of urban form etc.

2.3.3 Result of Incorporation

In the era of big data exploration, Computer Vision technology will be an important tool for the urban planner to explore and analyze of urban data. The result of cooperation will be widely using in urban policy making, urban design guidance, the perfection of social environment, cultural interpretation and urban scene recognition etc.

Currently, the main cooperation limitation is technical exploration; besides urban planning is a subject relying on city image; thus there is less application in urban planning. The trend of these two fields' cooperation is to using deep learning (neural network) to segment image content, and results in great improvement in research efficiency and range, as well as the accuracy.

References

Davreu, R. (1978). Cities of Mystery: The Lost Empire of the Indus Valley. The World's Last Mysteries, 121-129.

Ellis, C. (2011). History of cities and city planning. Recuperado de http://www. art. net/~ hopkins/Don/simcity/manual/history. html.

Lewis, M. (1961). The city in history. San Diego, New York, London: A Harvest Book Harcourt.

Laurence, R., Cleary, S. E., & Sears, G. (2011). The City in the Roman West, c. 250 BC–c. AD 250. Cambridge University Press.

Owens, E. J. (1989). Roman town planning. Roman Public Buildings, Exeter, 7-30.

Adams, T. (1935). Outline of town and city planning.

Gibberd, F. (1967). TOWn Design, New York: The Frederick/L Praeger.

Le Dantec, J. P. (1991). For a Baroque Approach to Cities and Architecture. Architecture and Behaviour/Architecture et Comportement, 7(4).

Grant, J. (2000). Planning Canadian cities: context, continuity and change. Canadian cities in transition: The twenty-first century.

Akkerman, A. (2001). Urban planning in the founding of Cartesian thought. Philosophy & Geography, 4(2), 141-167.

Hall, P., & Tewdwr-Jones, M. (2010). Urban and regional planning. Routledge.

Sitte, C., & Stewart, C. T. (1945). The art of building cities: city building according to its artistic fundamentals. New York: Reinhold Publishing Corporation.

Roseland, M. (1997). Dimensions of the eco-city. Cities, 14(4), 197-202.

Forsyth, D., & Ponce, J. (2011). Computer vision: a modern approach. Upper Saddle River, NJ; London: Prentice Hall.

Szeliski, R. (2010). Computer vision: algorithms and applications. Springer Science & Business Media.

Debevec, P. E., Taylor, C. J., & Malik, J. (1996, August). Modeling and rendering architecture from photographs: A hybrid geometry-and image-based approach. In Proceedings of the 23rd annual conference on Computer graphics and interactive techniques (pp. 11-20). ACM.

Goldenberg, R., Kimmel, R., Rivlin, E., & Rudzsky, M. (2001). Fast geodesic active contours. IEEE Transactions on Image Processing, 10(10), 1467-1475.

Websites:

http://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_imgproc/py_pyramids/py_pyramids.html, retrieved on 13-12-2017

https://cn.mathworks.com/discovery/image-enhancement.html, retrieved on 13-12-2017

https://cn.mathworks.com/help/images/apply-gaussian-smoothing-filters-to-images.html, retrieved on 13-12-2017

Chapter Content

3 Case Study

3 Case Study

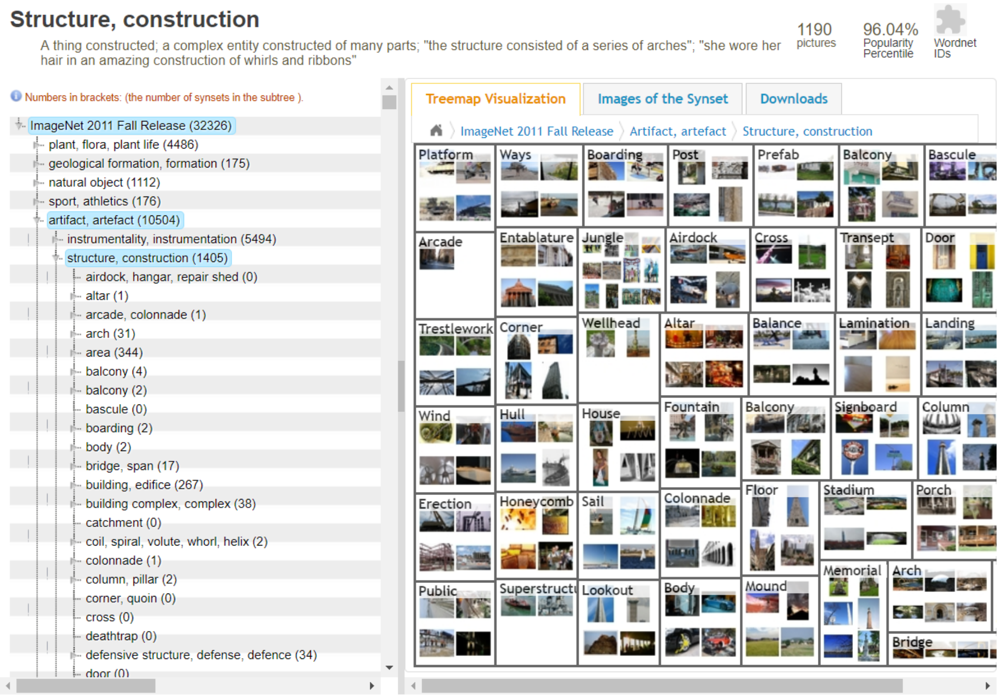
This chapter includes two cases. The first, ImageNet, the most important image annotation dataset in the past decade, also hosted the Large Scale Visual Recognition Challenge, which became the most popular competition in computer vision. It successfully promoted deep learning and the rapid development of computer vision. The second is Treepedia, aims to raise a proactive awareness of urban vegetation improvement, using computer vision techniques based on Google Street View. Treepedia provides a way to use computer vision techniques to study urban environments.

3.1 ImageNet and ILSVRC

> *"ImageNet is an image database organized according to the WordNet hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images. Currently we have an average of over five hundred images per node. We hope ImageNet will become a useful resource for researchers, educators, students and all of you who share our passion for pictures."*
>
> *"ImageNet is an ongoing research effort to provide researchers around the world an easily accessible image database."*
>
> *(Online via http://image-net.org, retrieved on 01-12-2017)*

The ImageNet project is a large visual database designed for use in visual object recognition software research. The database of ImageNet was proposed firstly in an article of CVPR 2009 (Deng, J., Li, K., Do, M., Su, H., & Fei-Fei, L. (2009). Construction and analysis of a large scale image ontology. Vision Sciences Society, 186, 2.) and replaced the PASCAL database (which was less diversity) and LabelMe database (which was less normalization).

ImageNet, not only is an important promoter for the development of Computer Vision, but also one of the key driving force for the Deep Learning fervor. Until 2016s, ImageNet has included over



Structure, construction

A thing constructed; a complex entity constructed of many parts; "the structure consisted of a series of arches"; "she wore her hair in an amazing construction of whirls and ribbons"

1190 pictures | 96.04% Popularity Percentile | Wordnet IDs

Numbers in brackets: (the number of synsets in the subtree).

- ImageNet 2011 Fall Release (32326)
  - plant, flora, plant life (4486)
  - geological formation, formation (175)
  - natural object (1112)
  - sport, athletics (176)
  - artifact, artefact (10504)
    - instrumentality, instrumentation (5494)
    - structure, construction (1405)
      - airdock, hangar, repair shed (0)
      - altar (1)
      - arcade, colonnade (1)
      - arch (31)
      - area (344)
      - balcony (4)
      - balcony (2)
      - bascule (0)
      - boarding (2)
      - body (2)
      - bridge, span (17)
      - building, edifice (267)
      - building complex, complex (38)
      - catchment (0)
      - coil, spiral, volute, whorl, helix (2)
      - colonnade (1)
      - column, pillar (2)
      - corner, quoin (0)
      - cross (0)
      - deathtrap (0)
      - defensive structure, defense, defence (34)
      - door (0)

Treemap Visualization | Images of the Synset | Downloads

ImageNet 2011 Fall Release 〉 Artifact, artefact 〉 Structure, construction

Figure 3.1 Some examples of "construction" in the ImageNet dataset

http://image-net.org/explore.php

15 millions URL of a manually annotated image, which is labeled image; and the label describes the image content and there are more than 22 thousand categories. At least, there are 1 million images with bounding box among of them.

Since the 2010s, ImageNet has held annual software competition, which is ImageNet Large Scale Visual Recognition Challenge (ILSVRC); and the champion program has the highest accuracy for the classification and recognition of object and scene.

> *This challenge evaluates algorithms for object localization/detection from images/videos at scale.*
>
> *Object localization for 1000 categories.*
>
> *Object detection for 200 fully labeled categories.*
>
> *Object detection from video for 30 fully labeled categories.*
>
> *(Online via http://image-net.org/challenges/LSVRC/2017/index, retrieved on 01-12-2017)*

ILSVRC evaluates mainly on the effect of the algorithm to object detection and image classification in large scale. One of the goals of competition is through a mass of manual labeled trained data to stimulate the researcher to compare their algorithm in the detected effect in the various object; besides, another aim is to check out the improvement of Computer Vision technology apply in image retrieval and image annotation in large scale.

In 2012s, Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton invented a “large deep convolutional neural network”, which was AlexNet and was famous for the champion of ILSVRC in the same year, also its performance was the best ever. Some methods mentioned in the article, such as data enhancement and dropout, which was still useful, and the article “ImageNet Classification with Deep Convolutional Networks” was cited seven thousand times and regarded as one of the most important articles in this field, which showed the advantages of CNN for real and is supported by the record-breaking grade in competition.

2012 is an important year that CNN firstly realized the 15.4% error rate in Top 5, and the error rate of the second one is 26.2%. It was shocked the whole Computer Vision field, in another word, CNN has become a widely known name.

After that, many network structures have had a good performance in this competition, such as VGG, GoogleNet, Deep Residual Network etc. Until to the 2016s, the error rate of image recognition achieved by around 2.9%, which was less than human (5.1%), as well as the meaning of competition. Currently, no matter image classification, object detection and object recognition, the correct rate has been exceeded far away than human. The problem of Computer Vision in perceptive aspect is solved in a good way.

The CVPR 2017 seminar “Beyond ImageNet Large Scale Visual Recognition Challenge” announced that 2017 is the last year for the ImageNet competition. It is no doubt that this represents an end of the era, but also a beginning for the new trip, in future the main point of Computer Vision will be image understanding, and one of the candidate, WebVision Challenge, which was proposed jointly by Swiss Federal Institute of Technology Zurich (ETU Zurich) and Google etc. and it will be held in 2017 and the main content will focus on the study and understanding of network data.

3.2 Treepedia

Treepedia is duly credited as a project by the MIT Senseable City Lab in 2016. It focuses on exploring the green canopy in cities around the world. Treepedia measures the canopy cover in cities, rather than count the individual number of trees. They analysis the amount of green perceived while walking down the street. The group developed a metric—the Green View Index—by which to evaluate and compare canopy cover.

> *“The Green View Index (GVI) was calculated using Google Street View (GSV) panoramas. This method considers the obstruction of tree canopies and classifies the images accordingly. By using GSV rather than satellite imagery, we represent human perception of the environment from the street level. The GVI presented here is on a scale of 0-100, showing the percentage of canopy coverage of a particular location. Explore the maps above to see how the GVI changes across a city, and how it compares across cities and continents.”*
>
> *(Online via http://senseable.mit.edu/treepedia, retrieved on 01-12-2017)*
>
> *GVI calculation:*
>
> *Yang et al. (2009) proposed a “Green View” index to evaluate the visibility of urban forests. Their GVI was defined as the ratio of the total green area from four pictures taken at a street intersection to the total area of the four pictures.*
>
> *The Trepedia group modified GVI calculation formula, using six images covering the 360◦ horizontal surroundings to calculate the index for*

each sample site along streets.

Furthermore, to effectively represent the surrounding greenness that pedestrians can see, one vertical view angle is apparently insufficient. Three different vertical view angles were considered at each direction for calculating the GVI in this study.

Consequently, the final modified GVI used in this study was actually calculated using 18 GSV images for each site.

The modified GVI calculation formula is written as:

$$\text{Green View} = \frac{\sum_{i=1}^{6}\sum_{j=1}^{3} Area_{g\_ij}}{\sum_{i=1}^{6}\sum_{j=1}^{3} Area_{t\_ij}} \times 100\%$$

where $Area_{g\_ij}$ is the number of green pixels in one of these images captured in six directions with three vertical view angles (−45◦, 0◦, 45◦) for each sample site, and $Area_{t\_ij}$ is the total pixel number in one of the 18 GSV images.

(Source: Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., & Zhang, W. (2015). Assessing street-level urban greenery using Google Street View and a modified green view index. Urban Forestry & Urban Greening, 14(3), 675-685.)

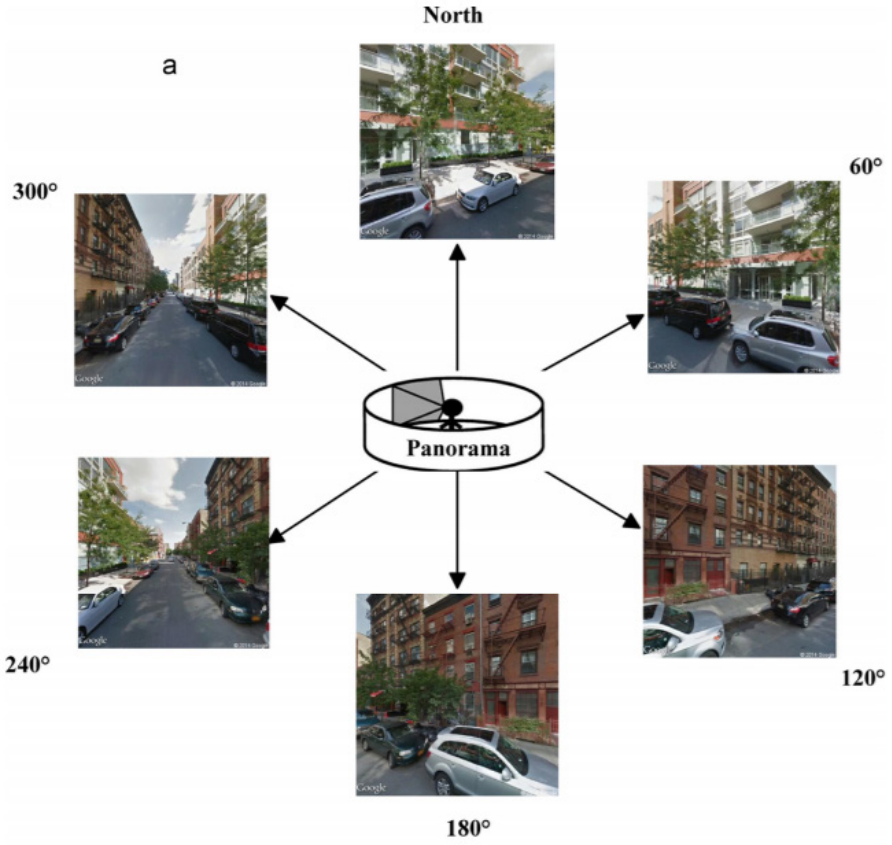


Figure 3.2.a GSV images captured in six directions at a sample site in the study area

source: Assessing street-level urban greenery using Google Street View and a modified green view index, Urban Forestry & Urban Greening 14 (2015) , 678-680

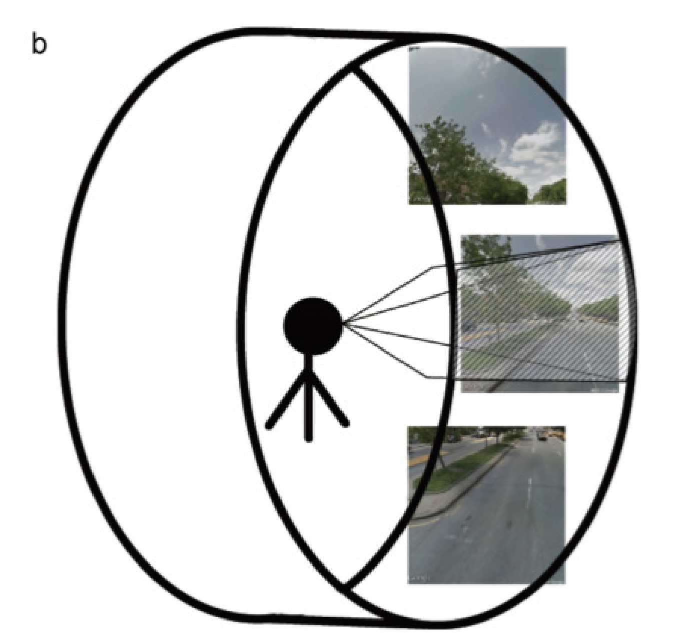


Figure 3.2.b GSV images captured at three vertical view angles at a sample site

source: Assessing street-level urban greenery using Google Street View and a modified green view index, Urban Forestry & Urban Greening 14 (2015) , 678-680

Treepedia aims to raise a proactive awareness of urban vegetation improvement, using computer vision techniques based on Google Street View images (http://senseable.mit.edu/treepedia, retrieved on 01-12-2017). So the visualization maps street-level perception only and parks aren't included.

The maps from Treepedia website[1] show how the GVI changes across a city, and how it compares across cities and continents. The group has been developing the database to span cities all over the world. Now, the database includes 27 cities around the world.

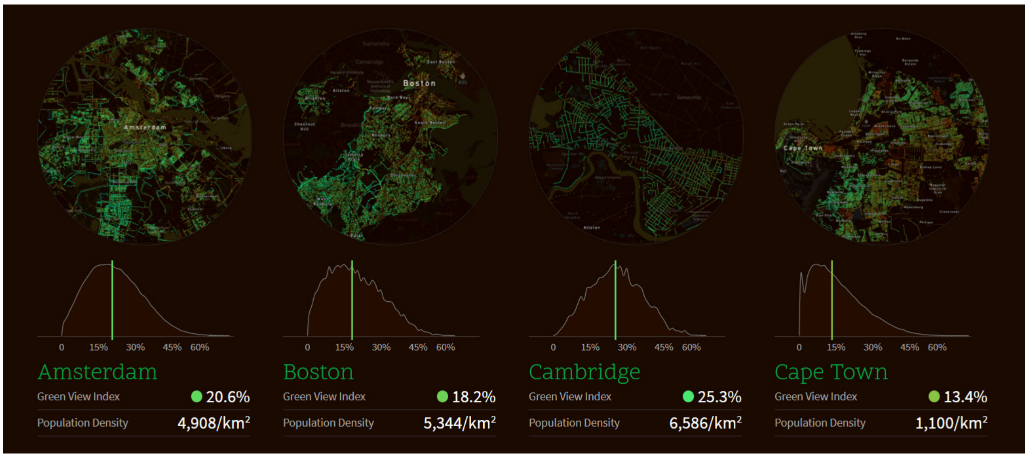


Figure 3.3 Some examples of GVI used in cities

http://senseable.mit.edu/treepedia

[1] *http://senseable.mit.edu/treepedia, retrieved on 01-12-2017*

References

Deng, J., Li, K., Do, M., Su, H., & Fei-Fei, L. (2009). Construction and analysis of a large scale image ontology. Vision Sciences Society, 186, 2.

Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., & Zhang, W. (2015). Assessing street-level urban greenery using Google Street View and a modified green view index. Urban Forestry & Urban Greening, 14(3), 675-685.

Websites:

http://image-net.org, retrieved on 01-12-2017

http://image-net.org/challenges/LSVRC/2017/index, retrieved on 01-12-2017

http://image-net.org/explore.php, retrieved on 01-12-2017

http://senseable.mit.edu/treepedia, retrieved on 01-12-2017

Chapter Content

4 Technology Exploration

4 Technology Exploration

This chapter mainly introduces our exploration and realization of urban street analysis technology. The chapter is divided into seven parts: first, introduce the Semantic Segmentation Technology and how to achieve Urban Street Analysis (UrSA); then, introduce the specific content and function of the UrSA, including Image Database Acquisition, Image Annotation, and Classification, Semantic Segmentation Using SegNet and Model Assessment and perfection. Final is three examples of UrSA implementation by visualization.

4.1 Semantic Segmentation Technology

4.1.1 The Image Recognition and Classification

Comparing with character, an image is more readable and vivid; both of them are the important source of delivering and exchanging information from one to others. Our target is analyzing the spatial feature of the urban street through the image semantic segmentation, which means to classify the different urban street through its images segmented by semantic information.

Normally, the first step of image classification is to do the image description by manual labeling feature or method of feature learning, second, using the classifier to distinguish the type of object, hence how to extract features is essential in this process.

Before the deep learning in the algorithm, the method of classification using frequently is based on the model of Bag of Words. The Bag of Words is introduced from Natural language processing, which means the features of a sentence are represented by a bag of “words” which is vocabularies, phrases or characters. The bag-of-words model is commonly used in methods of document classification where the (frequency of) occurrence of each is used as a feature for training a classifier. In terms of image, to operate this method, we need to construct a “dictionary”, and the simplest model framework could be designed to three processes: low-level feature extraction, feature coding and design of classifier.

However the method of classification based on deep learning replaced work of manual design or selecting image feature, the former learns the hierarchical description of feature by supervised learning or unsupervised learning. Convolutional Neural Network (CNN), one of the deep learning model, has achieved an amazing breakthrough in the field of computer vision. CNN using image pixel information as input, which is to keep all the information from input images on the maximum level; extracting feature and high-level abstract by convolutional calculation; the output of the model is just the result of image recognition. This learning method based on “input-output” and end-to-end is very effective and has been widely applied.

Before 2012s the traditional classified method can be achieved by three steps that mentioned before, but in general, to construct a whole image recognition model includes low-level feature learning, feature coding, spatial constraint, the design of classifier, model ensemble and etc.

> *1). Low-level feature extract: in general extracting a mass of local feature according to fix step length and size from image. The frequently used of local feature includes SIFT[1] (Scale-Invariant Feature Transform); HOG[2] (Histogram of Oriented Gradient), LBP[3] (Local Binary Pattern) and etc., also using various feature descriptor in case of over losing effective information.*
>
> *2). Feature coding: the low-level feature includes a numerous redundancies and a noise; for the sake of improving the robustness of feature expression needs to code low-level feature by feature transform, which is feature coding. The common using of feature coding are vector quantitation coding[4], sparse coding[5], locally linear feature encoding[6], Fisher vector coding[7] and etc.*
>
> *3). Spatial feature constraint: after feature coding, spatial feature constraint appears that called feature convergence. Feature convergence refers to take the maximum or average value for each dimension in a space can obtain the non-deformation feature expression. Spatial Pyramid Matching is one of the common using feature convergence methods; it is came up with the idea that divide image evenly and uses feature convergence in every part of image.*
>
> *4). Classification by classifier: passing by the former process, an image can be descripted by fix-dimension vector, and then classifier works continually. The common classifiers are SVM[8] (Support Vector Machine), Random Forests and etc., but the SVM based on kernel method is widest using classifier and good performance on traditional image classification task.*

This method is applied widely in the image classification algorithm of PASCAL VOC competition. NEC laboratory employed the features of SIFT and LBP in the ILSVRC2010, as well as non-linear encoder and SVM classifier, then they were the champion in the image classification.

[1] *Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. International journal of computer vision, 60(2), 91-110.*

[2] *Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on (Vol. 1, pp. 886-893). IEEE.*

[3] *Ahonen, T., Hadid, A., & Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. IEEE transactions on pattern analysis and machine intelligence, 28(12), 2037-2041.*

[4] *Sivic, J., & Zisserman, A. (2003, October). Video Google: A text retrieval approach to object matching in videos. In null (p. 1470). IEEE.*

[5] *Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1?. Vision research, 37(23), 3311-3325.*

[6] *Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., & Gong, Y. (2010, June). Locality-constrained linear coding for image classification. In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on (pp. 3360-3367). IEEE.*

[7] *Perronnin, F., Sánchez, J., & Mensink, T. (2010). Improving the fisher kernel for large-scale image classification. Computer Vision–ECCV 2010, 143-156.*

[8] *Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine learning, 20(3), 273-297.*

In 2012s, the CNN model that has been introduced by Alex Krizhevsky in ILSVRC achieved a history-making breakthrough, the effect was much more than the traditional method and he won the champion in ILSVRC2012, therefore this model was named AlexNet. In addition, this was the first time that deep learning was employed in image classification on a large scale.

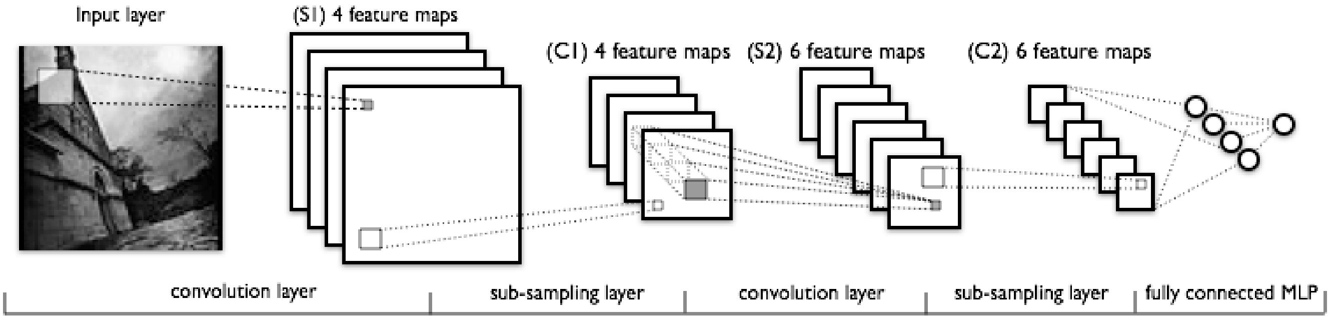


Figure 4.1 The Full CNN Model: LeNet

Source: http://deeplearning.net/tutorial/lenet.html

The traditional CNN contains convolution layer, fully-connected layer and so forth, and employs softmax, a multi-classifier, and the loss function. Figure 4.1 shows a typical Convolution Neural Network.

Common components for CNN construction:

Convolution layer: implementing convolutional calculate for extracting the features from bottom layer to top layer, and unveil the property of local relationship and space invariance feature.

Pooling layer: operating down sampling, which means takes max-pooling or avg-pooling in locality of output image. Down sampling is also a common operation in image processing, which can filter the unimportant high-frequency information.

Fully-connected layer/fc layer: all neuron in from input layer to hidden layer are connected.

Non-linear variation layer: in general, after convolution layer and fully-connected layer, non-linear variation layer followed, such as, Sigmoid, Tanh, and ReLu, etc., which is to enhance the ability of expression of network, in CNN, the ReLu is the most using function.

Dropout: in the process of model training, making some hidden layer nodes weight unused in randomly for improving the ability of generalization and to avoid over-fitting.

Moreover, in the model training, each layer's parameters are updating continuously, so the input distribution varies that needs to design the hyper-parameter in training process. In 2015s, Sergey Ioffe and Christian Szegedy put forward the Batch Normalization (BN)[9] algorithm; each batch uniforms each layer's feature so that the input distribution is relatively stable. BN algorithm plays a role in both regularization and weakening the hyper-parameter design. After experimental verification, BN algorithm speeds up the convergence procedure, as well as applied widely in some deep model.

4.1.2 The Classic CNN Model - VGG

After Alex Krizhevsky came up with the AlexNet model, a numerous CNN models have obtained a great grade in ImageNet. The deeper model is, the less error rate it has, the figure 4.2 shows that the error rate in top5 reduced to the 3.5% around. However, in the same dataset from ImageNet, the error rate of identifying objects by human eyes is around 5.1%, which means the ability of recognition in deep learning model is exceeding in human eyes. The most classical CNN models are VGG[10], GoogleNet[11], and ResNet[12] (Residual Network); they have their own advantages and characteristics in the neural network structure. No matter what VGG, GoogleNet and ResNet, their recognition ability is beyond human eyes and have a high accuracy.

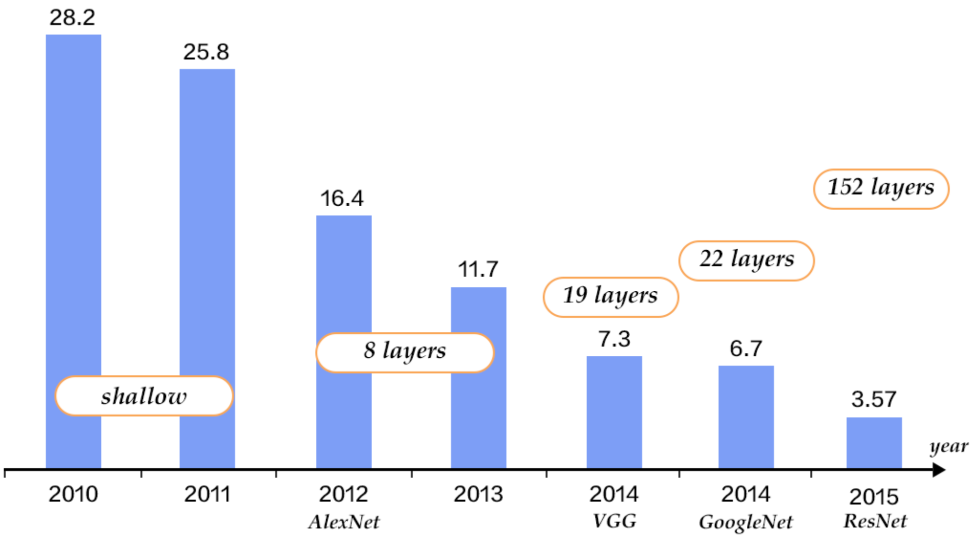


Figure 4.2 ILSVRC Image Classification Top-5 Error Rate

Source from ImageNet

[9] Ioffe, S., & Szegedy, C. (2015, June). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International Conference on Machine Learning (pp. 448-456).

[10] Chatfield, K., Simonyan, K., Vedaldi, A., & Zisserman, A. (2014). Return of the devil in the details: Delving deep into convolutional nets. arXiv preprint arXiv:1405.3531.

[11] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

[12] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

ford proposed a model in ILSVRC, it was named VGG model. The VGG model architecture, presented by Simonyan and Zisserman, is described in the paper on Very Deep Convolutional Networks for Large Scale Image Recognition.

VGG model has a more deep and wider network structure comparing to the old. Its core has five groups of convolutional operation, and every two groups have Max-Pooling and termspacereduction. 3X3 convolutional operation in the same group, the number of convolution kernel is from 64 in the shallow group to 512 in the deep group, and the same number of convolution kernel in a group. After convolution, it links with two fully connected layers and then classification layer. Because of the different convolution layer, it has 11, 13, 16, and 19 in models; figure 4.3 shows a 16 layers model. VGG model is relatively simple in structure, also there is much research based on it, for instance, the first public model that better than human eyes took VGG model's structure as a reference.

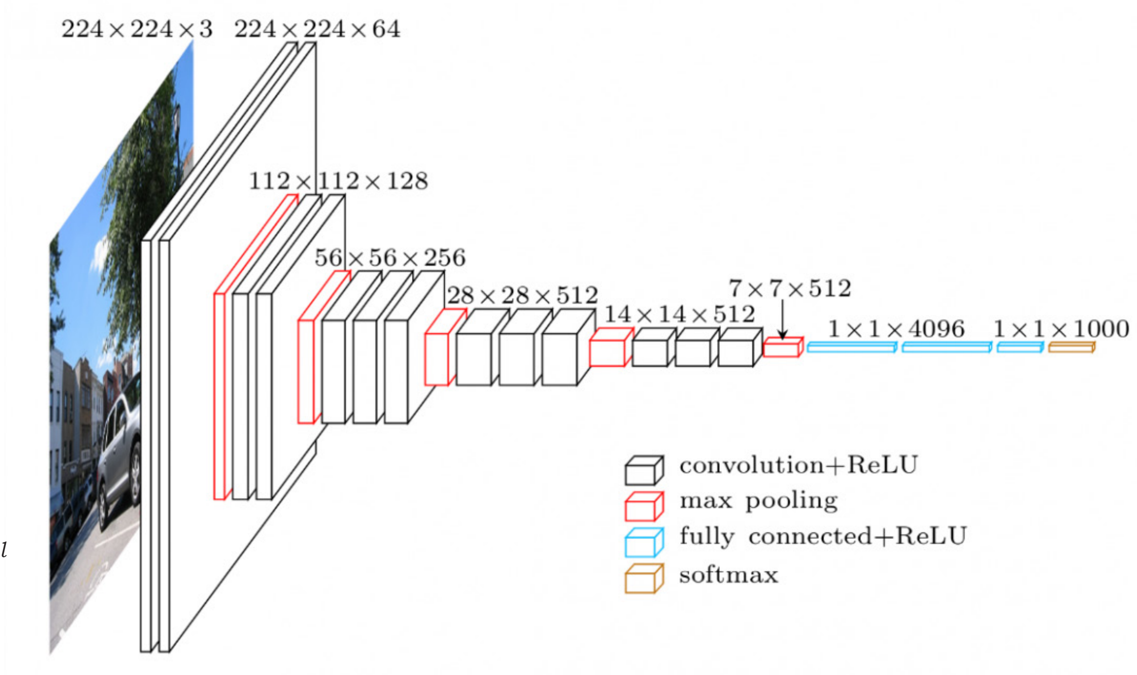


Figure 4.3 VGG16 Model

Source: https://www.cs.toronto.edu/~-frossard/post/vgg16/

4.1.3 From VGG16 to SegNet[13]

Semantic segmentation is a method to understand an image at the pixel level, in simple terms, we assign each pixel in the image an object class. SegNet, one type of convolutional neural network (CNN) was designed for semantic image segmentation.

SegNet, proposed by Cambridge, is an open-source deep web of image semantic segmentation, which aim is to solve automatic drive and intelligent robot. SegNet is based on FCNN and a result of modification of VGG-16. There are two kinds of SegNet, normal and Bayesian version, at the same time, the author of SegNet offered a basic version according to the deep of network.

Figure 4.4 shows the structure of SegNet, "Input" represents original image and "Output" represents the image after segmentation, different colors represent classification. Semantic segmentation not only recognizes the object, more important it also reflects the position of the image. We can see the asymmetric network, divided by green pooling layer and red upsampling layer in the middle, the left part is Encoder named by SegNet's author, which is extracting the high dimension features through convolution, and diminish image by pooling. The right side is de-convolution (here the convolution and de-convolution are the same) and upsampling, the effect of de-convolution is to reappear the feature after classification, and upsampling larges the image, which is called Decode. Final, through the Softmax, output the maximum value among the different classifications.

Convolution: during the Encoder process, convolution has a role in extracting features, which does not change the size of image and the convolution used in SegNet is the same with the convolution used in traditional CNN.

Batch Normalisation: the main effect of standardization is to accelerate learning speed, using before the function. In SegNet, before each convolutional layer it is BN layers and ReLU active layers.

ReLU: ReLU is the modification of traditional active function sigmoid, and main effect is to solve gradient problem.

Pooling: pooling is a means to diminish image in a half, there are two approaches in common: max pooling and mean pooling.

Upsamping: upsamping is a inverse process of Pooling, is to large image in 2 fold.

[13] *Badrinarayanan, V., Kendall, A., & Cipolla, R. (2015). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. arXiv preprint arXiv:1511.00561.*

Deconvolution: during the Decoder process, deconvolution works on padding missing content, also using "same" convolution, which to enrich the information in upsampling image, thus the missing information in pooling process can obtain in Decoder process.

Output: the last convolution output all the classification, and the network adds a softmax layer. Because the softmax calculate the maximum probability in each pixel among all the classification, which is as a label, ultimately, it has a classification result in image pixel level.

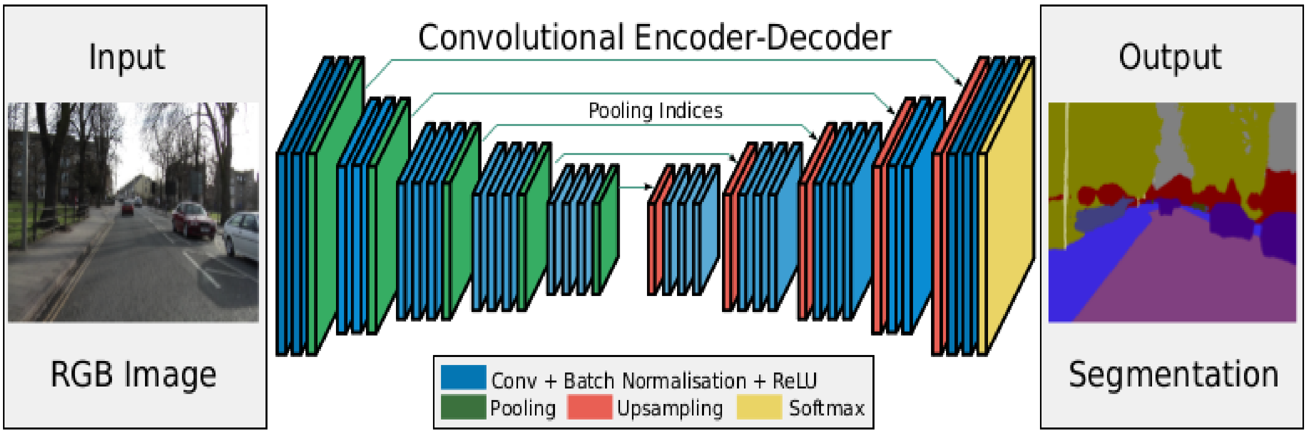

Input
RGB Image
Convolutional Encoder-Decoder
Pooling Indices
Output
Segmentation
🟦 Conv + Batch Normalisation + ReLU
🟩 Pooling 🟥 Upsampling 🟨 Softmax

Figure 4.4 The SegNet Model

Source: SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation

4.2 Technology Practice Framework

A clear working framework is a prerequisite for carrying out specific projects. In order to achieve Urban Street Analysis (UrSA), we have done a lot of technical exploration, summed up an effective framework, including working path and technology system.

4.2.1 UrSA Technology Exploration

Our exploration of technology focused on image acquisition, image processing, Semantic Segmentation and image analysis and visualization, which formed the core technology box of UrSA. Figure 4.5 shows the detail of technology framework.

Image acquisition refers to the process of collecting images in a specific city or region. The amount and range of image acquisition can be controlled. In this work, the captured images specifically refer to the Google Street View (GSV) images.

Traditional image processing refers to the use of computer algorithms to perform digital image analyzing and processing to meet the special requirements. In this work, it includes annotate the image contents with multi-class labels in pixel level as CNN training database and extracts the color information from segmented images. Image processing does typically involve filtering or enhancing an image using various types of functions in addition to other techniques to extract information from the images.

In computer vision, semantic image segmentation is the process of partitioning a digital image into multiple segments by assigning a label to each pixel in an image so that pixels with the same label share certain features. Image segmentation is typically used to locate objects and boundaries in images. Currently, CNN is widely used for semantic image segmentation.

The semantic segmentation technology can provide a database with segmented images, the image processing technology can extract the digital information from the database. Image analysis focuses on using mathematical models and statistical methods to analyze, interpret and recognize image information with some intelligence. Data visualization and image analysis are inextricably linked, and excellent visualization can convey the analysis results intuitively and effectively to the users.

4.2.2 The Main Work of UrSA

The neural network learns image features and outputs the feature image. After obtaining the features information from images, we can perform the street-level UrSA using the GSV images.

The main work :

- *Data Acquisition: Download street level city image via GSV's API port and name the file according to 'City + Coordinates + Camera Angle + .jpg' and save it to the corresponding folder.*
- *Processing Data: The pre-processed and manually annotated images are used as the basic data for neural network training and testing.*
- *Select Semantic Image Segmentation Network: SegNet (based on VGG16).*
- *Train Data: Train 80% of images with annotation randomly and obtain the feature information model.*
- *Test Data: Enter the remaining data and output the test results.*
- *Evaluation: Through compare the test results with ground truth results, we can evaluate the learning effects, and identify problems.*
- *Optimize data and adjust parameters*
- *Test and evaluation again*
- *Final output result and visualize*

4.2.3 Computer Technology System for UrSA

The technical system refers to the various interrelated technologies in society constitute a whole, according to a certain purpose and a certain structure. Computer Technology System for UrSA refers to a variety of computer technologies according to the working path to achieve UrSA.

> *The main technology category:*
>
> *Image Acquisition Technology, Image Processing (Annotation) Technology, Semantic Segmentation Technology (VGG16, SegNet), Evaluation Technology, Statistical Technology, Visualization Technology.*

4.2.4 The Framework of UrSA

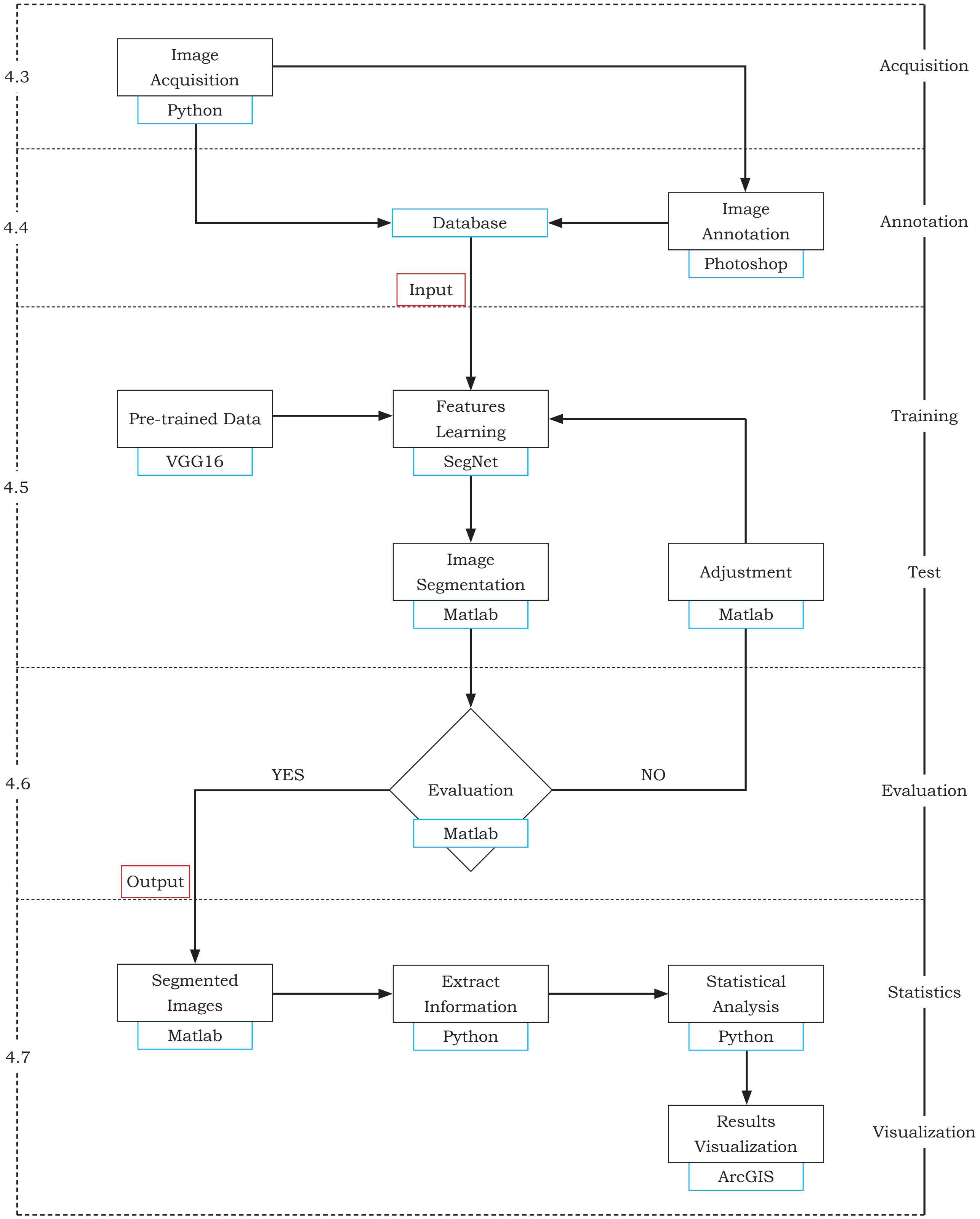
The framework is a workflow which integrates the technology system and main works. It is purpose-oriented and arranges corresponding computer technology in each phase of work.

The workflow of UrSA:

> *1) Image Database Acquisition: In this step, we make a GSV image database including the original GSV images and the annotation images. Each original image is labeled by 11 classes correspond 11 colors as annotation image.*
>
> *2) Input data to the CNN model.*
>
> *3) Training: After inputting the data, CNN model will learn the image features from the database (80%), we call the step as "training". The CNN model is named SegNet, a specific semantic segmentation model. The SegNet model build by a Pre-trained VGG16 model with over 1.2 million data from ImageNet.*
>
> *4) Test: In test step, we input the other images to the model, and get segmented images.*
>
> *5) Evaluation and Adjustment: In order to evaluate the accuracy of CNN model, we compare each segmented image with corresponding annotation image. If the accuracy is low, we adjust the parameters in the SegNet model until get a excellent result. This is an iterative process.*
>
> *6) Output segmented images through the trained SegNet model. Each image is segmented by 11 classes correspond 11 colors.*
>
> *7) Statistics and Visualization: Through count the amount of color in segmented images, we can extract the classes information in each image. So we can analysze the status and distribution of different classes in specific areas in cities. The details shown through visualization techniques.*

The specific work flow will be described in detail later. Figure 4.5 shows the framework of UrSA.

Figure 4.5 The Framework of UrSA



4.3 Image Database Acquisition

In the age of digital information, the traditional textual information cannot meet the requirements of the urban planning study for huge data and precise analysis. Because of the limitations of data acquisition technology, urban planners can hardly study the characteristics of the city in the street-level. As Google opens up the APIs for street view images, it becomes possible for planners to obtain large and comprehensive urban street image data. If used the images directly in the analysis of the city, the effect is so limited that we need to extract the features information of images. Computer vision technology provides the possibility of extracting image features. Before that, we need to prepare the basic database for training and test. We take the image data obtained from GSV as the original image database, and then take the processed and labeled images data as the label database.

4.3.1 Data Sources: Google Street View Images API

The first step in the analysis of urban scenes is to obtain the image data and Google Street View is the most comprehensive urban streetscapes website. Google opens the Google Street View API to the public, the user can simply enter the URL and parameters to get the specific image, for users of the standard API, 25,000 times per day, 640 x 640 maximum image resolution. Figure 4.6 shows the example of acquiring images from GSV.

Acquiring images from Google Street View Image API :

1) Get API KEY: AIzaSyC8_qoTIudMzCMSWpdzhhHnbsjo2FMWE4o

2) URL: https://maps.googleapis.com/maps/api/streetview?parameters

3) URL Parameters:

Required parameters:

location: can be either a text string (such as Chagrin Falls, OH) or a lat/lng value (40.457375,-80.009353).

or pano: is a specific panorama ID. These are generally stable.

size: specifies the output size of the image in pixels. Size is specified as {width}x{height} - for example, size=600x400 returns an image 600 pixels wide, and 400 high.

key: allows you to monitor the Application's API usage in the Google API console.

Optional parameters:

signature (recommended): is a digital signature used to verify that any *site generating requests using your API key is authorized to do so*

heading: indicates the compass heading of the camera.

fov (default is 90): determines the horizontal field of view of the image.

pitch (default is 0): specifies the up or down angle of the camera relative to the street view vehicle.

Example:

Address1: Piazza Duca d'Aosta, 20125 Milano

URL1: https://maps.googleapis.com/maps/api/streetview?size=600x600&location=45.484803,9.202369&heading=90&pitch=10&key=%20AIzaSyC8_qoTIudMzCMSWpdzhhHnbsjo2FMWE4o

Address2: Piazza del Duomo, 20121 Milano

URL2: https://maps.googleapis.com/maps/api/streetview?size=600x600&location=45.463737,9.190050&heading=90&pitch=15&key=%20AIzaSyC8_qoTIudMzCMSWpdzhhHnbsjo2FMWE4o

Address3: Piazza Leonardo Da Vinci, 20131 Milano

URL3: https://maps.googleapis.com/maps/api/streetview?size=600x600&location=45.479294,9.226501&heading=90&pitch=20&key=%20AIzaSyC8_qoTIudMzCMSWpdzhhHnbsjo2FMWE4o

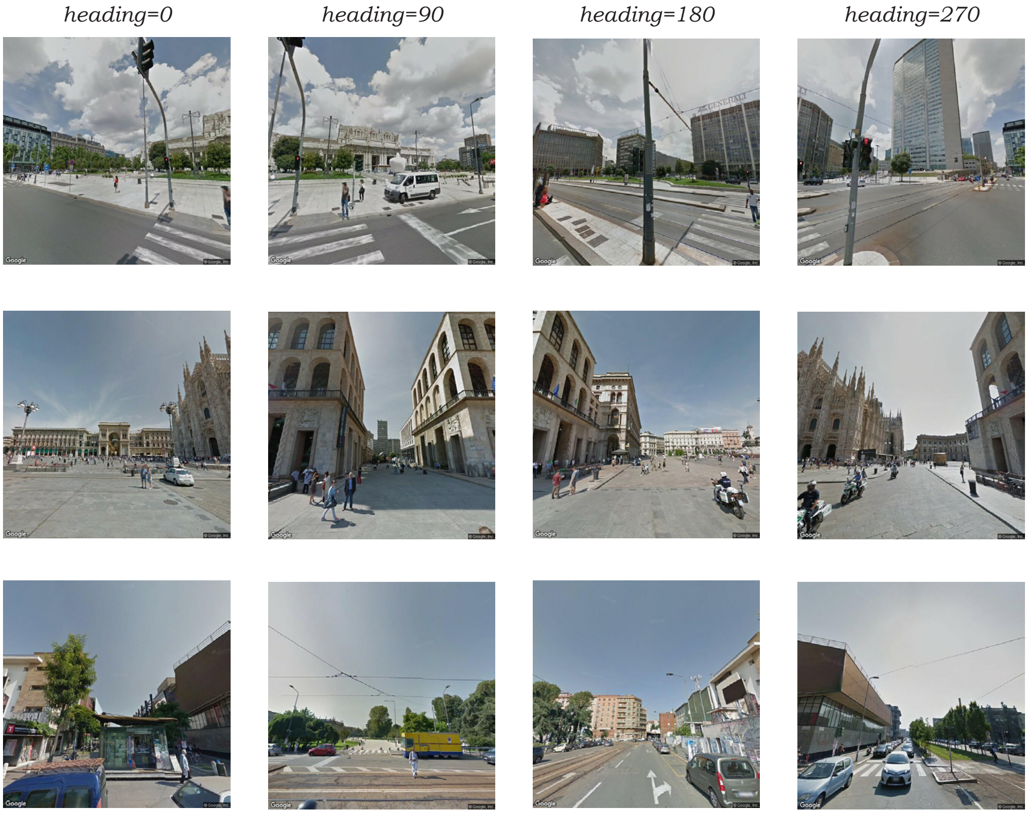


Figure 4.6 The example of acquiring images from GSV

Address1
Milano Centrale
size=600x600, pitch=10
location=
45.484803,9.202369

Address2
Duomo di Milano
size=600x600, pitch=15
location=
45.463737,9.190050

Address3
Polimi
size=600x600, pitch=20
location=
45.479294,9.226501

4.3.2 Image Standard

UrSA is used to study the urban street features from Google Street View Images. Image size is one of the important parameters of image data. The size of the image considers the computational ability of the computer and the requirements of the neural network framework. According to the VGG16 model's performance in the Large Scale Visual Recognition Challenge (ILSVRC), 360*480 is suitable.

The standard of image size:

*1) each image: PNG format, size is 360*480 pixels, aspect ratio is 4:3.*

2) naming rules: latitude + longitude + camera angle (heading) + .png

When pitch = 7, the image fits the human perspective. So we set the pitch parameter to 7 and the four camera angles (0, 90, 180, 270) also ensure the integrity of the same location information collection.

Finally, the image sample looks like this:

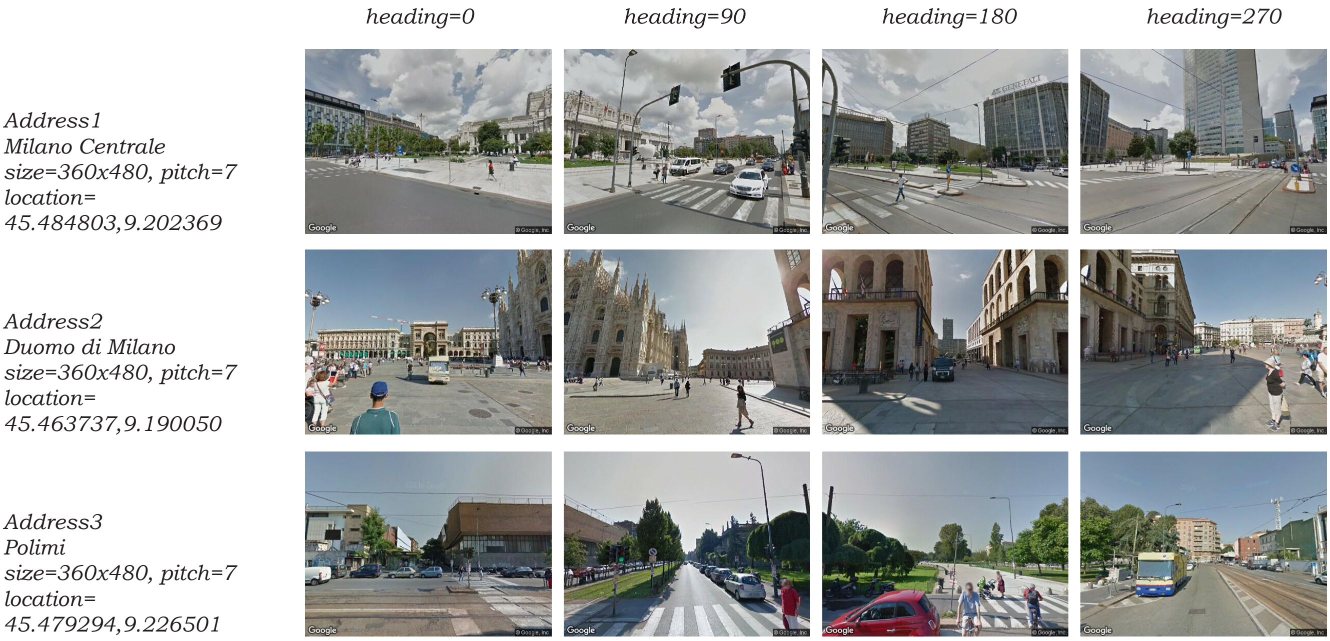


Figure 4.7 The example of standard image

4.3.3 Acquiring Images in Specific Area

The amount of data that UrSA requires is enormous, so we write a GSV Images acquisition program in Python, which can request street view images in a specific area. This work provides two methods of image acquisition - request images in a circle or in a polygon. The Python code references Lezhi Li's study in the Streetview (https://github.com/Firenze11/cv_streetview/tree/master/scripts).

In both approaches, there is a very important parameter - step. "Step" refers to the difference between the reference coordinate value and the new coordinate value, and each new coordinate value becomes the reference value of the next coordinate value. In other words, the difference between adjacent coordinate values is the same, equal to the value of step. The step determines the number of coordinates to get within a specific range. By changing the step size, we can control the number of images downloaded.

In computational geometry, the point-in-polygon (PIP) problem asks whether a given point in the plane lies inside, outside, or on the boundary of a polygon. One of the best solutions to this issue is the "Ray Casting Method".

The method can be described as drawing an imaginary line from the point in question and stop drawing it when the line leaves the polygon bounding box. Along the way, we can count the number of times the line crossed the polygon's boundary. If the count is an odd number the point must be inside.

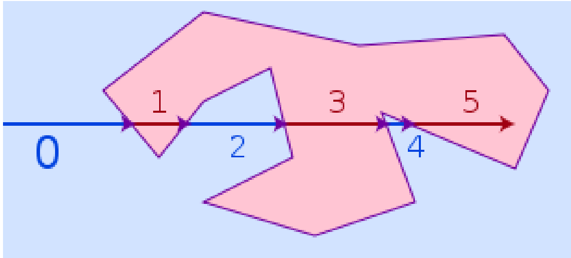
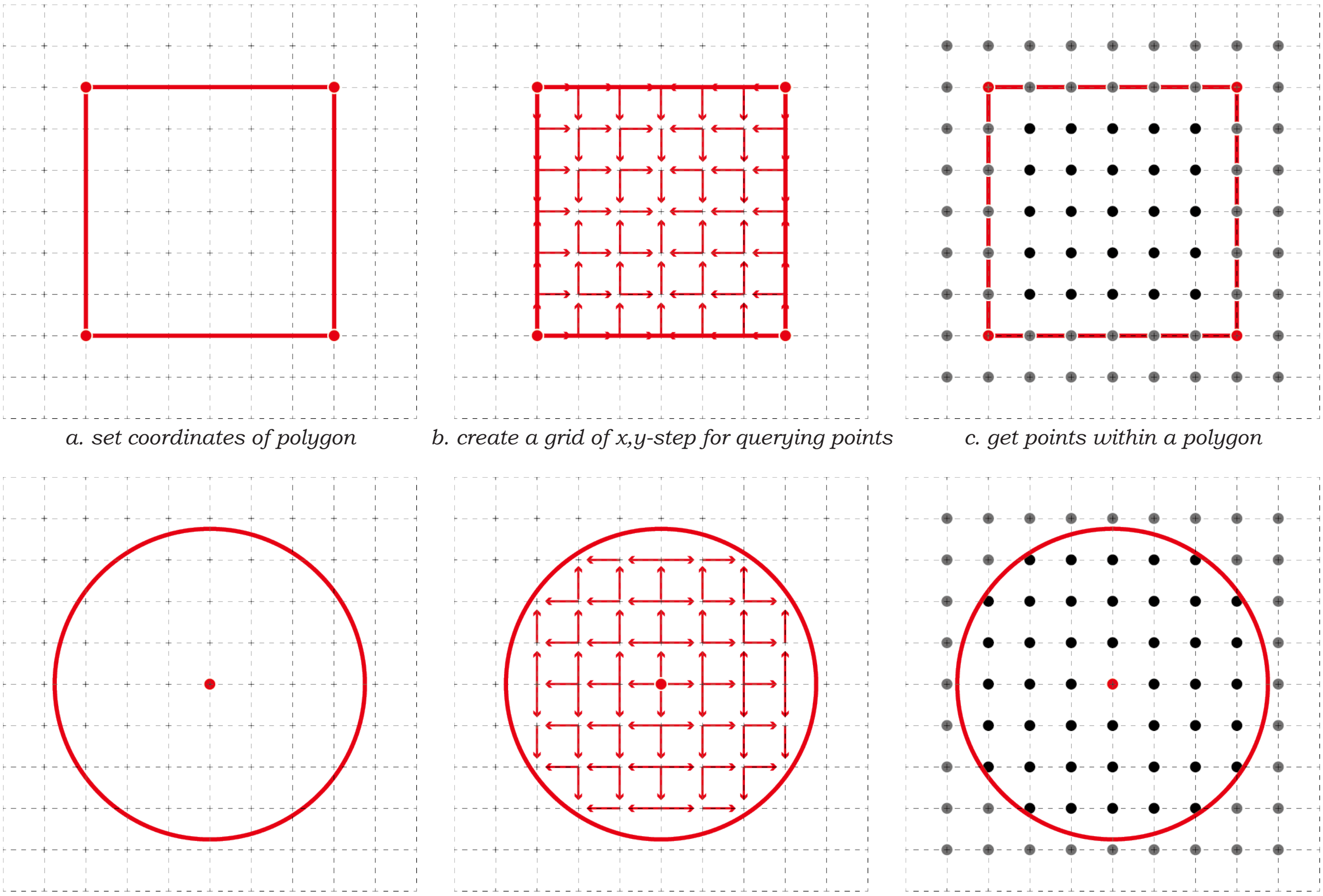


Figure 4.8 Ray Casting Method

The number of intersections for a ray passing from the exterior of the polygon to any point; if odd, it shows that the point lies inside the polygon. If it is even, the point lies outside the polygon; this test also works in three dimensions. The algorithm is described at Wise, Stephen (2002). GIS Basics.

If it's an even number the point is outside the polygon.The other method defines a circle and creates a grid of lat-long for querying points. A function named 'distance on unit sphere' can converts lat & long to spherical coordinates in radians and compute the spherical distance from spherical coordinates. Through defined the function, we can filter out points that are not in the region. After that, we can get points in the specific circle.



a. set coordinates of polygon

b. create a grid of x,y-step for querying points

c. get points within a polygon

d. set coordinates of center and radius

e. create a grid of x,y-step for querying points

f. get points within a circle

● Reference Point

● Selected Point

● Filtered Point

⟵ Step Direction

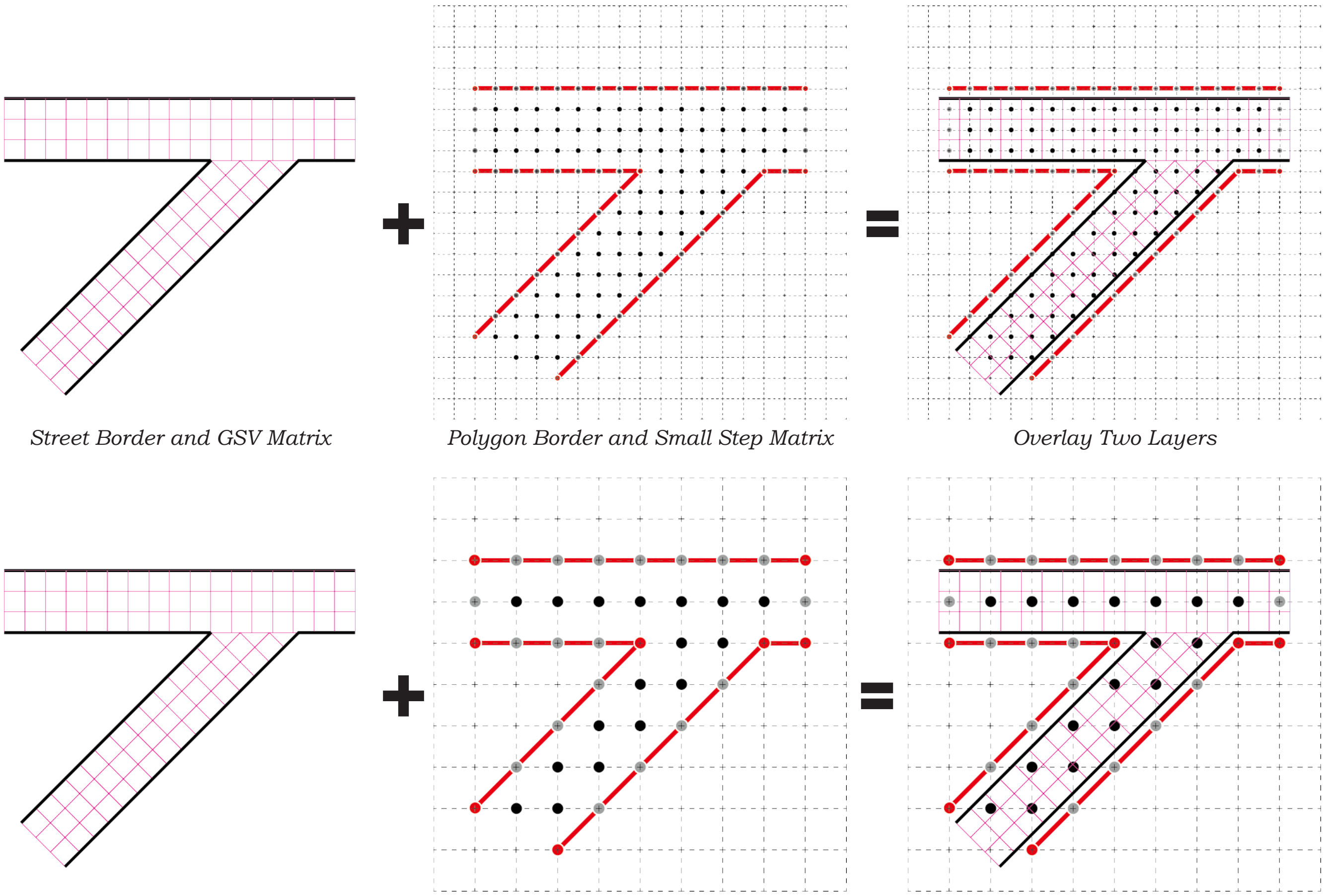
□ Polygon Border

○ Circle Border

Figure 4.9 The method of acquiring images in specific area

a, b and c in figure 4.9 show the schematic to get points in a polygon. First of all, we set the location of polygon corners as the reference coordinates. Secondly, we set the “step” parameters to control the number of querying points. According to the value of step, the difference between adjacent coordinate values are the same, equal to the value of step, and each new coordinate value becomes the reference value of the next coordinate value. Finally, we can get the points within a polygon. Each point responses a coordinate of GSV image in the polygon area.

The other three maps (d, e and f) show how to get points in a circle. Different from the first method, we determine the scope of point collection by setting the center of the circle and the radius. Through the function named ‘distance on unit sphere’, we can determine whether the point within the circle. “Step” is still an important decision parameter. In this method, the coordinate of the center is the reference coordinate. Through iterative calculations, new points are generated until the boundaries of the circle.



Street Border and GSV Matrix *Polygon Border and Small Step Matrix* *Overlay Two Layers*

Street Border and GSV Matrix *Polygon Border and Big Step Matrix* *Overlay Two Layers*

The essence of the GSV image capture tool is to automatically enter a large number of coordinate values as parameters into the Google Street View Image API to capture images. The GSV image corresponding to each automatically generated point coordinate is unique, but the corresponding coordinate value of the GSV image is not unique. This is because the range of the GSV matrix does not exactly coincide with the matrix formed by "step". Therefore, we need to iteratively test the value of "step" until it matches the GSV matrix. This avoids excessive duplication of the image. In our study, the value of "step" is between 0.00025-0.00030, which gives better results.

Figure 4.10 shows the difference between different steps, where "Small Step" obviously gets more duplicate images than "Big Step".

● Reference Point

● Selected Point

● Filtered Point

— Polygon Border

— Street Border

□ GSV Matrix

□ Step Matrix

Figure 4.10 Application of Image Acquisition Tools in GSV

1) Example of image acquisition in a Polygon Area(Python code)

```
import numpy as np
import requests
import random
api_key = ‘My Google Street View Images API key’

# The algorithm is called the “Ray Casting Method”.
def point_in_poly(p,poly):
    n = len(poly)
    inside = False
    p1x,p1y = poly[0]
    for i in range(n+1):
        p2x,p2y = poly[i % n]
        if p[1] > min(p1y,p2y):
            if p[1] <= max(p1y,p2y):
                if p[0] <= max(p1x,p2x):
                    if p1y != p2y:
                        xints = (p[1]-p1y)*(p2x-p1x)/(p2y-p1y)+p1x
                    if p1x == p2x or p[0] <= xints:
                        inside = not inside
        p1x,p1y = p2x,p2y
    return inside
```

The algorithm is called the “Ray Casting Method”. Source: http://geospatialpython.com/2011/01/point-in-polygon.html

```
# defined coordinates
def strcoor(coor):
    return str(coor[0])+’,’+str(coor[1])

# get imgs within ‘polygon’ and save them to ‘cityname’ folder
def getim(polygon,cityname):
```

defined the step

```
    xstep=0.00025
    ystep=0.00025
```

create a grid of lat-long for querying images

```
    xs=np.arange(min([x for x,y in polygon]),max([x for x,y in polygon]),xstep)
    ys=np.arange(min([y for x,y in polygon]),max([y for x,y in polygon]),ystep)
    xx,yy = np.meshgrid(xs,ys)
```

flatten the meshgrid and create tuples of lat-long

```
    allcoors=list(zip(xx.ravel(),yy.ravel()))
```

```
valid_coors=[coor for coor in allcoors if point_in_poly(coor,polygon)]
dir_count=[‘0’,’45’,’90’,’135’,’180’,’225’,’270’,’315’]
for i in range(len(valid_coors)):
    print (valid_coors[i], i)
    for heading in dir_count:
            params=’size=480x360&location=’ + strcoor(valid_coors[i]) + ‘&fov=100&heading=’+ str(heading) + ‘&pitch=7’
              url=’https://maps.googleapis.com/maps/api/streetview?’+params + ‘&key=’ + api_key
          response = requests.get(url, stream=True)
            with open(‘E:\Data_Acquisition_Milano/’+cityname+’/’+strcoor(valid_coors[i])+’_’+str(heading)+’.png’, ‘wb’) as out_file:
              out_file.write(response.content)
          del response
# define the parameters
polygon_po = [(45.480292,9.223608), (45.480728,9.234553), (45.475961,9.223772), (45.476769,9.225887), (45.476731,9.234593)]
getim(polygon_po,’Polimi’)
```

filter out points that are not in the region

number of directions we get from each point

set API parameters

download images request method

This code references the source code of Lezhi Li’s study in the streetview.

Original Source Code: https://github.com/Firenze11/cv_streetview/tree/master/scripts

The details of image acquisition:

Leonardo Campus, Politecnico di Milano

Coordinate of Polygon:

(45.480292,9.223608), (45.480728,9.234553),

(45.476731,9.234593), (45.476769,9.225887),

(45.475961,9.223772)

*Image size: 360*480   Step size: 0.00025 Number of directions: 8*

Amount of images: 5168   Road length: 4.2km

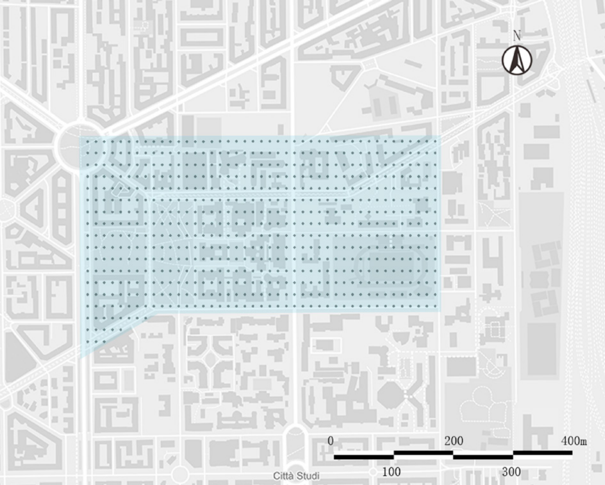


Figure 4.11 Image Acquisition Area - Leonardo Campus, Politecnico di Milano

2) Example of image acquisition in a Circular Area(Python code)

```
import numpy as np
import requests
import math
import os
api_key = ‘My Google Street View Images API key’
# defined coordinates
def strcoor(coor):
    return str(coor[0])+’,’+str(coor[1])

# defined distance on unit sphere
def distance_on_unit_sphere(lat1, long1, lat2, long2):
    degrees_to_radians = math.pi/180.0
    phi1 = (90.0 - lat1)*degrees_to_radians
    phi2 = (90.0 - lat2)*degrees_to_radians
    theta1 = long1*degrees_to_radians
    theta2 = long2*degrees_to_radians
    cos = (math.sin(phi1)*math.sin(phi2)*math.cos(theta1 - theta2) + math.cos(phi1)*math.cos(phi2))
    arc = math.acos(cos)*6371
    return arc

# get imgs within a circle and save them to ‘Areaname’folder
def getim_circle(center_lat_long, radius, Areaname):
    global I_NUM
    xstep=0.0003 / math.cos(center_lat_long[0] * math.pi/180.0)
    ystep=0.0003
    ys=np.arange(center_lat_long[0]-0.1,center_lat_long[0]+0.1,ystep)
    xs=np.arange(center_lat_long[1]-0.1,center_lat_long[1]+0.1,xstep)
    yy,xx = np.meshgrid(ys,xs)
    allcoors=list(zip(yy.ravel(),xx.ravel()))
    valid_coors=[coor for coor in allcoors if distance_on_unit_sphere(center_lat_long[0],center_lat_long[1],coor[0],coor[1]) < radius]
    dir_count=4
    for i in range(len(valid_coors)):
        if i >= I_NUM:
            I_NUM = i # for remembering where we left off
```

Converts lat & long to spherical coordinates in radians.

source: http://gis.stackexchange.com/questions/163785/using-python-to-compute-the-distance-between-coordinates-lat-long-using-havers

phi(Φ) = 90 - latitude

theta(Θ) = longitude

Compute the spherical distance from spherical coordinates.

radius of the earth in km

normalize x spacing

create a grid of lat-long for querying images

flatten the meshgrid and create tuples of lat-long

filter out poins that are not in the region

number of directions we get from each point

```
            for heading in range(dir_count):
                params='size=480x360&location=' + strcoor(valid_coors[i]) +
'&fov=100&heading='+ str(heading*360/dir_count) + '&pitch=7'
                    url='https://maps.googleapis.com/maps/api/street-
view?'+params + '&key=' + api_key
                response = requests.get(url, stream=True)
                with open('E:/Data_Acquisition_Milano/'+Areaname+'/'+str-
coor(valid_coors[i])+'_'+str(heading)+'.png', 'wb') as out_file:
                    out_file.write(response.content)
                del response


# define the parameters
Areaname = 'Duomo'
center = [45.464154, 9.191965]
newpath = 'E:/Data_Acquisition_Milano/'+Areaname
if not os.path.exists(newpath):
    os.makedirs(newpath)
getim_circle(center,0.4,Areaname)
```

set API parameters

download images request method

This code references the source code of Lezhi Li's study in the streetview.

Original Source Code: https://github.com/Firenze11/cv_streetview/tree/master/scripts

The details of image acquisition:

Name: Duomo di Milano (commercial center)

Radius= 400m

Area= 502654.825m^2

Coordinate of Center: (45.464154,9.191965)

*Image size: 360*480*

Step size: 0.0003   Number of directions: 4

Amount of images: 1804   Road length: 9.2km

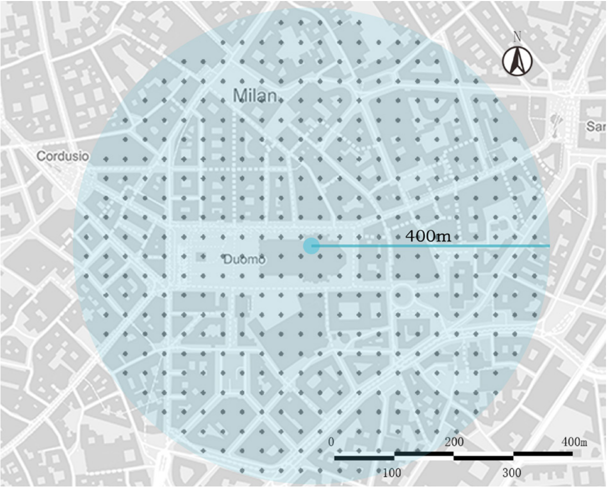
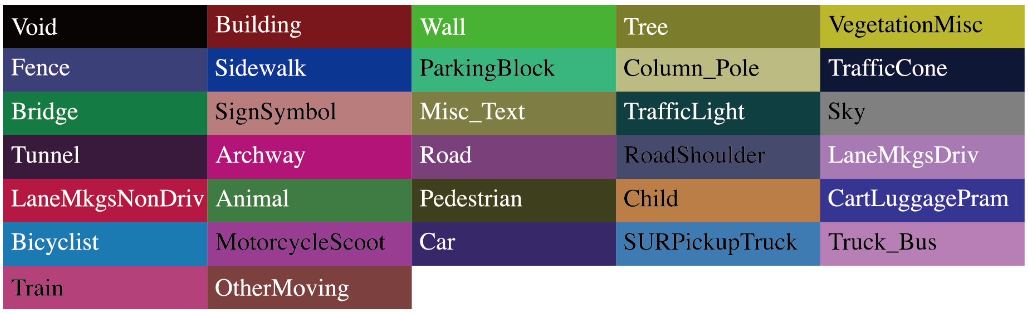


Figure 4.12 Image Acquisition Area - Duomo di Milano

4.4 Image Annotation and Classification

4.4.1 Image Classification of CamVid Dataset[14]

The Cambridge-driving Labeled Video Database is a ground truth dataset that can be freely used for research work in object recognition in video, it depicts a moving driving scene in the city of Cambridge filmed from a moving car. The dataset is taken around Cambridge, UK, and contains day and dusk scenes. Our classification standard follows the CamVid dataset's rule which contains 701 images of road scenes for training and testing. It consists of 960x720 pixel images in which each pixel was manually assigned to one of the following 32 object classes that are relevant in a driving environment.

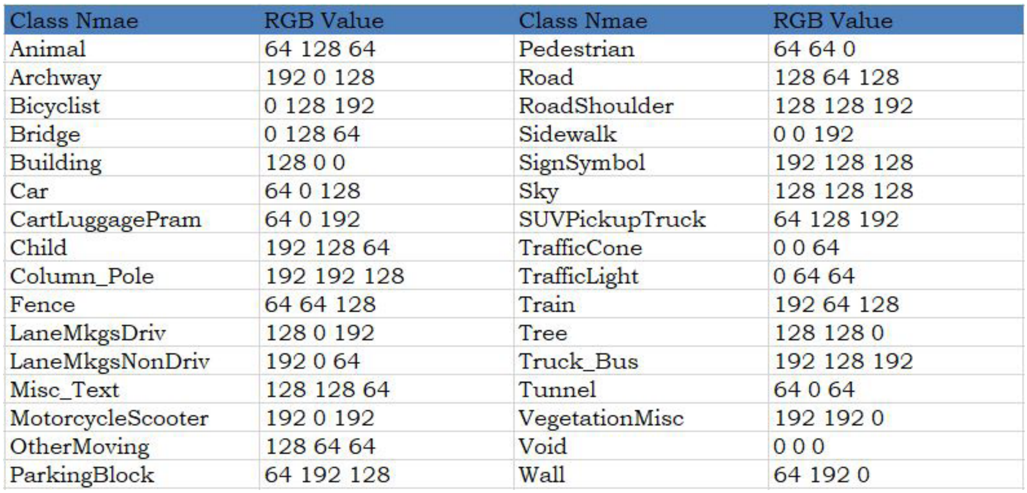


Void	Building	Wall	Tree	VegetationMisc
Fence	Sidewalk	ParkingBlock	Column_Pole	TrafficCone
Bridge	SignSymbol	Misc_Text	TrafficLight	Sky
Tunnel	Archway	Road	RoadShoulder	LaneMkgsDriv
LaneMkgsNonDriv	Animal	Pedestrian	Child	CartLuggagePram
Bicyclist	MotorcycleScoot	Car	SURPickupTruck	Truck_Bus
Train	OtherMoving			

Figure 4.13 list of class labels and corresponding colours in CamVid Dataset

Source: Cambridge Labeled Objects in Video

All images (original and ground truth) are in uncompressed 24-bit color PNG format. For each frame from the original sequence, its corresponding labeled frame bears the same name, with an extra "_L" before the ".png" extension.



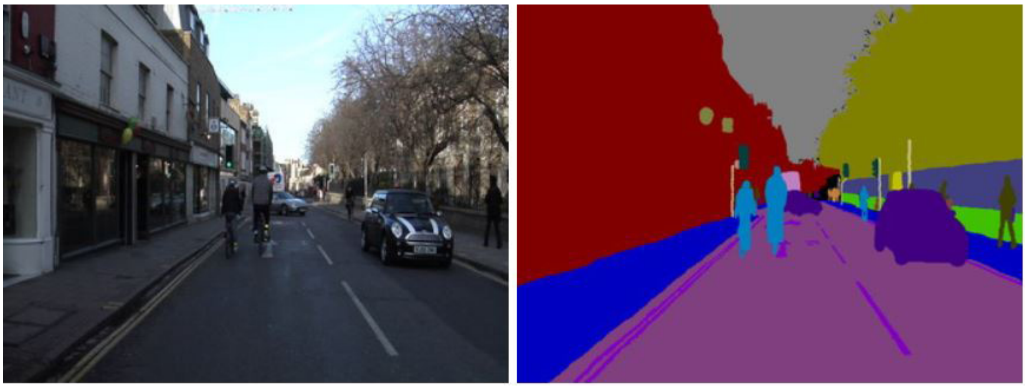
Class Nmae	RGB Value	Class Nmae	RGB Value
Animal	64 128 64	Pedestrian	64 64 0
Archway	192 0 128	Road	128 64 128
Bicyclist	0 128 192	RoadShoulder	128 128 192
Bridge	0 128 64	Sidewalk	0 0 192
Building	128 0 0	SignSymbol	192 128 128
Car	64 0 128	Sky	128 128 128
CartLuggagePram	64 0 192	SUVPickupTruck	64 128 192
Child	192 128 64	TrafficCone	0 0 64
Column_Pole	192 192 128	TrafficLight	0 64 64
Fence	64 64 128	Train	192 64 128
LaneMkgsDriv	128 0 192	Tree	128 128 0
LaneMkgsNonDriv	192 0 64	Truck_Bus	192 128 192
Misc_Text	128 128 64	Tunnel	64 0 64
MotorcycleScooter	192 0 192	VegetationMisc	192 192 0
OtherMoving	128 64 64	Void	0 0 0
ParkingBlock	64 192 128	Wall	64 192 0

Table 4.1 Name of class labels and RGB Value

[14] *http://mi.eng.cam.ac.uk/research/projects/VideoRec/CamVid/, retrieved on 15-11-2017*

In order to build the groundtruth, the CamVid Group hired people to produce manually the labeled maps for each of the 101 frames. In the labeled images, each object has been painted with a given class color by human operators (Figure 4.14).

> *"They painted the areas corresponding to a predefined list of 32 objects of interest given a specific palette of colors. . . . . . . By logging and timing each stroke, we were able to estimate the hand labeling time for one frame to be around 20-25 minutes (this duration can vary greatly depending on the complexity of the scene)" (Source from: http://mi.eng.cam.ac.uk/research/projects/VideoRec/CamSeq01/)*



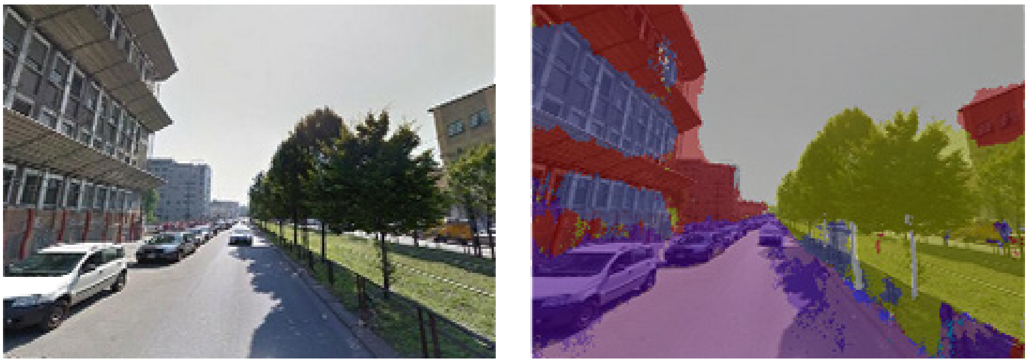
original image **labeled image**

Figure 4.14 Example of original and labeled image

Source: Cambridge Labeled Objects in Video

4.4.2 Image Classification of UrSA

SegNet learns to predict pixel-wise class labels from supervised learning. Therefore we require a dataset of input images with corresponding ground truth labels before we train a SegNet model for multi-class pixel-wise classification. And SegNet will take a color image, such as the one on the left (Figure 4.15), and produce a labeled output (segmentation) like the image on the right. Label images should be RGB channel, with each pixel labeled with its class.

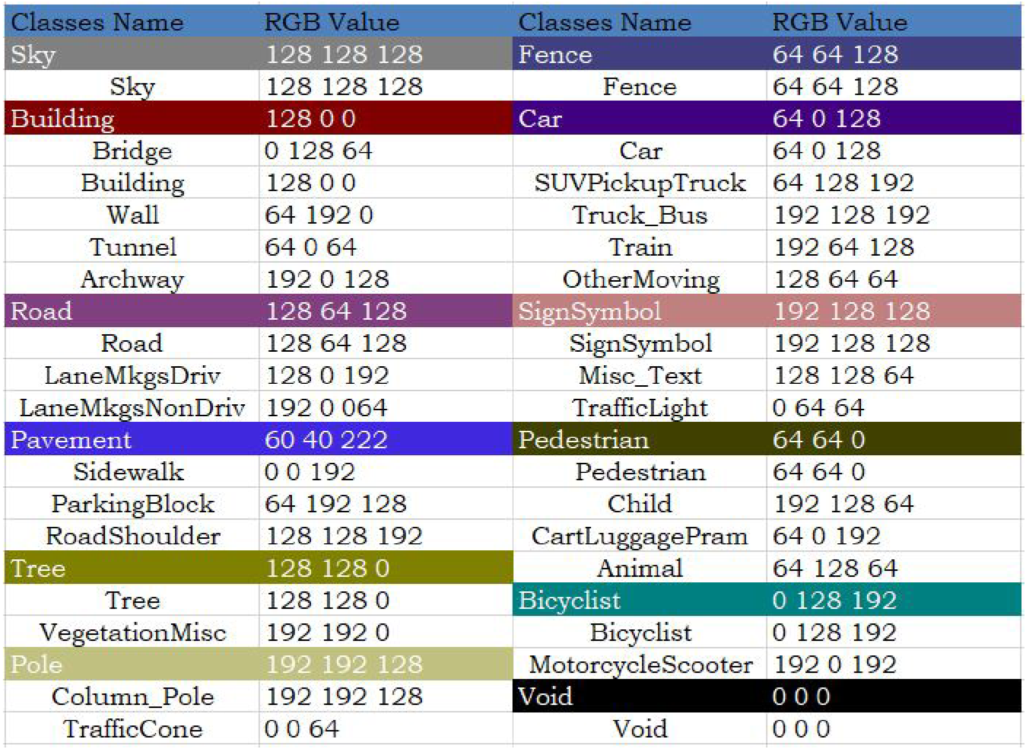


original image **output image**

Figure 4.15 Example of original and output image

Source: Made by Author

We are going to group the 32 original classes in CamVid to 11 classes. The “void” label indicates an area which ambiguous or irrelevant in this context. The color/class association is given in Figure 4.16. Each line has the R G B values (between 0 and 255) and then the class name.



Classes Name	RGB Value	Classes Name	RGB Value
Sky	128 128 128	Fence	64 64 128
Sky	128 128 128	Fence	64 64 128
Building	128 0 0	Car	64 0 128
Bridge	0 128 64	Car	64 0 128
Building	128 0 0	SUVPickupTruck	64 128 192
Wall	64 192 0	Truck_Bus	192 128 192
Tunnel	64 0 64	Train	192 64 128
Archway	192 0 128	OtherMoving	128 64 64
Road	128 64 128	SignSymbol	192 128 128
Road	128 64 128	SignSymbol	192 128 128
LaneMkgsDriv	128 0 192	Misc_Text	128 128 64
LaneMkgsNonDriv	192 0 064	TrafficLight	0 64 64
Pavement	60 40 222	Pedestrian	64 64 0
Sidewalk	0 0 192	Pedestrian	64 64 0
ParkingBlock	64 192 128	Child	192 128 64
RoadShoulder	128 128 192	CartLuggagePram	64 0 192
Tree	128 128 0	Animal	64 128 64
Tree	128 128 0	Bicyclist	0 128 192
VegetationMisc	192 192 0	Bicyclist	0 128 192
Pole	192 192 128	MotorcycleScooter	192 0 192
Column_Pole	192 192 128	Void	0 0 0
TrafficCone	0 0 64	Void	0 0 0

Figure 4.16 list of class labels and corresponding colours

Source: Made by Author

4.4.3 UrSA Dataset

As we know, SegNet learns each pixel label of an image from supervised learning. According to the actual test results, we found that the model has obvious differences in the accuracy of image segmentation when analyzing street scenes in different cities. Sometimes, facing unknown areas, the image segmentation accuracy was significantly reduced. Therefore, we try to increase the coverage of the labeled dataset. Finally, we manually labeled a total of 150 images in 15 cities around the world.

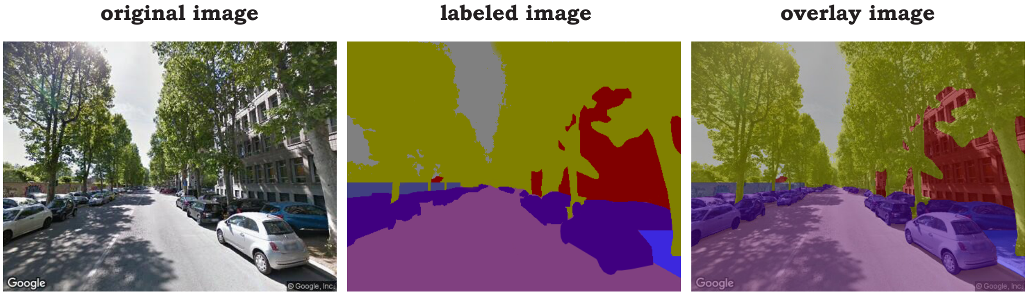


Figure 4.17 Example of original and labeled image

Source: Made by Author

4.5 Semantic Segmentation Using SegNet

MATLAB (matrix laboratory) is a multi-paradigm numerical computing environment, a proprietary programming language developed by MathWorks. MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, the creation of user interfaces, and interfacing with programs written in other languages, including C, C++, C#, Java, Fortran, and Python.

Although MATLAB is mainly used for numerical calculations, it is also suitable for a wide range of applications such as Deep Learning, Computer Vision, Signal Processing, Data Analytics. The version released by Matlab in March each year is denoted by “a”, while the version released by September is denoted by “b”. For example, R2006a represents the version released in March 2006 and R2009b the version released in the second half of 2009. The Matlab version used in this article is the R2017b, released in September 2017.

4.5.1 Build SegNet Network

In Matlab 2017b, we can create SegNet layers for semantic segmentation, the SegNet network with weights initialized from the VGG-16 network. We use “segnetLayers[15]” function to create a SegNet network initialized using VGG-16 weights. The function automatically performs the network surgery needed to transfer the weights from VGG-16 and adds the additional layers required for semantic segmentation.

The segnetLayers’s syntax:

lgraph = segnetLayers(imageSize, numClasses, model)

imageSize — Network input image size, based on the size of the images in the dataset

numClasses — Number of classes, based on the classes in UrSA

model — Pretrained network model, vgg16 is selected

4.5.2 The Pretrained VGG16 Model

The common image classification datasets are CIFAR, ImageNet, COCO, etc. Commonly used fine-grained image classification datasets include CUB-200-2011, Stanford Dog, Oxford-flowers, etc. A large number of research results based on ImageNet large-

[15] *https://cn.mathworks.com/help/vision/ref/segnetlayers.html, retrieved on 02-12-2017*

scale dataset. The ImageNet data has been changing slightly since 2010. Now it is common to use the ImageNet-2012 dataset, which contains 1000 categories: the training set contains 1,281,167 images, each ranging in size from 732 to 1,300 and the validation set contains 50,000 images, an average of 50 pictures per category.

In practical projects, training is usually not directly from the first layer, but rather through the training of a good model on a large dataset (such as ImageNet), the parameters of the previous layers fixed in the application of our new problem. On the last two layers changes, with our own data to fine tune (finetuning), the general effect is also very good. The so-called finetuning, that is, we have trained for a similar task model, such as CaffeNet, VGG-16, ResNet, and then through our own dataset weight update.

If the amount of data is relatively small, we can update the last layer, the other layers of the same weight; if the amount of data is medium, we can train the behind layers; if the amount of data is large, training directly from scratch, but training time, we need to spend more. After training the network, we only need to forward the process to make predictions. Of course, we can also use this network directly as a feature extractor, which can directly use any layer of output as the feature.

According to the R-CNN paper, the experimental results of Alexnet, if not fine-tuning, the effect of characteristics of the pool5 and fc6 and fc7 is not very strong, so if it used directly as the feature extractor, the last layer of the pool output directly.

Due to the large ImageNet dataset, downloading and training are slow. However, the VGG model has excellent generalization ability, and the pre-training model can be loaded directly to reduce the training time of the model. MathWorks Neural Network Toolbox Team provide a pretrained VGG-16 network model for image classification that has been trained on approximately 1.2 million images from the ImageNet Dataset (http://image-net.org/index). To get VGG-16, we should install the Neural Network Toolbox™ Model for VGG-16 Network. After installation is complete, run the following code to verify that the installation is correct.

```
vgg16();
```

4.5.3 Analyze Dataset Statistics and Balance Classes

After building the SegNet network and loading the pre-trained model, we need to make statistical analysis to see the distribution of class labels in UrSA dataset. In Matlab, we use the "countEachLabel[16]" function to count the number of pixels by the class label.

The countEachLabel's syntax:

tbl = countEachLabel(datasource)

frequency = tbl.PixelCount/ sum(tbl.PixelCount)

datasource — Create datasource using ground truth and pixel labeled images.

tbl — Pixel label information, returned as a table. The table contains three variables:

Name - Pixel label class name

PixelCount - Number of pixels in class

ImagePixelCount - Total number of pixels in images that had an instance of a class

Ideally, all classes would have an equal number of observations (same frequency). However, the classes in UrSA are imbalanced, which is a common issue in datasets of street scenes. Such scenes have more sky, building, and road pixels than pedestrian and bicyclist pixels because sky, buildings, and roads cover more area in the image. If not handled properly, this imbalance can be detrimental to the learning process, because learning is biased in favor of the dominant classes.

In practice, the distribution of classes in a dataset are not balanced, so we use class weighting to balance the classes to improve training. In detail, we use the pixel label counts computed with "countEachLabel" function to calculate the median frequency class weights and specify the class weights using "pixelClassificationLayer[17]" function.

The formula of Class Weights:

imageFreq = tbl.PixelCount ./ tbl.ImagePixelCount;

classWeights = median(imageFreq) ./ imageFreq

Layer = pixelClassificationLayer('Name','labels','ClassNames', tbl.Name, 'ClassWeights', classWeights)

[16] https://cn.mathworks.com/help/vision/ref/pixellabelimagesource.counteachlabel.html, retrieved on 02-12-2017

[17] https://cn.mathworks.com/help/vision/ref/nnet.cnn.layer.pixelclassificationlayer.html, retrieved on 02-12-2017

After balance the classes, we should update the SegNet network by removing the current pixelClassificationLayer and adding the new layer. In SegNet network, the current pixelClassificationLayer is named ‘pixelLabels’. We remove it using “removeLayers[18]”, add the new one using “addLayers[19]”, and connect the new layer to the rest of the network using “connectLayers[20]”.

Update the SegNet network:

```
lgraph = removeLayers(lgraph, ‘pixelLabels’);
lgraph = addLayers(lgraph, Layer);
lgraph = connectLayers(lgraph, ‘softmax’ ,’labels’);
```

4.5.4 Select Training Options and Start Training & Test

When you have completed the previous steps, we need to set training options and start training. The optimization algorithm used for training is stochastic gradient descent with momentum (SGDM). We use “trainingOptions” to specify the hyperparameters[21] used for SGDM.

1) Training Options:

```
options = trainingOptions(‘sgdm’, ...
    ‘Momentum’, 0.9, ...
    ‘InitialLearnRate’, 1e-3, ...
    ‘L2Regularization’, 0.0005, ...
    ‘MaxEpochs’, 100, ...
    ‘MiniBatchSize’, 4, ...
    ‘Shuffle’, ‘every-epoch’, ...
    ‘VerboseFrequency’, 2);
```

2) Start Training:

```
datasource = pixelLabelImageSource(imds,pxds)
[net, info] = trainNetwork(datasource,lgraph,options);
```

3) Start Test:

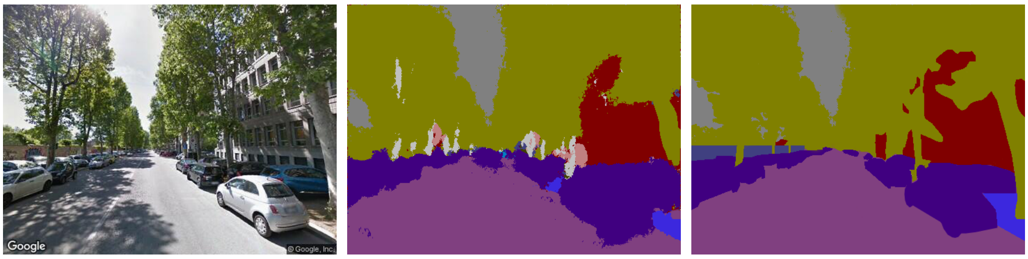
```
I = read(Test_Image);
C = semanticseg(I, net);
```

[18] *https://cn.mathworks.com/help/nnet/ref/removelayers.html, retrieved on 02-12-2017*

[19] *https://cn.mathworks.com/help/nnet/ref/addlayers.html, retrieved on 02-12-2017*

[20] *https://cn.mathworks.com/help/nnet/ref/connectlayers.html, retrieved on 02-12-2017*

[21] *In the context of machine learning, hyperparameters are parameters whose values are set prior to the commencement of the learning process. By contrast, the values of other parameters are derived via training.*



original image **output image** **ground truth image**

Figure 4.18 The Ground truth image and Output image

Source: Made by Author

4.6 Model Assessment and perfection

4.6.1 Evaluate one trained image

Figure 4.18 shows one result of semantic segmentation. Visually, the results overlap well for classes such as road, sky, and building. However, smaller objects like pedestrians and cars are not as accurate. The amount of overlap per class can be measured using the intersection-over-union (IoU) metric, also known as the Jaccard index. In Matlab, we use the "Jaccard[22]" function to measure IoU.

The one measured result:

```
    ans =

          classes         iou
           "Sky"          0.90831
         "Building"       0.81071
         "Pole"           0
         "Road"           0.94595
         "Pavement"       0.45793
         "Tree"           0.82933
         "SignSymbol"     0.12162
         "Fence"          0.21443
         "Car"            0.456803
         "Pedestrian"     0
         "Bicyclist"      0
```

4.6.2 Evaluate trained network

[22] https://cn.mathworks.com/help/images/ref/jaccard.html, retrieved on 02-12-2017

To measure accuracy for multiple test images, run “semanticseg[23]” on the entire test set. “semanticseg” returns the results for the test set as a pixelLabelDatastore object. The actual pixel label data for each test image in imdsTest is written to disk in the location specified by the ‘WriteLocation’ parameter. Use “evaluateSemanticSegmentation[24]” to measure semantic segmentation metrics on the test set results.

The multiple test images measured result:

Source: Documentation for evaluateSemanticSegmentation from MathWorks

```
ans =

    GlobalAccuracy      0.88236

    MeanAccuracy        0.85071

    MeanIoU             0.60986

    WeightedIoU         0.7988

    MeanBFScore         0.61248
```

- *GlobalAccuracy*

GlobalAccuracy is the ratio of correctly classified pixels, regardless of class, to the total number of pixels. Use the global accuracy metric if you want a quick and computationally inexpensive estimate of the percentage of correctly classified pixels.

- *Intersection over union (IoU)*

Intersection over union (IoU), also known as the Jaccard similarity coefficient, is the most commonly used metric. Use the IoU metric if you want a statistical accuracy measurement that penalizes false positives.

For each class, IoU is the ratio of correctly classified pixels to the total number of ground truth and predicted pixels in that class. In other words, IoU score = TP / (TP + FP + FN). The image describes the true positives (TP), false positives (FP), and false negatives (FN).

For each image, MeanIoU is the average IoU score of all classes in that particular image. For the aggregate data set, MeanIoU is the average IoU score of all classes in all images.

- *Weighted-iou*

Average IoU of each class, weighted by the number of pixels in that class. Use this metric if images have disproportionally sized classes, to reduce the impact of errors in the small classes on the aggregate quality score.

- *MeanBFScore*

[23] *https://cn.mathworks.com/help/vision/ref/semanticseg.html, retrieved on 02-12-2017*

[24] *https://cn.mathworks.com/help/vision/ref/evaluatesemanticsegmentation.html, retrieved on 02-12-2017*

The boundary F1 (BF) contour matching score indicates how well the predicted boundary of each class aligns with the true boundary. Use the BF score if you want a metric that tends to correlate better with human qualitative assessment than the IoU metric.

For each class, MeanBFScore is the average BF score of that class over all images. For each image, MeanBFScore is the average BF score of all classes in that particular image. For the aggregate data set, MeanBFScore is the average BF score of all classes in all images.

4.6.3 Operating Environment

Hardware:

Computer brand: MECHREVO X6TI-S

System version: Windows 10

System Type: 64-bit operating system, x64-based processor

CPU: Intel (R) Core (TM) i7-6700HQ CPU @ 2.60GHz x8

Graphics: NVIDIA GeForce GTX 965M

Memory: 8.00GB

The main software:

ArcGis 10.2

Matlab 2017b

Photoshop 2015

Python 3.6.3

4.7 Statistics and Visualization

Each image is segmented by SegNet with 11 classes correspond to 11 colors. Through count the amount of color in images, we can analyze the classes information in each image and the total street information in the specific area. So we can analyze the status and distribution of different classes in specific area or city. Finally, we will show the details by visualization technology. This part describes the application and visualization of UrSA in three levels:

Urban center area (three kilometers radius): used to study the overall street characteristics of the target city to show the uniqueness of the city.

Urban important function area (custom polygon border): used to identify the differences between functional areas, summarizing the features of different functional areas in the city.

Urban block / street: specific to each street, UrSA visualizes street-level detailed feature information.

4.7.1 UrSA in 15 famous cities

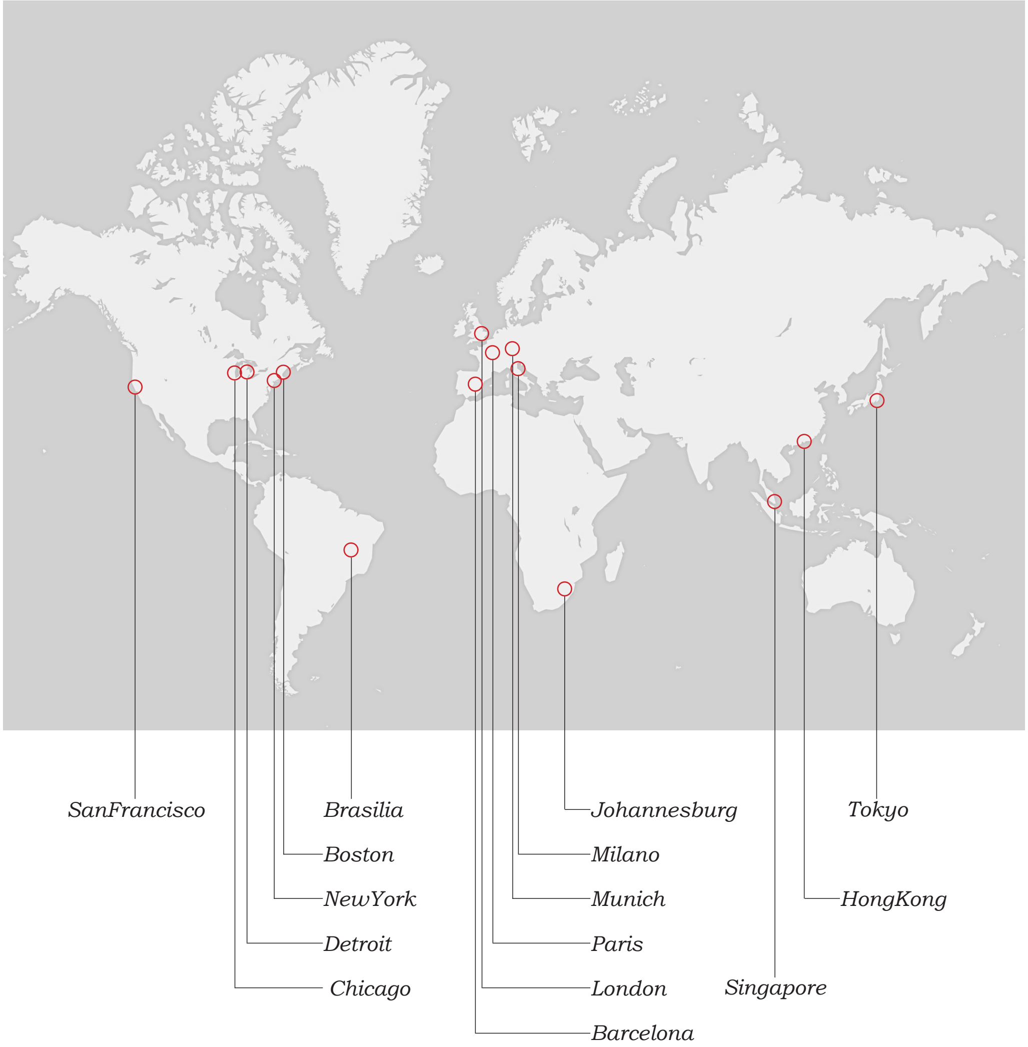


Figure 4.19
The 15 famous cities

We use “point in a circle” method to collect image information in 15 famous cities. Table 4.2 shows the details of image acquisition.

By running UrSA, we get the street profile of each city center (Figure 4.20a and Figure 4.20b). We found that in the city streets in Europe, “building” is more obvious characteristics, especially in Paris and Milan. As for Asian cities, Hong Kong and Singapore have very high “tree” rates. In contrast, Tokyo has a very high “building” rate. As the largest city in the world built in the late 20th century, Brasilia possesses particularly high “sky” and “tree”. This shows the distinctly different design concepts and spatial dimensions of Brasilia from other cities.

Table 4.2 The detials of image acquisition

City	Coordinate of Center	Radius
Barcelona	41.390298, 2.162001	3km
Boston	22.280556, 114.165278	3km
Brasilia	41.881944, -87.627778	3km
Chicago	40.747783, -73.968068	3km
Detroit	42.331389, -83.045833	3km
HongKong	37.783333, -122.416667	3km
Johannesburg	-26.204444, 28.045556	3km
London	51.507360, -0.127630	3km
Milano	45.464167, 9.190278	3km
Munich	48.139741, 11.565510	3km
NewYork	35.670226, 139.775122	3km
Paris	48.857527, 2.341560	3km
Sanfrancisco	1.302876, 103.829547	3km
Singapore	42.357778, -71.061667	3km
Tokyo	-15.797616, -47.891761	3km

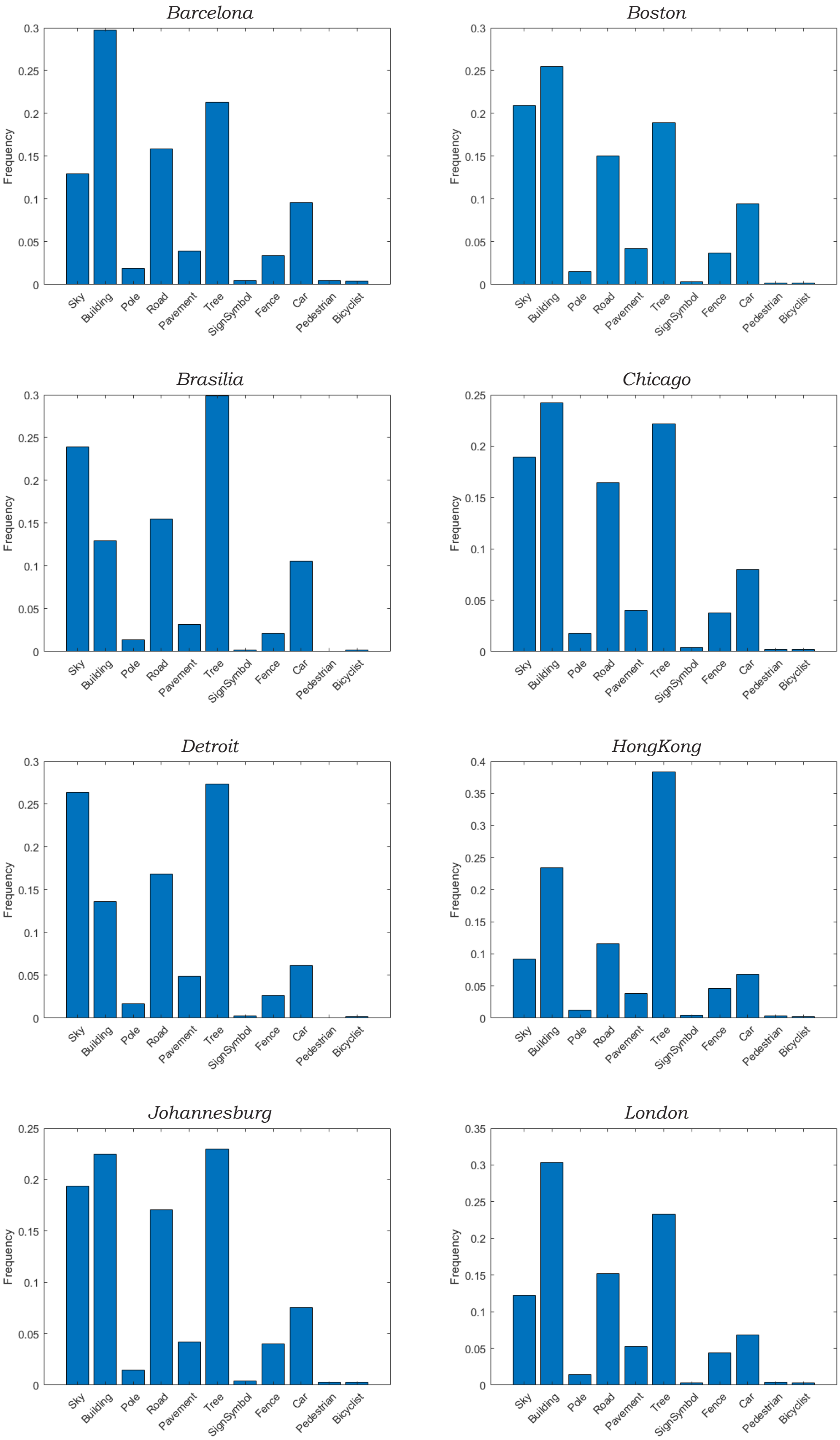


Figure 4.20.a
The UrSA Results

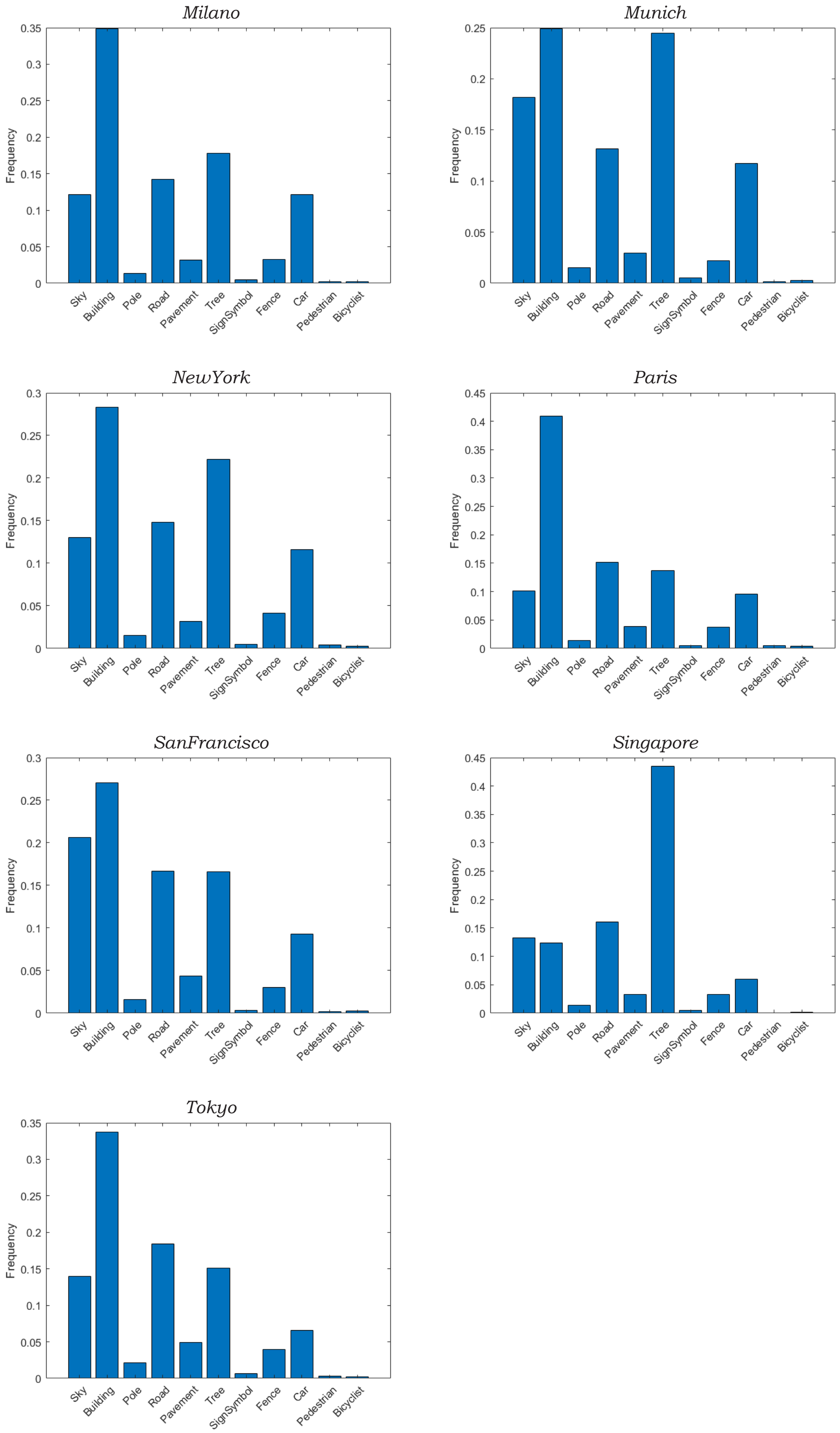


Figure 4.20.b
The UrSA Results

4.7.2 UrSA in Districts of Milano

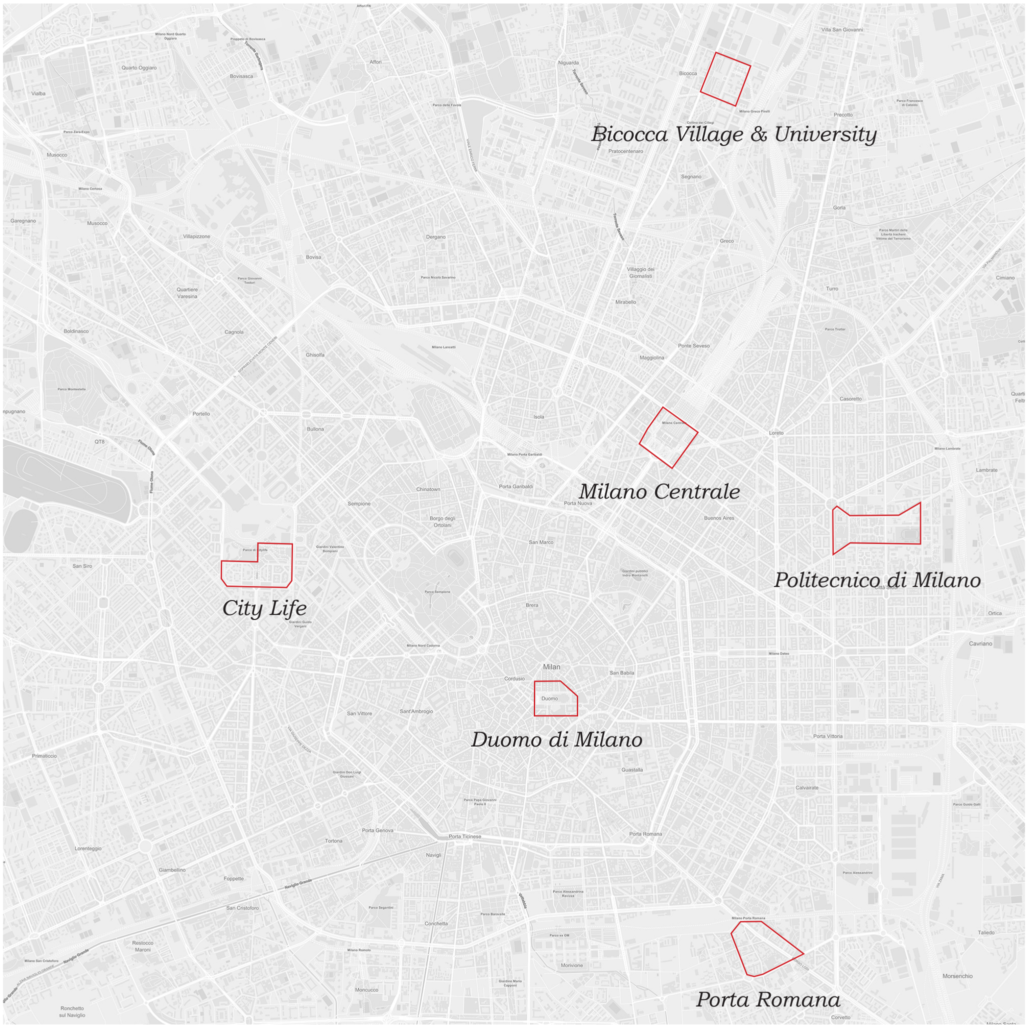


Figure 4.21
The 6 Districts of Milano

In this section, we used the “point in a polygon” method to capture GSV images from six regions of Milan and UrSA is used to identify the differences between functional areas. Each area corresponds to a city pattern.

Bicocca Village & University - Education and living in the suburbs

City Life - New town area

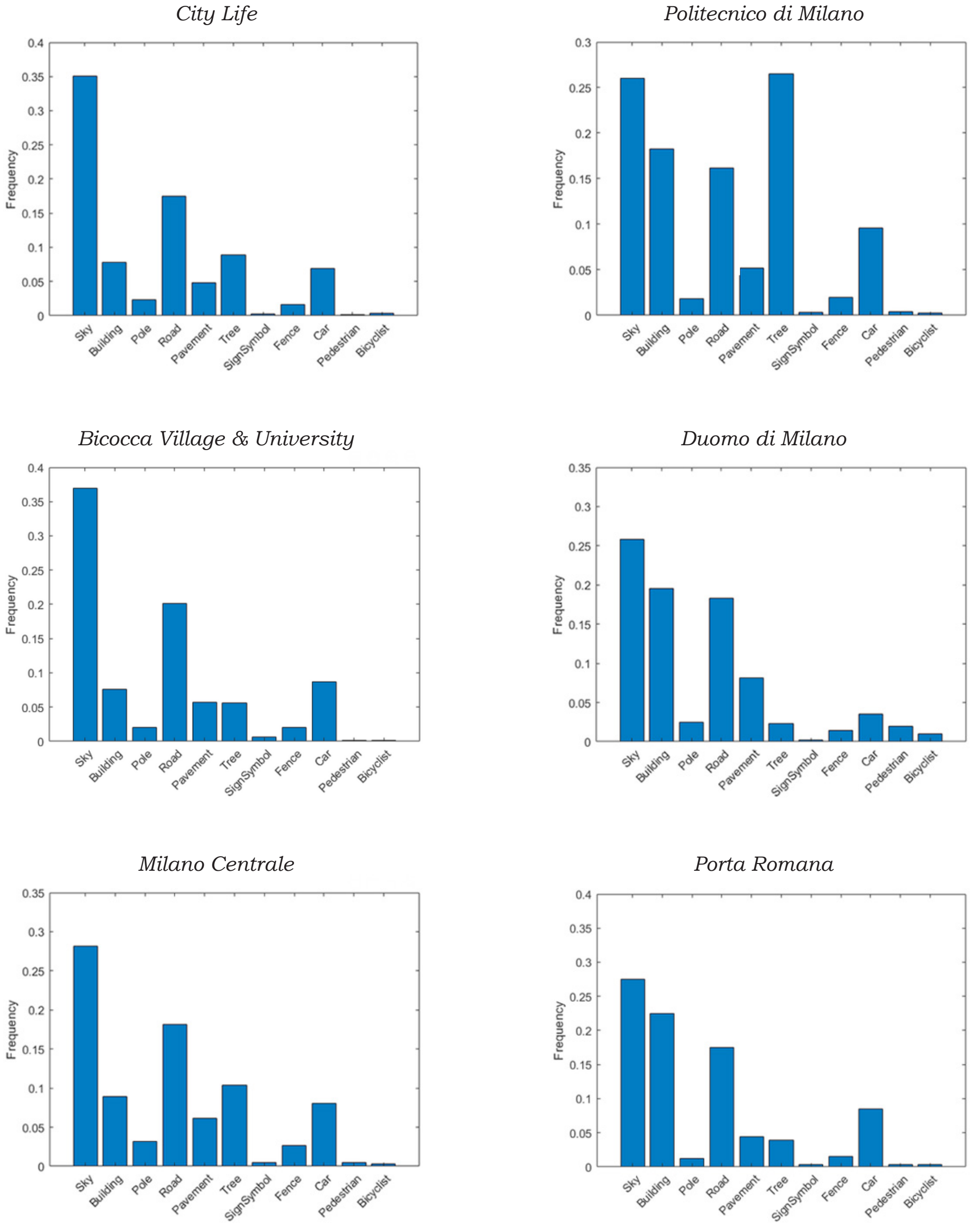
Duomo di Milano - Historical and commercial center

Milano Centrale - Train transportation hub

Politecnico di Milano - University

Porta Romana - Living in the suburbs

From UrSA's results, we found that Milan has a good "sky" ratio in each study area. At the same time, we found that there are very similar features of urban streets in City Life, Bicocca and Milano Centrale, similar low "building" ratio and "tree" ratio. This is very interesting that completely different functional area has a similar street environment. On the contrary, we can see more "building" from other three places. Also, we can find that there is the highest "tree" ratio in Politecnico di Milano. In the sense, here should be able to block more sunlight in the summer.



City Life

Politecnico di Milano

Bicocca Village & University

Duomo di Milano

Milano Centrale

Porta Romana

Figure 4.22
The UrSA Results

4.7.3 UrSA in Politecnico di Milano

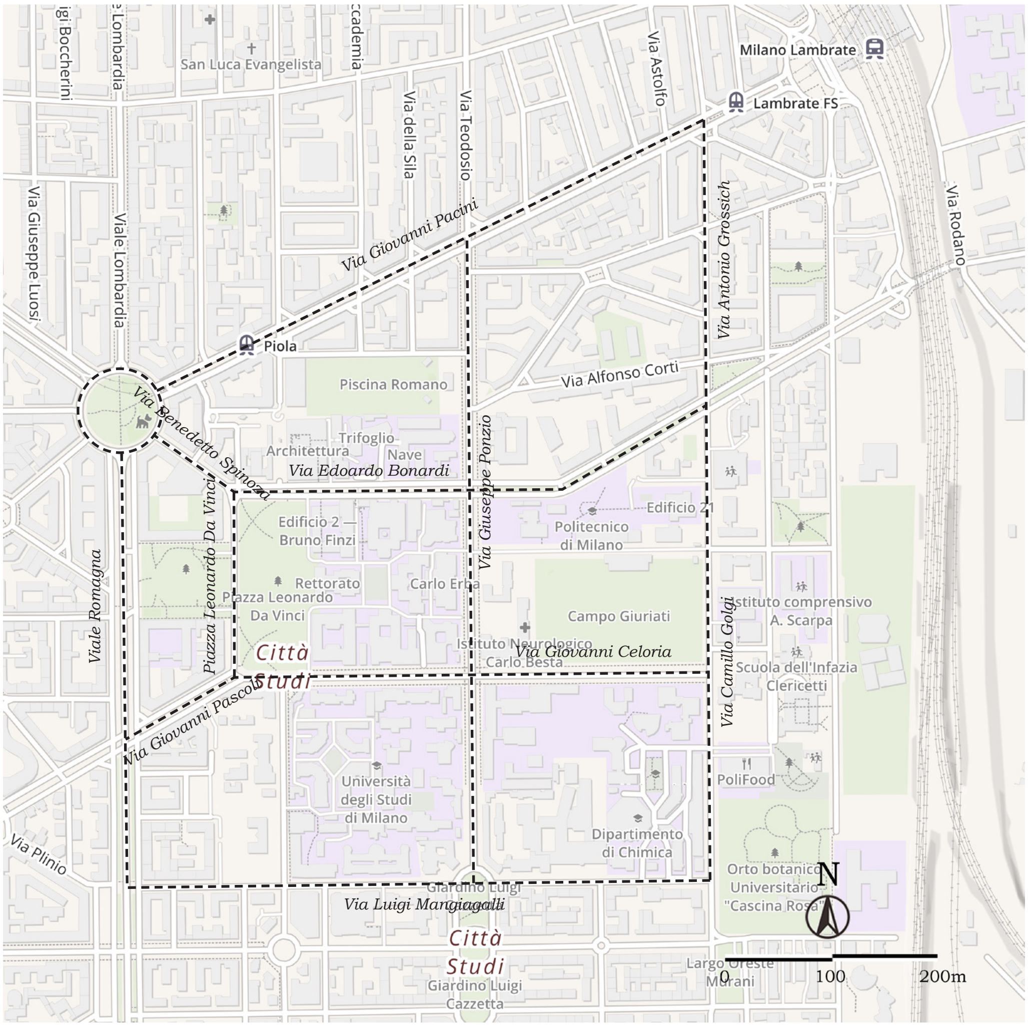


Figure 4.23
The 9 Streetss of Polimi

Again, we used the "point in a polygon" method to capture GSV images from ten streets around Politecnico di Milano. We focus on four classes, "tree", "sky", "car" and "building". Figure 4.24 - Figure 4.27 shows the details information of these classes. We hope to identify the spatial features of polimi from these details.

4.7.3 UrSA in Politecnico di Milano

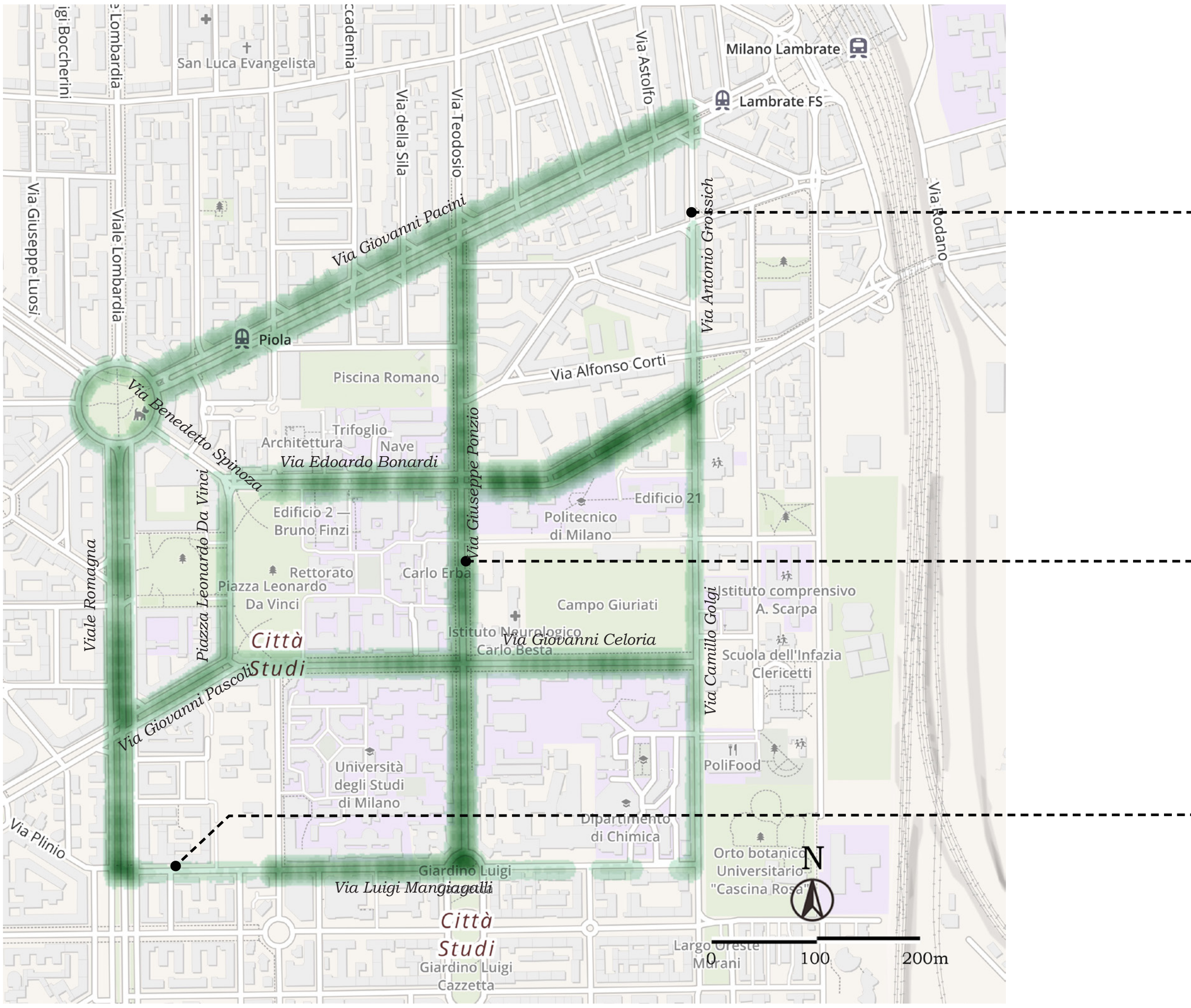


Figure 4.24
The Tree Value

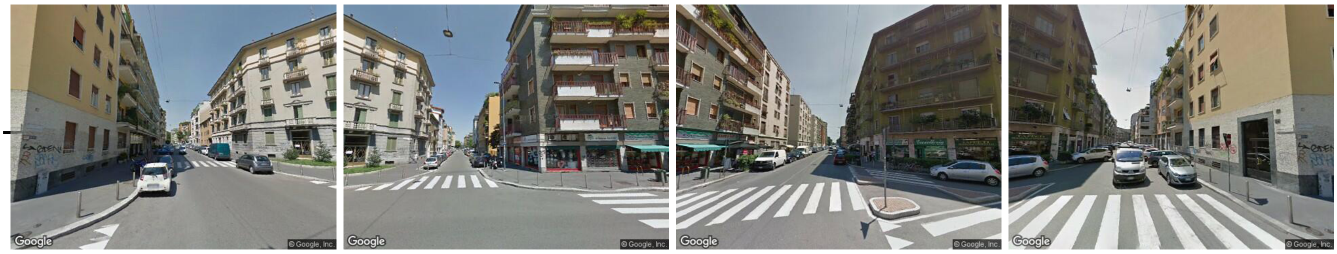
Low Medium High

The Tree Value

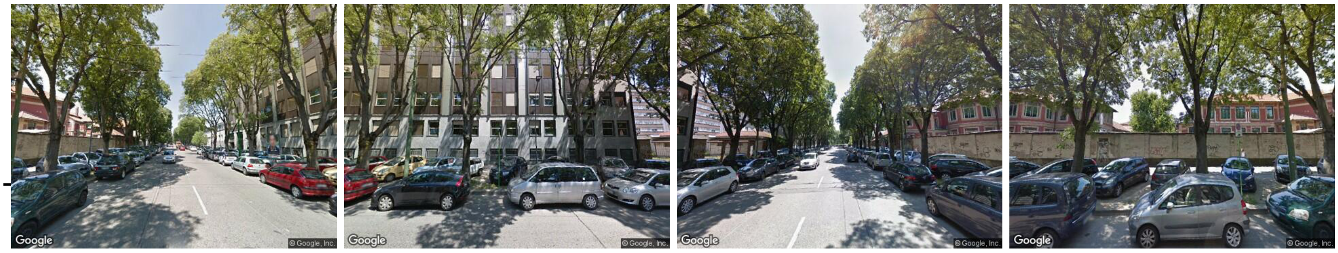
As you can see from Figure 4.24, the value of "tree" is different in different places. We can clearly identify where the "tree" is better. This is practical significance in guiding the urban greening design

The greening of Via Giuseppe Ponzio, Via Giovanni Celoria, Via Giovanni Pacini and Viale Romagna are obviously better than the other streets, while there is no continuous green interface in Via Luigi Mangiagalli and Via Edoardo Bonardi. Fifth Avenue and Sixth Street are moderate green coverage. Other streets are in a state of lack of "tree".

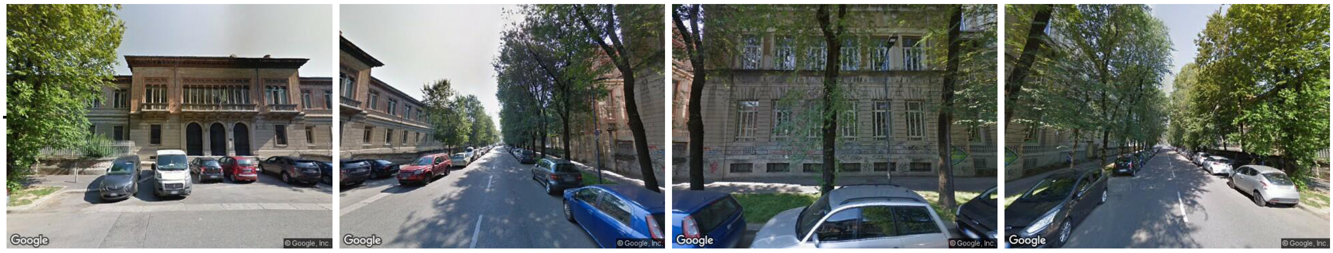
Via Antonio Grossich



Via Giuseppe Ponzio



Via Luigi Mangiagalli



4.7.3 UrSA in Politecnico di Milano

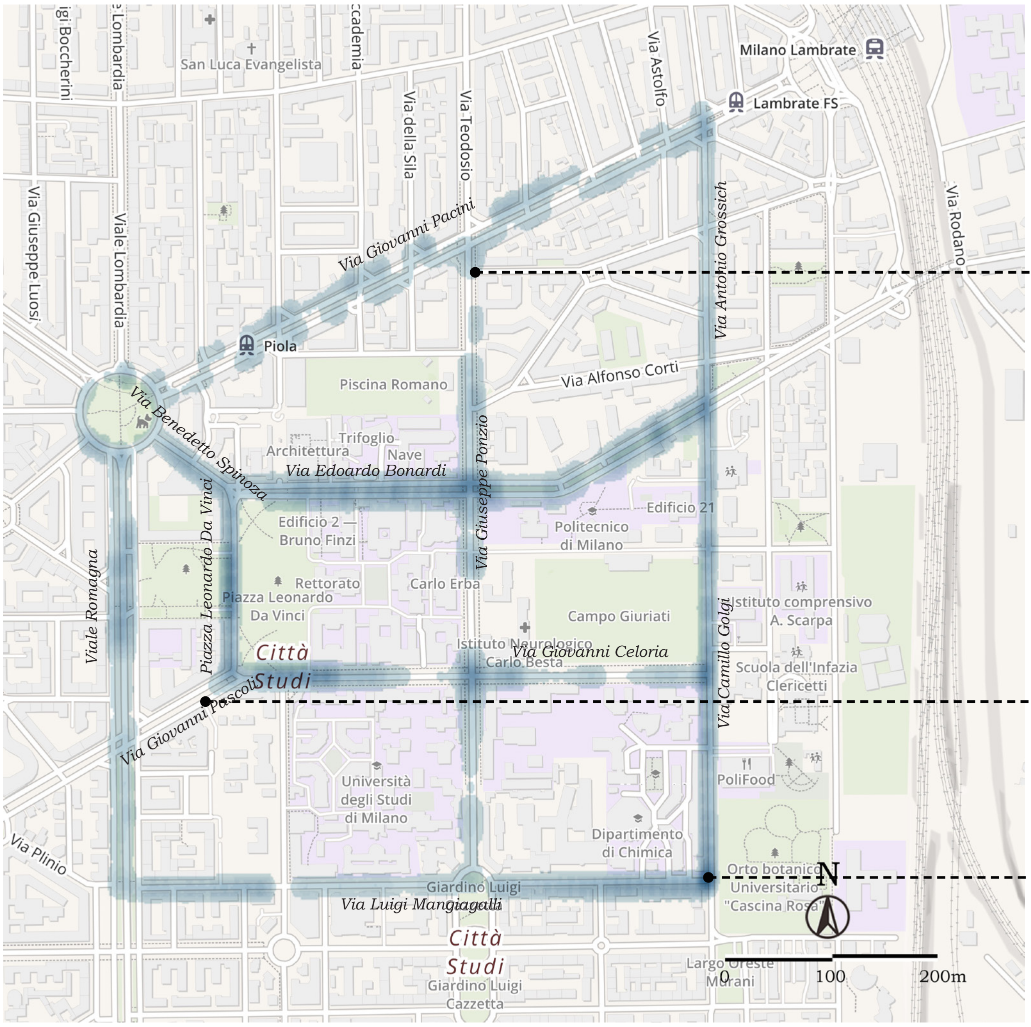


Figure 4.25
The Sky Value

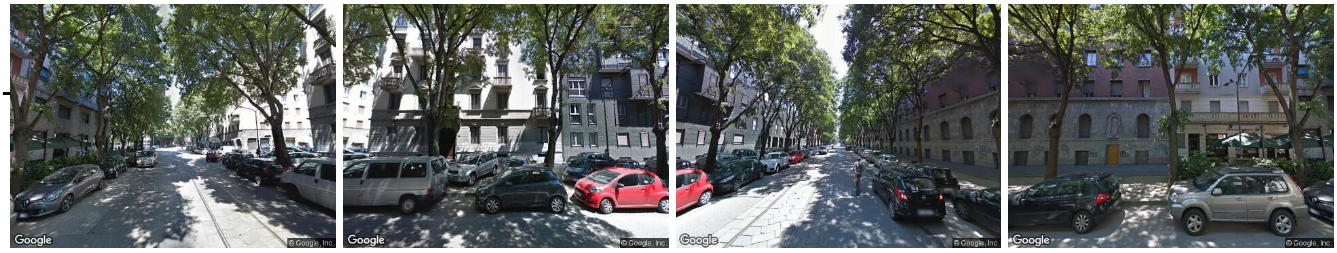
Low Medium High

The Sky Value

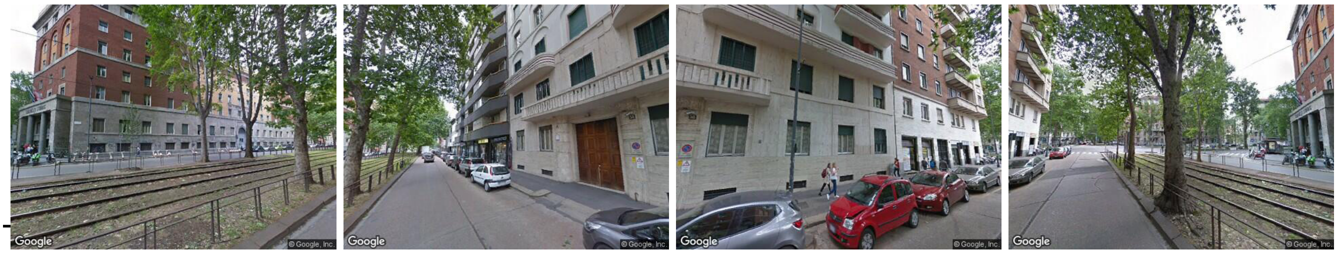
Figure 4.25 shows the distribution of “sky” in the street. In marked contrast to Figure 4.24, the value of “sky” is lower in streets with high “tree” values.

In Via Giuseppe Ponzio and Via Giovanni Pacini, the sky becomes partially visible. In the absence of trees in the streets, the “sky” becomes very high, which is to be expected, after all, the tree has a strong blockade function. Through UrSA, this could be a useful way to check the balance between “sky” and “tree”, when we make the urban street design.

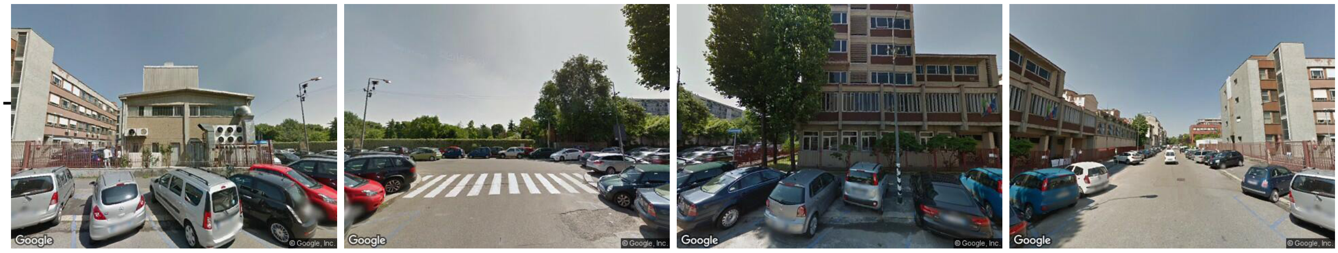
Via Giuseppe Ponzio



Via Giovanni Pascoli



Via Luigi Mangiagalli - Via Camillo Golgi



4.7.3 UrSA in Politecnico di Milano

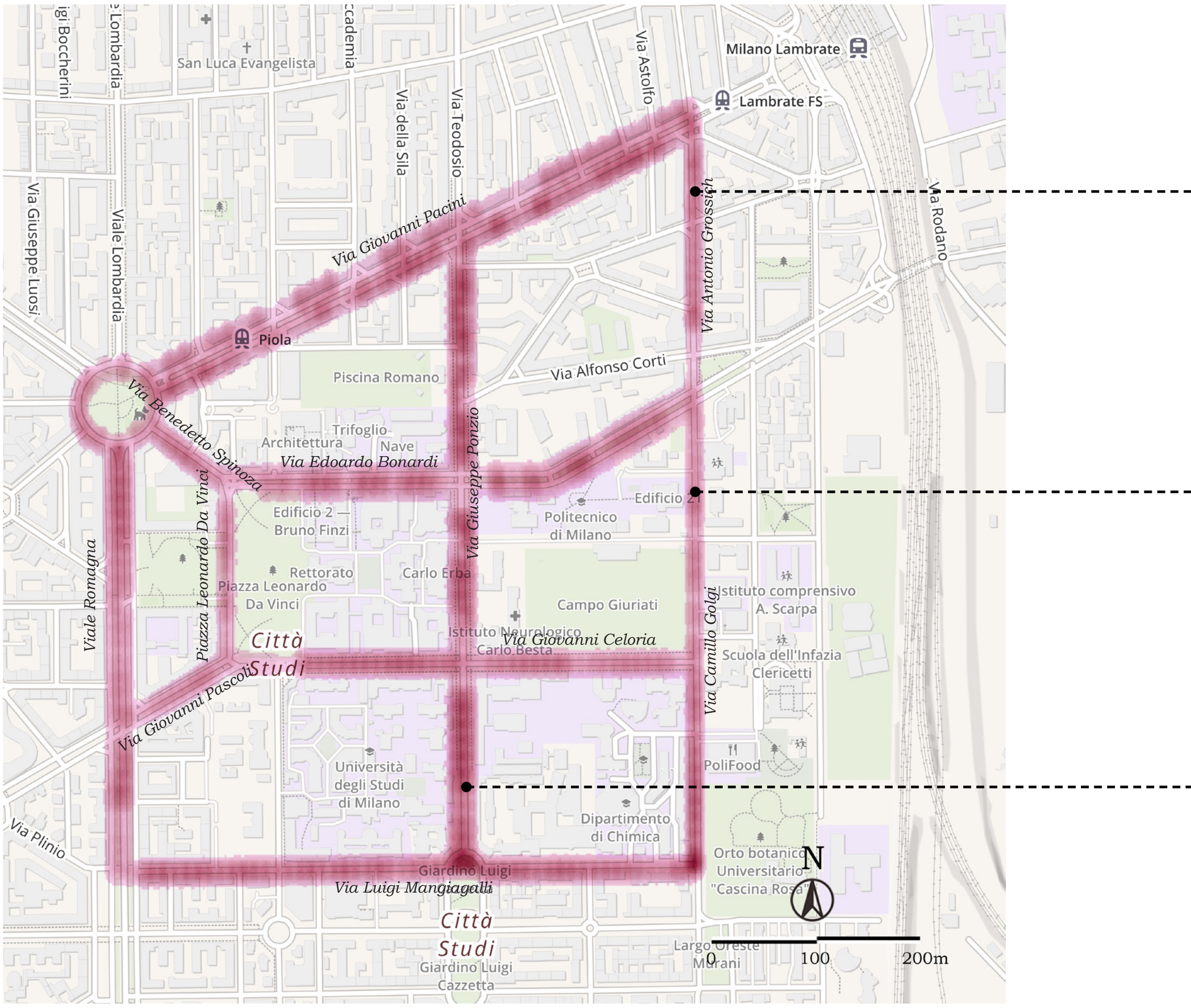
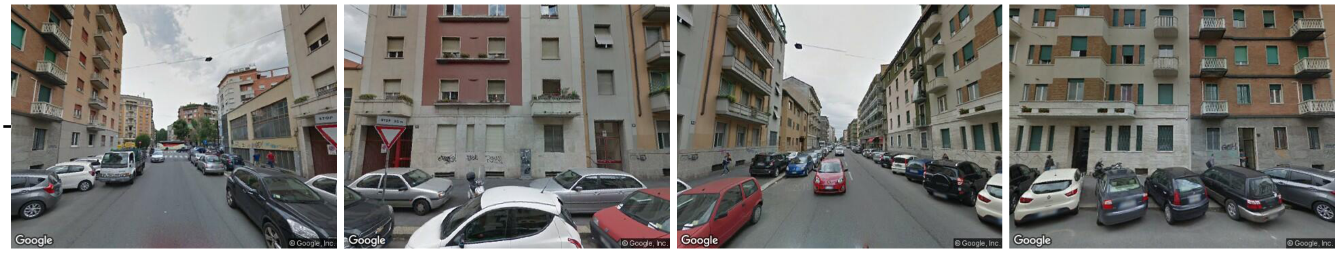


Figure 4.26
The Car Value

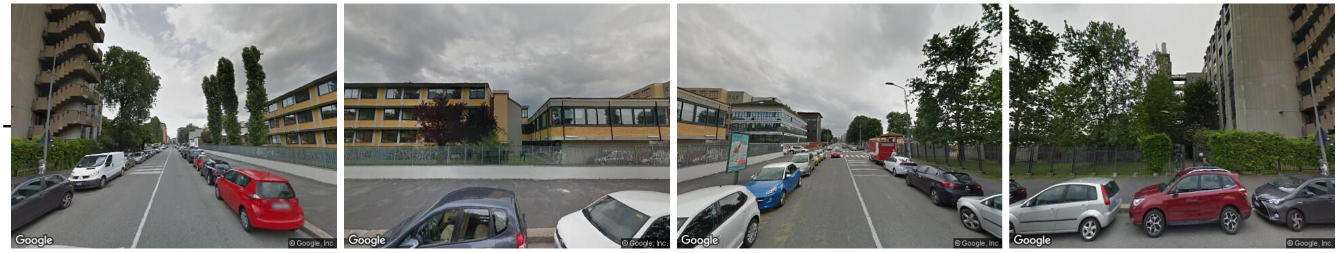
The Car Value

By the car value, we can analyze the parking situation in the city streets. From Figure 4.26, we can see parked full of cars on Polimi's street. This echoes the lack of indoor parking in the city. With UrSA, we can monitor street parking in the city. From the pictures in the next, we can see the true parking conditions in Via Antonio Grossich, Via Camillo Golgi and Via Giuseppe Ponzio.

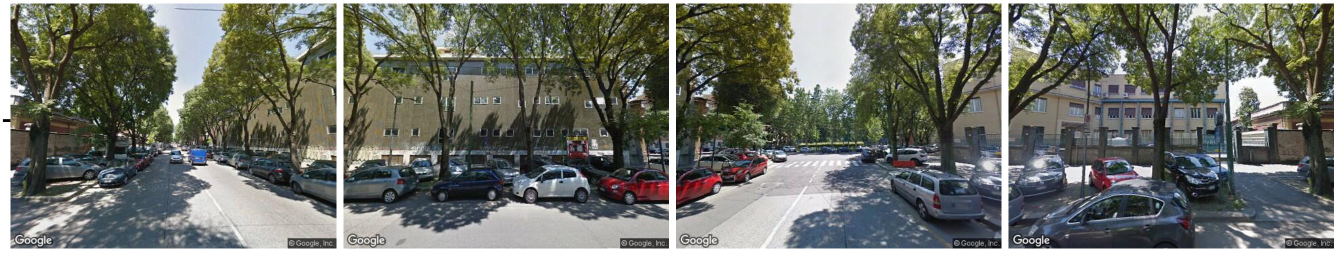
Via Antonio Grossich



Via Camillo Golgi



Via Giuseppe Ponzio



4.7.3 UrSA in Politecnico di Milano

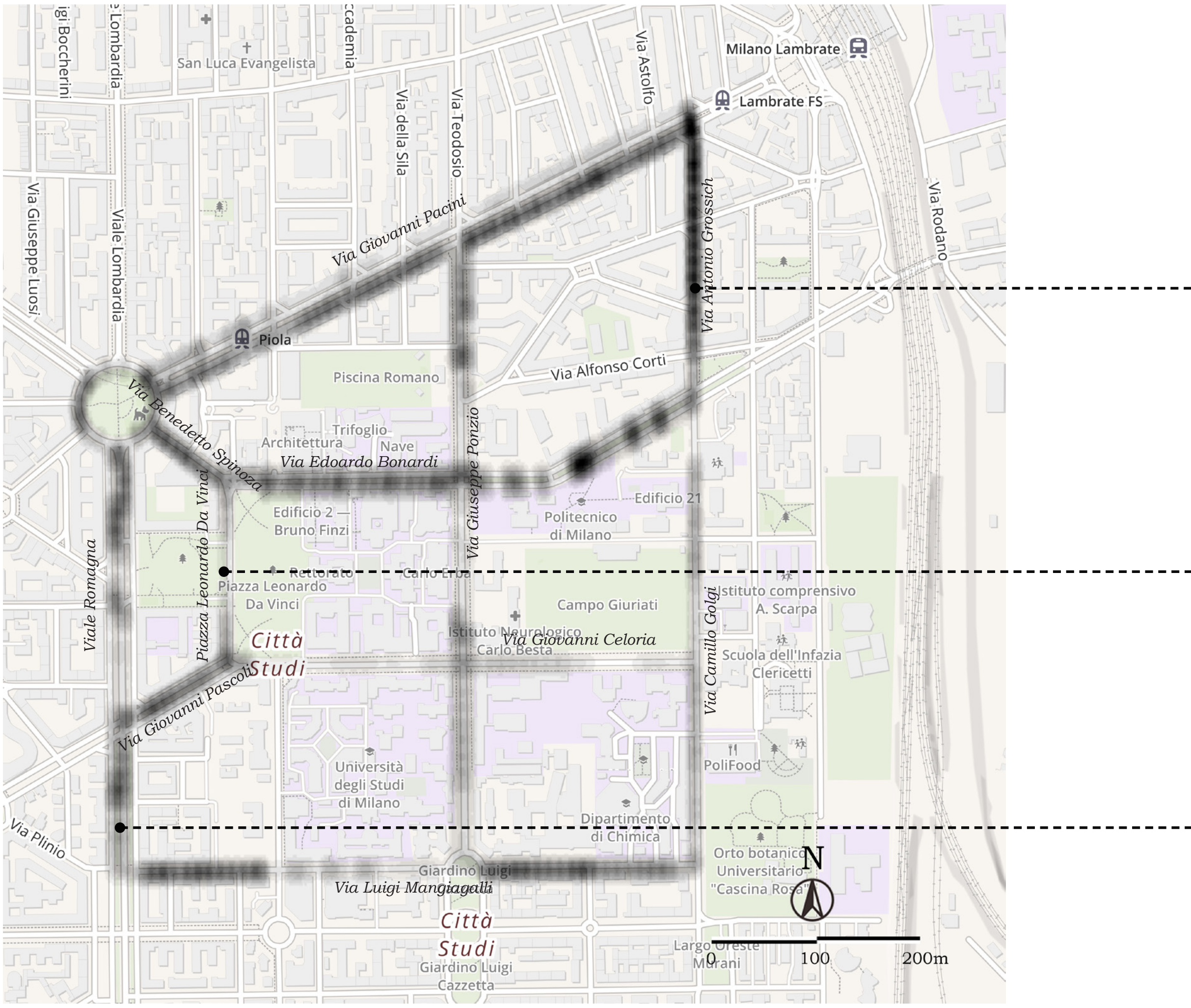


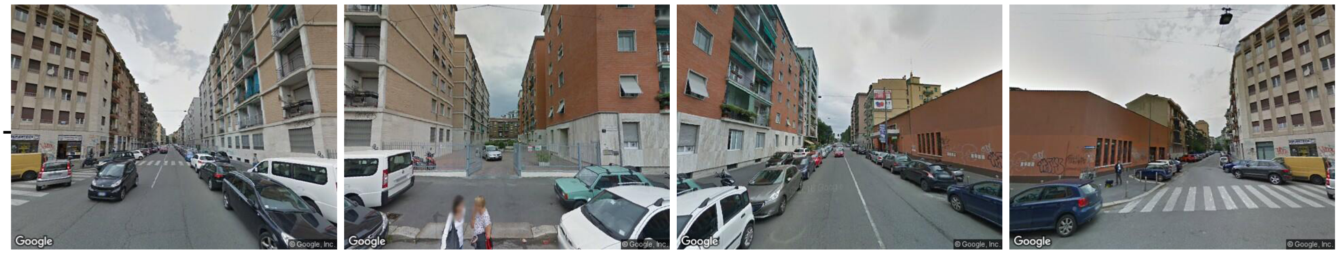
Figure 4.27
The Building Value

Low
Medium
High

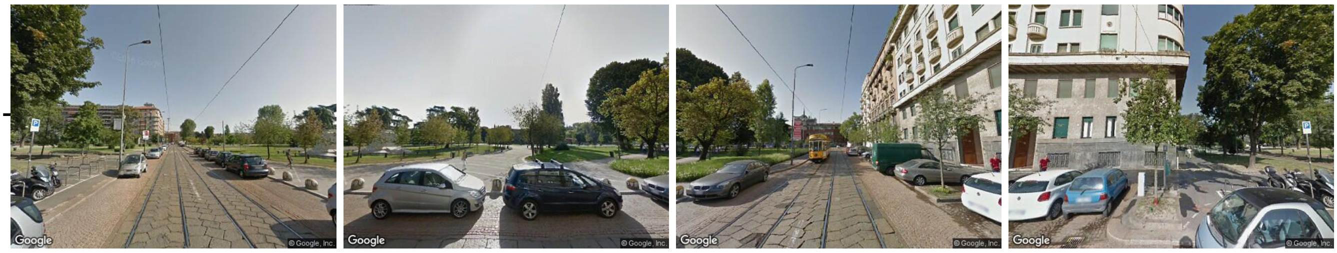
The Building Value

For the street environment, buildings around street are very important. Proper building dimensions help to create beautiful urban spaces. Figure 4.27 shows the building visible ratio. More dark color, the "building" value is higher, meanings there is a building closure space, such as Via Antonio Grossich. Of course, the low "building" value of Piazza Leonardo Da Vinci means that people can have a broader view there. And Viale Romagna shows a batter balance between "building", "sky" and "tree".

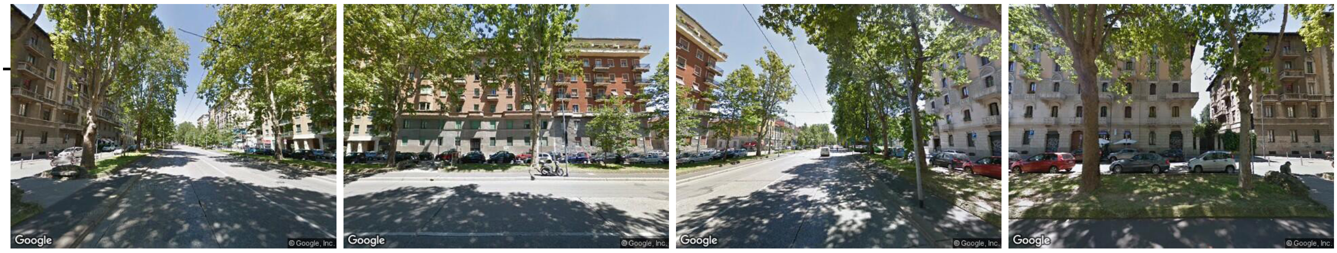
Via Antonio Grossich



Piazza Leonardo Da Vinci



Viale Romagna



References

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. International journal of computer vision, 60(2), 91-110.

Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on (Vol. 1, pp. 886-893). IEEE.

Ahonen, T., Hadid, A., & Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. IEEE transactions on pattern analysis and machine intelligence, 28(12), 2037-2041.

Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., & Gong, Y. (2010, June). Locality-constrained linear coding for image classification. In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on (pp. 3360-3367). IEEE.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine learning, 20(3), 273-297.

Ioffe, S., & Szegedy, C. (2015, June). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International Conference on Machine Learning (pp. 448-456).

Perronnin, F., Sánchez, J., & Mensink, T. (2010). Improving the fisher kernel for large-scale image classification. Computer Vision–ECCV 2010, 143-156.

Sivic, J., & Zisserman, A. (2003, October). Video Google: A text retrieval approach to object matching in videos. In null (p. 1470). IEEE.

Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1?. Vision research, 37(23), 3311-3325.

Chatfield, K., Simonyan, K., Vedaldi, A., & Zisserman, A. (2014). Return of the devil in the details: Delving deep into convolutional nets. arXiv preprint arXiv:1405.3531.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2015). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. arXiv preprint arXiv:1511.00561.

Websites:

http://mi.eng.cam.ac.uk/research/projects/VideoRec/CamVid/, retrieved on 15-11-2017

https://cn.mathworks.com/help/nnet/, retrieved on 02-12-2017

https://cn.mathworks.com/help/images, retrieved on 02-12-2017

https://cn.mathworks.com/help/vision, retrieved on 02-12-2017

Chapter Content

5 Project Practice

5 Project Practice

The content of chapter 5 elaborates the practice and application of Urban Street Analysis (UrSA), and what is the role of it; as well as the demonstration, analysis, and conclusion of the outcome. According to the analysis of street in which the social issues exist, and its human perception as a standard to explore the relationship in between the urban streets and social issues.

5.1 Methodological approach

For the technology part, we have a detail direction in chapter 4, here we explain the idea and its support concept and theoretical foundation.

5.1.1 Human Perception, Social Issues and Urban Street

First of all, urban planning is a comprehensive subject, it is related with many other related subjects, such as sociology, economics, politics, environment, ecology and so forth, and it also serves for urban development.

> *"In giving this sociological answer to the question: What is a City? [...] But if the city is a theatre of social activity, and if its needs are defined by the opportunities it offers to differentiated social groups, acting through a specific nucleus of civic institutes and associations, definite limitation on size follow from this fact. (Mumford, L. (1937). What is a city. Architectural record, 82(5), 59-62.)*

In the research project, we divided three parts in order to clear the research direction, which are human perception, social issues, and urban street. In the one hand, social issues and urban street are relevant and interactive from each other through the human behaviors; for instance, in general, the high housing price is related with a good street condition where the house is surrounded. On the other hand, social issues and urban streets have an influence on human perception; so people have a subjective judgment on it, thus the perception is a bridge between them.

So from the view of urban planning, our aims are trying to work on the relationships between social issues and urban streets; however, the forms of the urban street are infinite, which means lack of exactly objective standard and value to measure. So we take the human perception of urban street as a standard to link with social issues.

> *"Relatively low attractiveness of green areas within study area is highlighted by the fact that two-thirds of respondents named a favourite place outside the area, compared to 42% that named one within it." (Tyrväinen, L., Mäkinen, K., & Schipperijn, J. (2007). Tools for mapping social values of urban woodlands and other green areas. Landscape and urban planning, 79(1), 10)*

This was written in one of the Tyrvainen et al's (2007) questionnaire study, which was in Finland, and it worked on usage and perception issue and discovered that favorite areas were not much used frequently.

> *People's preference in terms of greenery and green spaces are vitally important in order to understand if people are to use and thereby benefit from green spaces. (Beaney, K. (2009). Green spaces in the urban environment: uses, perceptions and experiences of Sheffield city centre residents (Doctoral dissertation, University of Sheffield), 49)*

However, human perception is not only for vision but also for the sense of safety and may have an effect on the decision of crime.

> *"In terms of safety, the respondents preferred no enclosure (i.e. one side of pathway was open) and less vegetation under storey. However interactions between enclosure and the denseness of the under storey were apparent." (Beaney, K. (2009). Green spaces in the urban environment: uses, perceptions and experiences of Sheffield city centre residents (Doctoral dissertation, University of Sheffield), 51)*
>
> *"They restructure the physical layout of communities to allow residents to control the areas around their homes. This includes the streets and a ground outside [...] it depends on resident involvement to reduce crime and remove the presence of criminals [...]. The idea was to keep the grounds and the first floor free for community activity. "A river of trees" was to flow under the buildings [...] all the grounds were common and disassociated from the units, residents could not identify with them. The areas proved unsafe." (Newman, O. (1996). Creating defensible space, us department of housing and urban development. Office of Policy Development and Research, Rutgers University, 9-10)*

It proved that the physical urban form influences people' lifestyle and plays a role in reducing crime, and also the Crime prevention through environmental design (CPTED) that was came up with by criminologist C. Ray Jeffery supports this idea. On the other hand, we can see there is a relationship between the urban street and social issues, generally speaking, the form of the urban street is the reason, and the social issues are the result. In our application, we may find a certain relationship between these two objectives or not, but in the final, it is to unveil the facts of the urban street in the different context of the social issue, which is regarded as references and guidance to urban planning field.

5.1.2 The Sense of Visual Closure

The second part is taking the sense of visual closure to the urban street as a standard to find the relationship between social issues and urban streets.

From the view of human perception in vision and mentality to the external environment, Yoshinobu Ashihara, a Japanese architects, studied on urban design and spatial dimension in

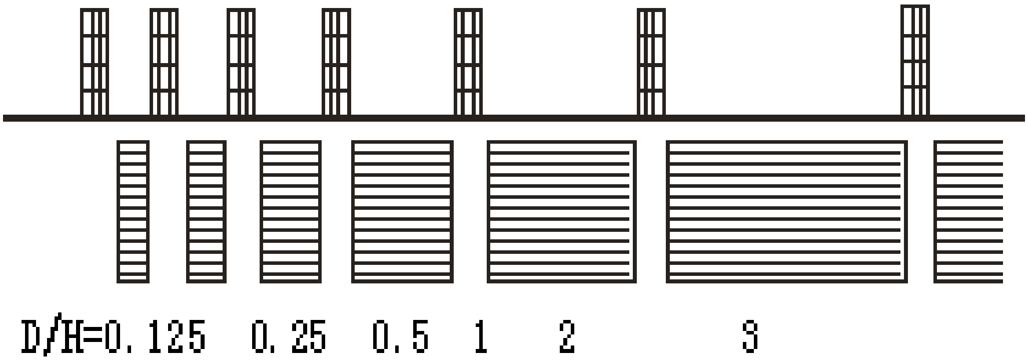


Figure 5.1
The index of street

Source: http://blog.sina.com.cn/s/blog_5938ea7101014tj0.html

1961s and pointed out that the external environment is created purposefully by people; it form from the interrelation of building and human perception, and the interrelation is mainly focused on visional and mentally perception. For instance, the book “Exterior Design in Architecture” written by Yoshinobu Ashihara, pointed out that “21-24m is the dimension that people can see each others’ face clearly if beyond this dimension can not. Thus in the street that the width is around 10m can increase the social elements, in another word, recognizing the pedestrian.” The spaciousness is the feature that enclosed by architecture, so a space enclosed by people is a certain enclosure space.

> *The index of street:*
>
> *H: the height of façade of architecture*
>
> *D: the width of street*
>
> *When D/H<1, perception of closure is intensive, and the void between the buildings is prefer to be seen as exit and entrance, the narrow space has a forward movement and depth of focus.*
>
> *When D/H=1, the height of building and width of street produce an symmetry, and the quality of space is a turning point, but for the perception is not apparently.*
>
> *When D/H>1, the sense of closure is few, and the bigger of index, the more commodious.*
>
> *(ASHIHARA, Y. (2006). The aesthetic townscape, 35)*

5.1.3 Visual Closure Index (VCI)

Closure sense of visual represents the human visional and metal perception to an urban street. In order to analyze the relationship between people and urban street, we propose Visual Closure Index (VCI) to estimate human perception to each Google Street View (GSV). We found that the most effective elements are building, sky and tree and these three elements play different roles enclosure space. The main closure effect due to the cover of building entity, and the semi-closure effect owing to tree and green that is semitransparent, and a street without cover produce a totally transparent environment. Thus, the more building, the more closure space in street; the more sky, the more transparent; and the tree and green is for adjusting the open degree and reducing the pressure, increasing the layers to space. From this point, VCI has these three parameters, which are the building view ratio (BVR), tree view ratio (TVR) and sky view ratio (SVR) and are based on a human angle of view, so we first obtained the eligible GSV, and then analyzed VCI of the image.

1) Building View Ratio

Definition: the ration of building coverage of a particular location in the common human eye diopter.

Formula: BVR = numbers of building pixel / total numbers of pixel of labeled image

Manifestation: the BVR presented here is on a scale of 0-0.7, "0" presents no buildings, "0.7" presents a closure sense of vision by surrounding buildings, and "1" does not exist in common.

According to the analysis of street canyon[1] effect (also known as an urban canyon), the higher of BVR, the less losing of heat, which contribute to the urban heat island effect, and vice versa. The same concept for the sense of visual closure, the higher of BVR, the more closure in street, and the more pressure for mentality.

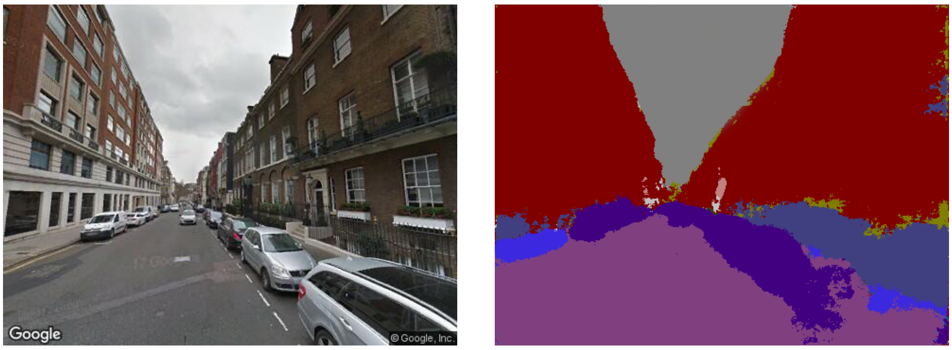


Figure 5.2
High Building Ratio

2) Sky View Ratio

Definition: the ration of sky view of a particular location in the common human eye diopter.

Formula: SVR = numbers of sky pixel / total numbers of pixel of labeled image

[1] *https://en.wikipedia.org/wiki/Street_canyon, retrieved on 03-12-2017*

Manifestation: the SVR presented here is on a scale of 0-0.7, "0" presents the sky are blocked totally, "0.7" presents a maximal sky view, and "1" does not exist in common.

According to the analysis of street canyon effect, the higher of SVR, the more of heat loss, so the city is easy to cool due to the sky absorbing the heat produced by building. On the contrary, the lower of SVR, the stronger of urban heat island effect. In terms of human perception, the higher SVR represents a high open degree of a street, but the extremely high SVR makes street alienated.

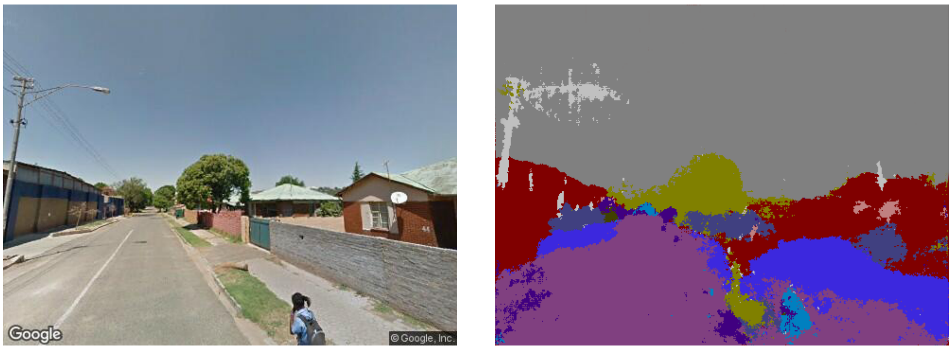


Figure 5.3
High Sky Ratio

3) Green View Ratio

Definition: the ration of tree coverage of a particular location in the common human eye diopter.

Formula: GVR = numbers of tree pixel / total numbers of pixel of labeled image

Manifestation: the GVR presented here is on a scale of 0-0.7, "0" presents the no trees, "0.7" presents a total coverage of tree, and "1" does not exist in common.

From the perspective of urban climate, the main part of natural elements is urban afforestation, and the high GVR means the high ability that urban streets absorb radiation to paly a role in cooling, as well as the photosynthesis. For the human perception, the high GVR represents the street space close to semi-en-

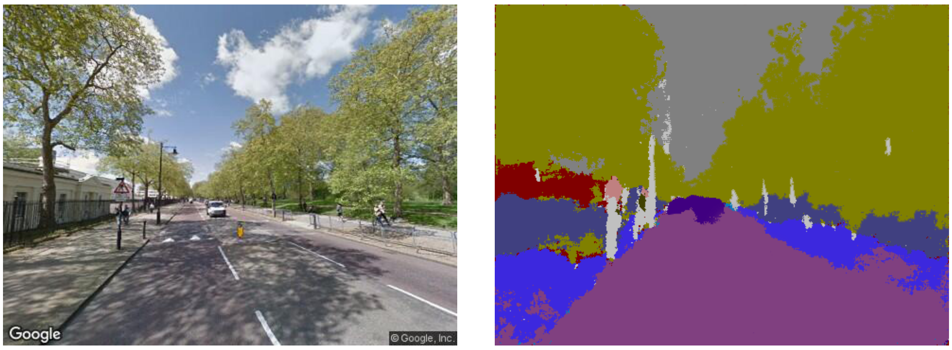


Figure 5.4
High Green Ratio

closed, but too much high GVR leads to lack of light, which making people feel like unsafe and a little depression.

4) Visual Closure Index (VCI)

No matter how high or low the closure degree is, it makes bad human perception in vision and mentality. Therefore, as shown in the figure.5.5 the lines of perception and degree of closure are in a single-peak way, with the rising degree of closure continuously, the comfort degree increasing in gradually until to the peak, and then decreasing.

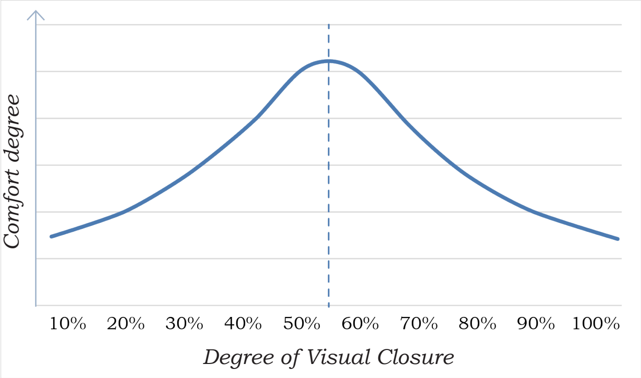


Figure 5.5 Diagram of feeling changed by the Visual Closure Index

By studying the information of more than 60,000 images, we conclude that there are four distinct combinations of the three parameters with VCI. They represent visual building closure, visual tree closure, visual open (sky) and visual open (tree).

Visual building closure

London	
file	51.50636,-0.150510924527_90_L.png
Bicyclist	0.12%
Building	46.35%
Car	7.44%
Fence	7.49%
Pavement	1.81%
Pedestrian	0.02%
Pole	0.16%
Road	22.09%
SignSymbol	0.14%
Sky	11.67%
Tree	2.72%

Visual tree closure

Milan	
file	45.472167,9.19293639668_180_L.png
Bicyclist	0.08%
Building	0.10%
Car	12.13%
Fence	2.05%
Pavement	0.10%
Pedestrian	0.01%
Pole	0.25%
Road	24.40%
SignSymbol	0.00%
Sky	1.91%
Tree	58.97%

Visual open (sky)

Johannesburg	
file	-26.19241,28.068416_90_L.png
Bicyclist	0.71%
Building	12.52%
Car	1.49%
Fence	3.06%
Pavement	7.91%
Pedestrian	0.10%
Pole	1.77%
Road	18.62%
SignSymbol	0.36%
Sky	49.08%
Tree	4.38%

Visual open (tree)

London	
file	51.50036,-0.136051097876_270_L.png
Bicyclist	0.05%
Building	3.43%
Car	2.41%
Fence	6.56%
Pavement	7.46%
Pedestrian	0.08%
Pole	2.59%
Road	23.07%
SignSymbol	0.44%
Sky	11.11%
Tree	42.79%

Table 5.1 Four typical visual closure classes

Table 5.2 The visual closure classes

Classes	Tree	Sky	Building
1	5.0%	5.0%	50.0%
2	50.0%	5.0%	5.0%
3	20.0%	20.0%	20.0%
4	5.0%	50.0%	5.0%
5	40.0%	15.0%	5.0%

We use the K-means algorithm to classify the image data, and Table 5.2 shows the K-means initialization settings. Classes are the number of clusters, Tree, Sky, and Building are the clusters' features.

Class 3 represents the best comfort situation of the human visual perception, and class1, 2, 4 and 5 correspond to "visual building closure", "visual tree closure", "visual open (sky)" and "visual open (tree)".

5.1.4 Step of analysis

Research factors:

Among all the social issues, we chose housing price and the crime rate that are relatively popular in the news and people care about much more than other issues because it is involved inhabitancy and safety issue.

Research setting:

According to the selected social issues, we found the global city ranking in housing price and crime rate and then decided the cities that near to the top and more typical, famous and powerful in the global level.

In each city, we found the distribution heat map of housing price and crime rate and selected 9 blocks in total in high, middle and low feature performance regions.

Street analysis:

Each block we took a certain number of street images from Google Street View (GSV), and then made a result, which is quantification and visualization of urban street analysis, through UrSA.

The sense of visual closure: according to the three parameters, which are BVR, TVR, and SVR, making a category of image.

Making a comparison with the heat map of social issues and the result of the sense of visual closure in the same street.

Conclusion:

Making a conclusion through the analysis.

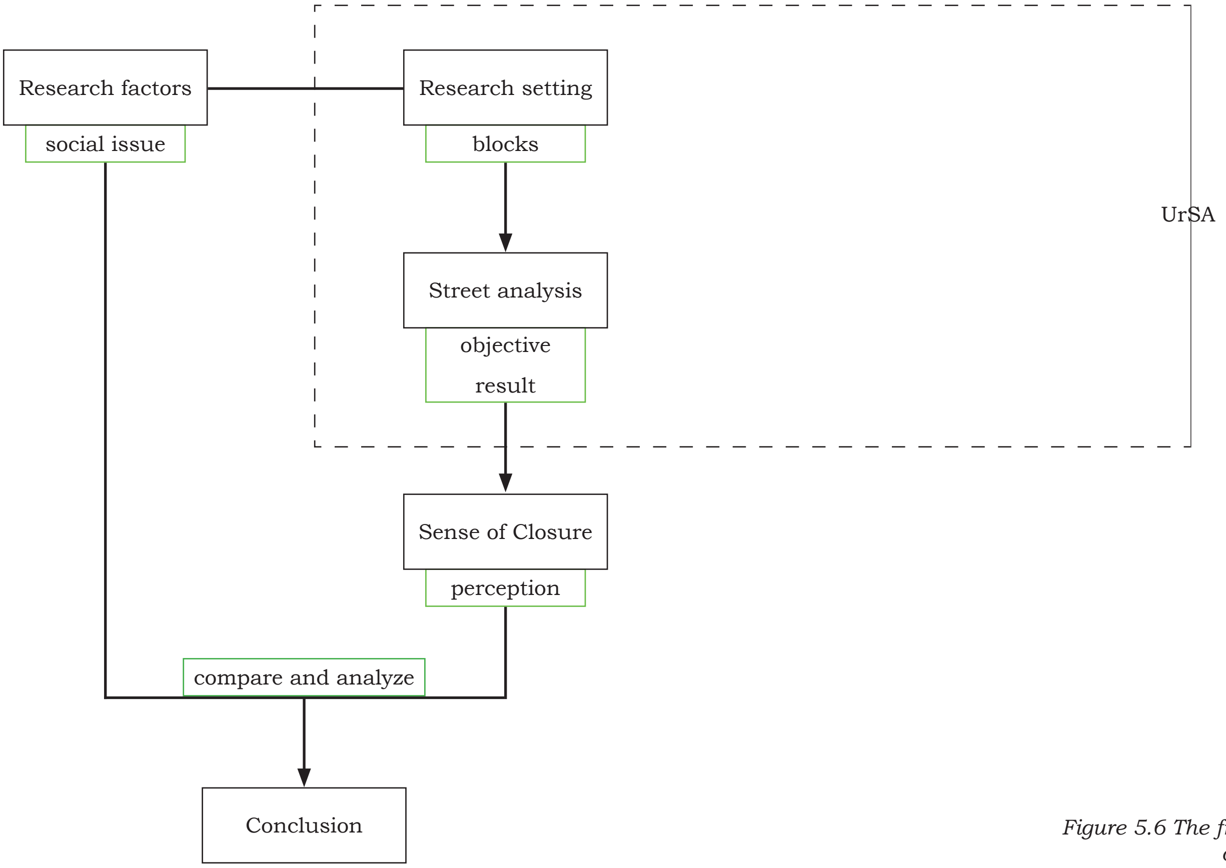


Figure 5.6 The framework of practice

5.2 Specific objectives of the research

The above paragraphs elaborated the general methodology in the research project. In this section we focused on specific objectives, which are housing price and crime rate, explain the reason that why we chose them, illustrate the state of them, as well as the city ranking of them.

5.2.1 Housing price

The high housing price has been one of the hottest topics in worldwide, the factors that influence housing price are divided into two parts in general.

The one is external factors, which is from population to society, from economy to policy, such as demography and the structure of the population, security of society, development of the economy, income of citizens, as well as policy and system of real estate and so forth.

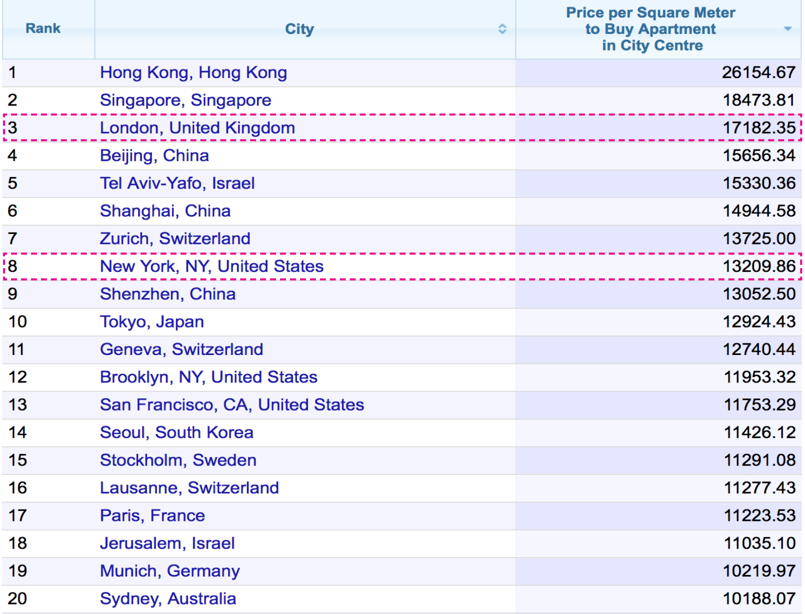
In the other hand, the factors depend on itself that includes location, entity, right, and interest, for instance, position, transportation, landscape, environment, story height and spatial features, etc. Our research is mainly in finding a relationship between housing price and urban street spatial features.

In the contrary, housing price produces an effect on the quality of life, economic competitiveness, social cohesion and the balance and diversity of the city. The essence of the city is active and healthy communities, and to satisfy the need of house at now and in future is the target of urban planning.

Thus, the housing price is interactive with urban planning in an intensive way.

According to the rules of United Nations Human Settlements Programme (UN-HABITAT) that the proportion of housing price and the income of local inhabitants is 3:1 at most, and 5:1 is from World Bank, which means the house price should be less than the three or five times of annual income of inhabitants. But in reality, there is a huge gap in housing affordability, which is variable in time and region.

According to numerous studies in the literature, urban planning norms have an important effect on housing prices. A large quantity of data and empirical studies focusing on building restrictions and housing prices [...] the relationship *between city-planning and the cost of housing [...] to analyze the effect of urban planning norms on housing (new and existing units) price variation. (Lorè, I. (2016). The Development and Promotion of the Inland Areas of the Metropolitan City of Reggio Calabria through the Enhancement and Restoration of the Calabro-lucane Railway Line–The Greenway Project, Train-hotel and Valorization of Former Railway Stations. Procedia-Social and Behavioral Sciences, 223, 363-370, 25-30)*



Rank	City	Price per Square Meter to Buy Apartment in City Centre
1	Hong Kong, Hong Kong	26154.67
2	Singapore, Singapore	18473.81
3	London, United Kingdom	17182.35
4	Beijing, China	15656.34
5	Tel Aviv-Yafo, Israel	15330.36
6	Shanghai, China	14944.58
7	Zurich, Switzerland	13725.00
8	New York, NY, United States	13209.86
9	Shenzhen, China	13052.50
10	Tokyo, Japan	12924.43
11	Geneva, Switzerland	12740.44
12	Brooklyn, NY, United States	11953.32
13	San Francisco, CA, United States	11753.29
14	Seoul, South Korea	11426.12
15	Stockholm, Sweden	11291.08
16	Lausanne, Switzerland	11277.43
17	Paris, France	11223.53
18	Jerusalem, Israel	11035.10
19	Munich, Germany	10219.97
20	Sydney, Australia	10188.07

Table 5.3 Prices by City of Price per Square Meter to Buy Apartment in City Centre (Buy Apartment Price)

Source: https://www.numbeo.com/cost-of-living/prices_by_city.jsp?itemId=100&displayCurrency=USD

Table 5.3 shows the prices by the city of price per square meter to buy an apartment in city center. Even the cities in top 20, there are a lot of different, especially the first, Hongkong ($26154.67 per square meter) and the second, Singapore ($18473.81 per square meter), the gap between those two cities is almost to ten thousand dollars. However from the first, the housing price of following cities are less in gradually, and the difference is not that much as it in top 2.

5.2.2 Crime rate

The crime issue has been always highlighted in society, in the 1970s, the book "Crime Prevention Through Environmental Design"[2] came out, the writer is C. Ray Jeffery, who was a criminologist. The CPTED strategies are based on the ability to influence offender decisions that before criminal acts and study in criminal behavior. The research shows that the decision to offend or not to is affected by cues of perceived risk. Moreover, the strategies cannot be fulfilled without the community's

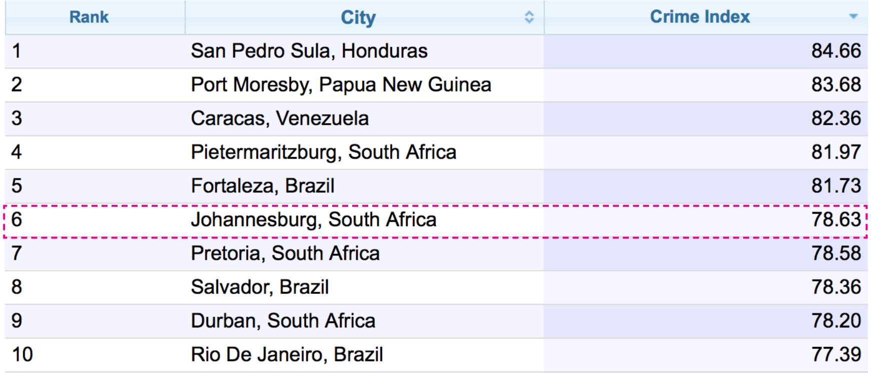
[2] *https://en.wikipedia.org/wiki/Crime_prevention_through_environmental_design, retrieved on 03-12-2017*

help, which means it needs the whole community was to make the environment a safe place together. A meta-analysis of multiple-component CPTED initiatives in the United States has found that they have decreased robberies between 30% and 84% (Casteel and Peek-Asa, 2000).

> *The rapid pace of urbanization coupled with the growth in city size and density is associated with increased crime and violence. Poor urban planning, design and management play a role in the shaping of urban environments that put citizens and property at risk. The fabric and layout of cities impact on the movements offenders and victims and on opportunities for crime. (Habitat, U. N. (2007). 3. Enhancing Urban Safety and Security: Global Report on Human settlements. Earth Scan.)*

The report from UN-HABITAT pointed out that the crime rate is linked to urban planning. Poor urban planning, design, and management have been playing a role in the shaping of urban and its environment that put citizens at risk, as well as property. Thus, the physical fabric and layout of a city have an influence on routine movements of victims and also on the possibility of crime.

The effective urban planning and government should maintain the environment in the ways that prefer to reduce or eliminate the opportunities for crime. Especially landscape maintenance, lighting and other physical elements for public activities, which have variable effects on the crime rate.



Rank	City	Crime Index
1	San Pedro Sula, Honduras	84.66
2	Port Moresby, Papua New Guinea	83.68
3	Caracas, Venezuela	82.36
4	Pietermaritzburg, South Africa	81.97
5	Fortaleza, Brazil	81.73
6	Johannesburg, South Africa	78.63
7	Pretoria, South Africa	78.58
8	Salvador, Brazil	78.36
9	Durban, South Africa	78.20
10	Rio De Janeiro, Brazil	77.39

Table 5.4 Crime Index 2017 Mid-Year

Source: https://www.numbeo.com/cost-of-living/rankings.jsp

From the table, we can see that the most of high crime rate cities are in South Africa and Brazil. According to the crime rate ranking (Table 5.4), the majority of the city in top 10 are in South Africa, such as Pietermaritzburg, Johannesburg, Pretoria and Durban; and Fortaleza, Salvador, and Rio De Janeiro, which are in Brazil. The most famous risky country is South Africa, which is owing to greater access to firearms, high unemployment and wealth gap. As for Brazil, the causation not only many people live in poverty, but also the favelas. Besides, there are also some high crime rate cities in the United States, such as Detroit (15), New Orleans (19), Baltimore (20) and Chicago (34); and the lowest crime rate cities in America are Boise (327), Salt Lake City (286), Boston (242) and San Diego (231); and New York (152) is near to the high crime rate. Thus there is an obvious disparity in crime rate in the United States.

This part section we introduced simply specific research objectives and its global rankings and performances in cities. Next, we combined the specific research objectives and urban scene analysis technology, through a plenty of data collection, quantification and analysis to achieve the goal which is to “read” the city and its relationship with social problems, and also to assist planner to recognize the city and as a reference and guidance to urban planning.

5.3 Application of Research 1 - Housing Price and UrSA

Using UrSA to study the spatial characteristics of streets in different housing price areas in cities, we try to find out the relationship between housing prices and urban space. In order to demonstrate better the relationship between the two, we introduce VCI as a bridge between UrSA and urban housing prices.

VCI quantifies people’s visual perception of the street, so we can get the relationship between them by comparing VCI values and prices in the same area. Then we can get the link between house prices and street features.

We chose London and New York as research cities, and as a global metropolis, housing prices are prominent in both cities. In 2016, London was third in the list with a price of $17182.35 per square meter and New York City ranked eighth in the list with a price of $13209.86 per square meter, respectively, the highest housing priced cities in Europe and the America.

We select high-price area, mid-price area and low-price area in each city. Each price level has three regions, so as to maximize the sample’s comprehensiveness.

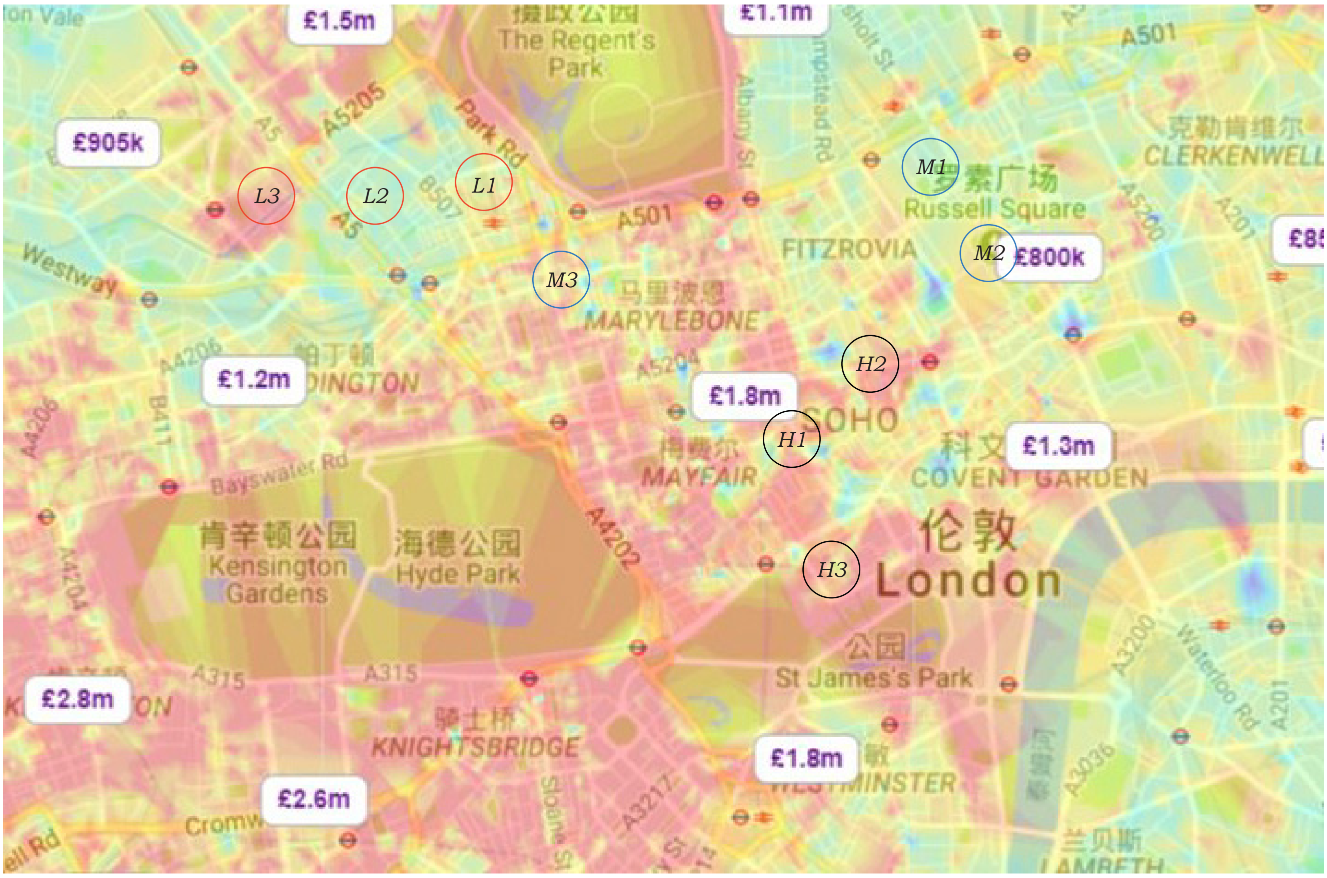


Figure 5.7 Heatmap of London's property values

Source: https://www.zoopla.co.uk/heatmaps/

○ Region of high housing price

○ Region of middle housing price

○ Region of low housing price

High Med Low

5.3.1 London Housing Price Distribution and UrSA

From the results of UrSA (Bar chart 5.1), we can see that the higher the price, the lower the "sky" and "tree" values. In the high-price area, the proportion of "building" is higher than 45%, H1 is more than 50%. Such a street environment leads to a higher VCI value.

In the mid-priced region, UrSA's results show that the "building" value here is significantly lower than the high-priced region, maintaining at about 35%. At the same time, the proportion of "tree" has improved, but "sky" is still at a low level. Such a space environment is reflected in Figure 5.8, that is, M1, M2, and M3 are middle VCI values for some areas only.

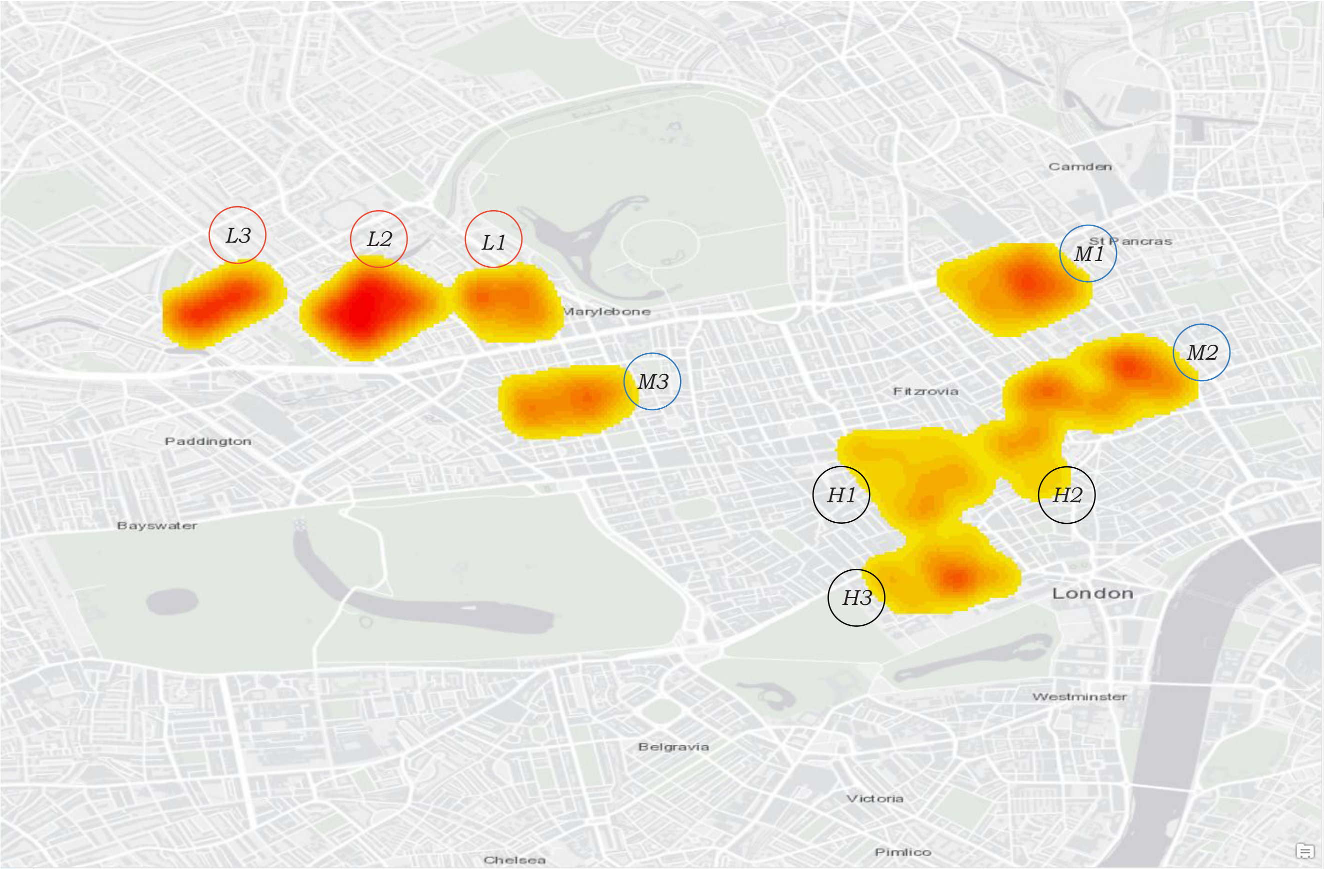


Figure 5.8 Heatmap of London's VCI values

○ Region of high housing price

○ Region of middle housing price

○ Region of low housing price

High Medium Low

In the low-price area, we found that the value of "sky" has increased a lot, revealing that the street spaces of L1, L2, and L3 are more open, especially at L2 and L3. Meanwhile, we can see from Bar chart 5.3.L3, there are more than 35% of the "tree" in the L3 region, which is unique across all nine regions. In low-price block, VCI values have also become much lower, reaching a visually open effect. This is a big difference between the visual enclosure of high-priced areas.

Figure 5.17 shows the street features distribution in different housing price region of London, it could be an overall analysis of this part.

● ○ ○ UrSA in high housing price region

selected region H1

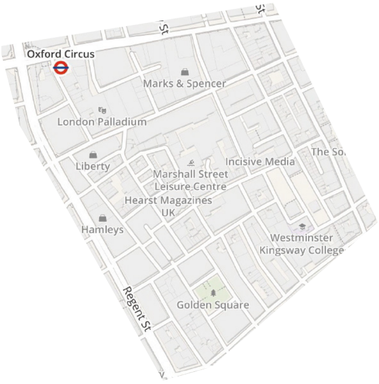


Figure 5.9.H1 Site of high housing price

Coordinate

51.515528, -0.142694
51.516139, -0.135500
51.512556, -0.133028
51.510472, -0.137417
51.510194, -0.138333

Area

213.655,14 m²

Number of image

Collecting: 2136
Processing: 1540

---

selected region H2

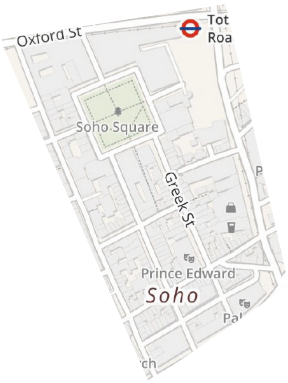


Figure 5.9.H2 Site of high housing price

Coordinate

51.516167, -0.134333
51.516583, -0.130222
51.513194, -0.128944
51.512167, -0.131556

Area

94.406,96 m²

Number of image

Collecting: 892
Processing: 624

---

selected region H3

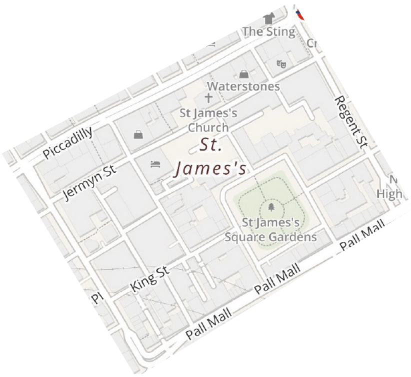


Figure 5.9.H3 Site of high housing price

Coordinate

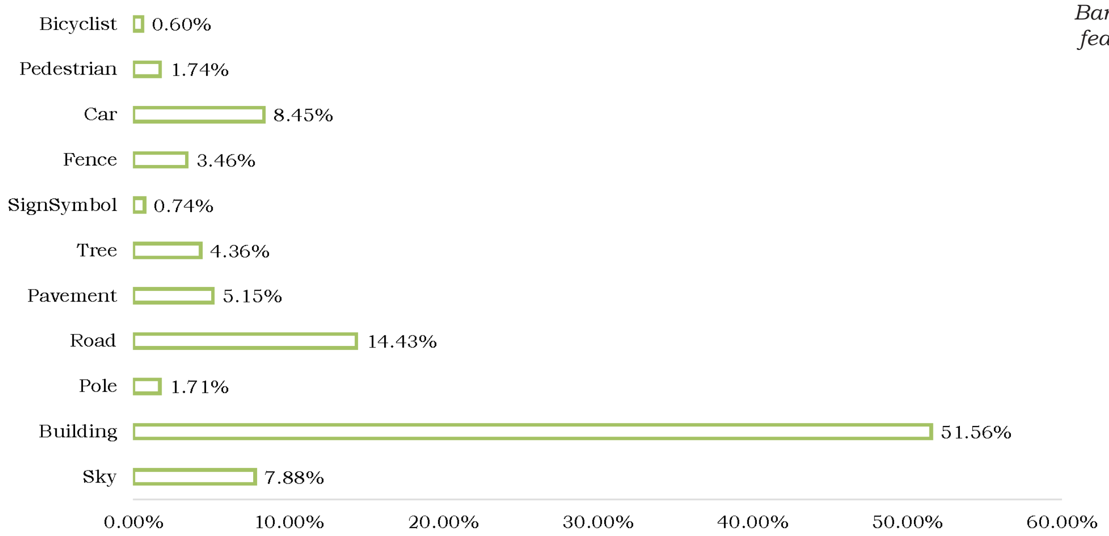
51.510056, -0.134861
51.507111, -0.132000
51.505000, -0.137972
51.507889, -0.140944

Area

157.008,90 m²

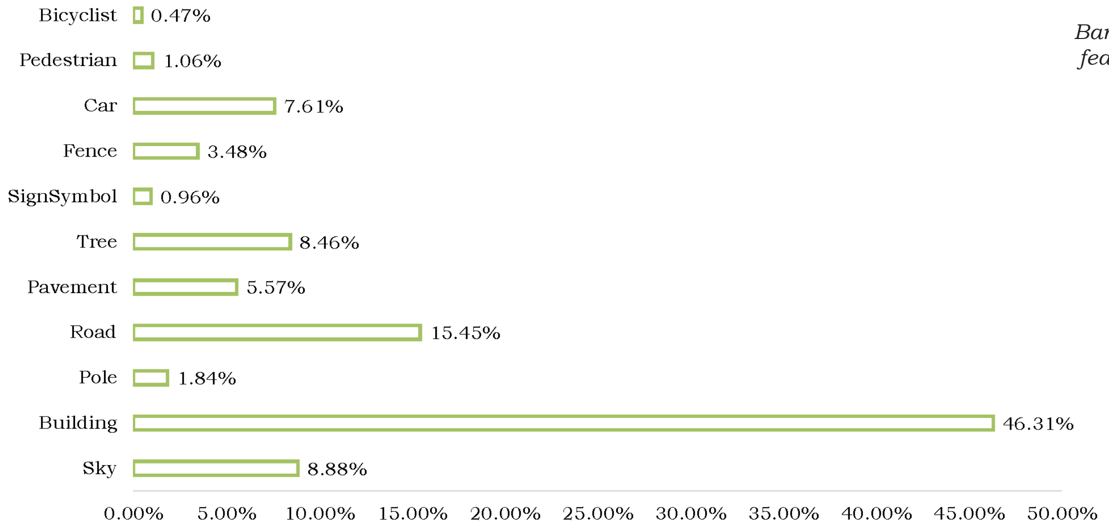
Number of image

Collecting: 1536
Processing: 1192



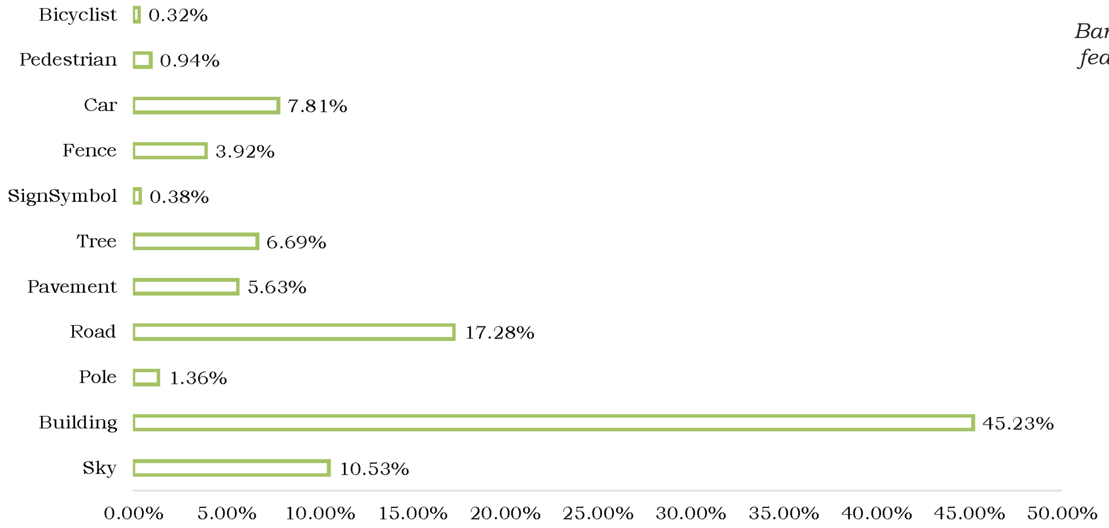
Bar chart 5.1.H1 Street features distribution in high housing price

H1



Bar chart 5.1.H2 Street features distribution in high housing price

H2



Bar chart 5.1.H3 Street features distribution in high housing price

H3

○ ● ○ UrSA in middle housing price region

selected region M1

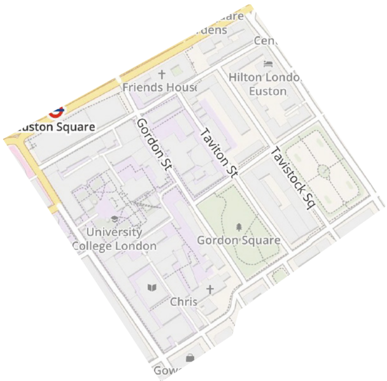


Figure 5.10.M1 Site of middle housing price

Coordinate

51.527658, -0.130691
51.524703, -0.127460
51.522360, -0.132653
51.525791, -0.136451

Area

172.382,36 m²

Number of image

Collecting: 1588
Processing: 1352

selected region M2

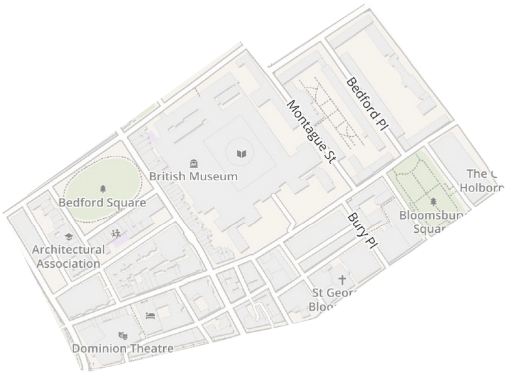


Figure 5.10.M2 Site of middle housing price

Coordinate

51.521682, -0.124237
51.519037, -0.120719
51.517006, -0.125406
51.516333, -0.130551
51.518578, -0.132444
51.521101, -0.126886
51.520674, -0.126328

Area

248.787,93 m²

Number of image

Collecting: 2212
Processing: 1492

selected region M3

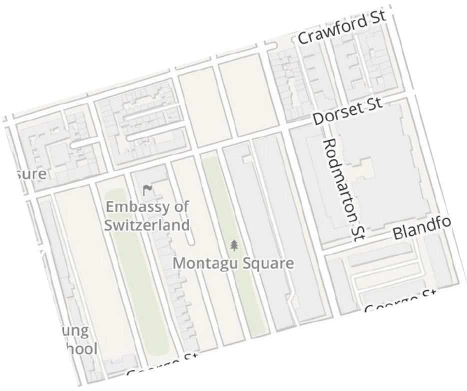


Figure 5.10.M3 Site of middle housing price

Coordinate

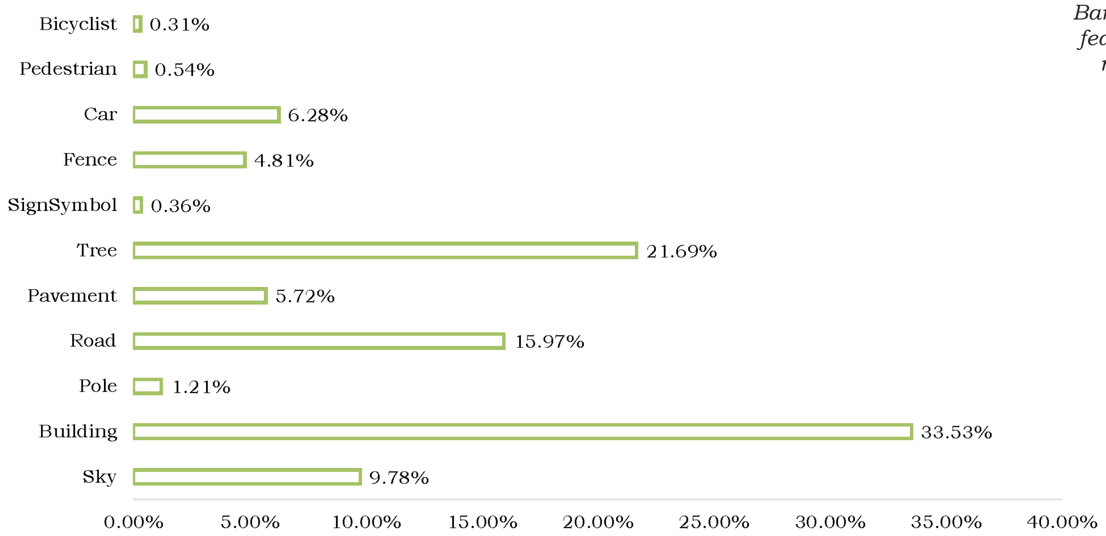
51.519611, -0.156155
51.517341, -0.155136
51.516313, -0.161884
51.519197, -0.163182

Area

170.628,47 m²

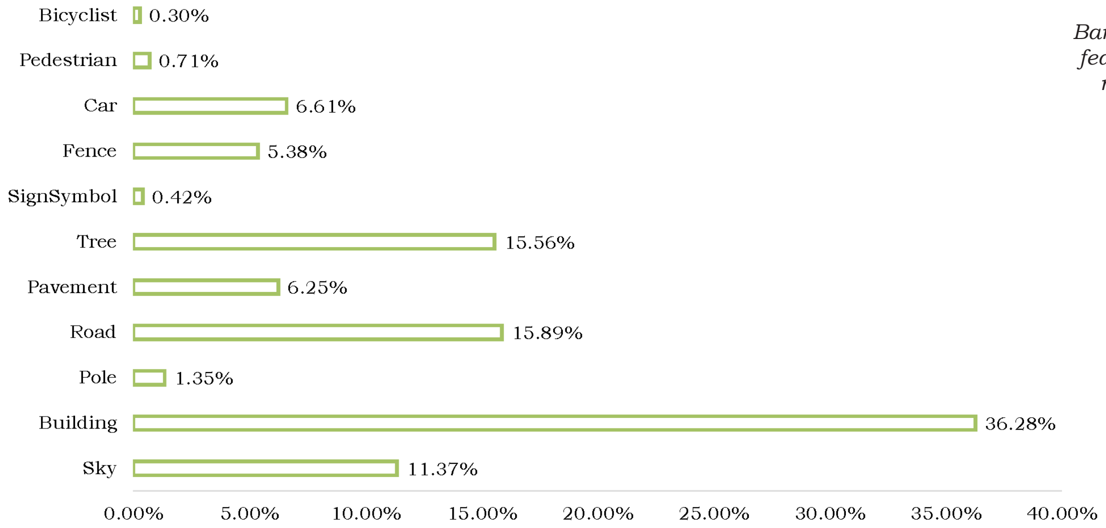
Number of image

Collecting: 1184
Processing: 1056



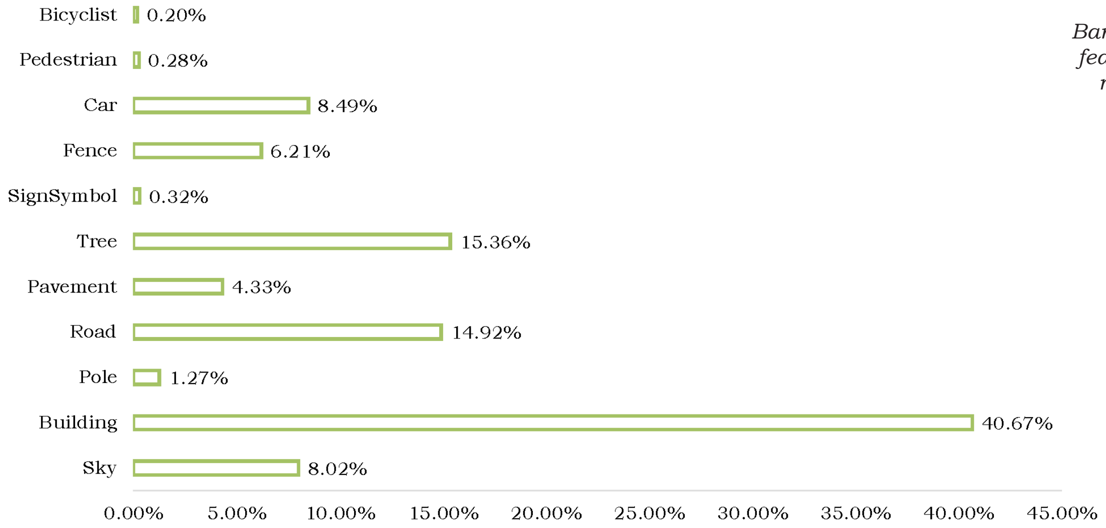
Bar chart 5.2.M1 Street features distribution in middle housing price

M1



Bar chart 5.2.M2 Street features distribution in middle housing price

M2



Bar chart 5.2.M3 Street features distribution in middle housing price

M3

UrSA in low housing price region

selected region L1

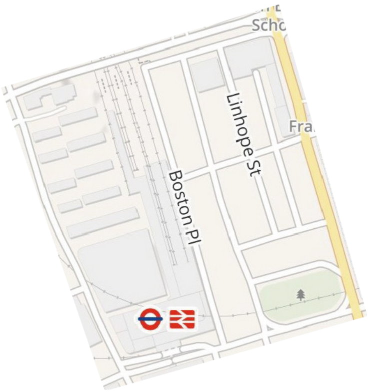


Figure 5.11.L1 Site of low housing price

Coordinate

51.525755,	*-0.161409*
51.522644,	*-0.159886*
51.522061,	*-0.163918*
51.524925,	*-0.165727*

Area

97.477,23 m²

Number of image

Collecting:	*880*
Processing:	*720*

selected region L2

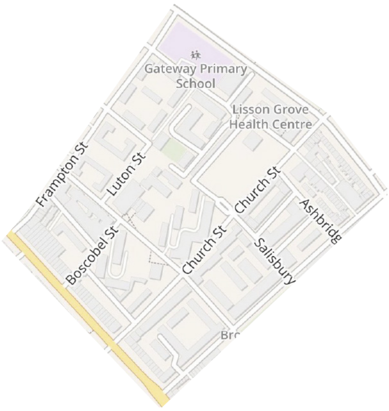


Figure 5.11.L2 Site of low housing price

Coordinate

51.526292,	*-0.171353*
51.524871,	*-0.168146*
51.523810,	*-0.166848*
51.521098,	*-0.171773*
51.523310,	*-0.175072*

Area

162.124,54 m²

Number of image

Collecting:	*1412*
Processing:	*1316*

selected region L3

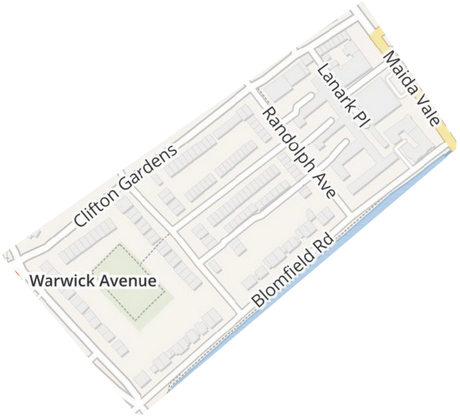


Figure 5.11.L3 Site of low housing price

Coordinate

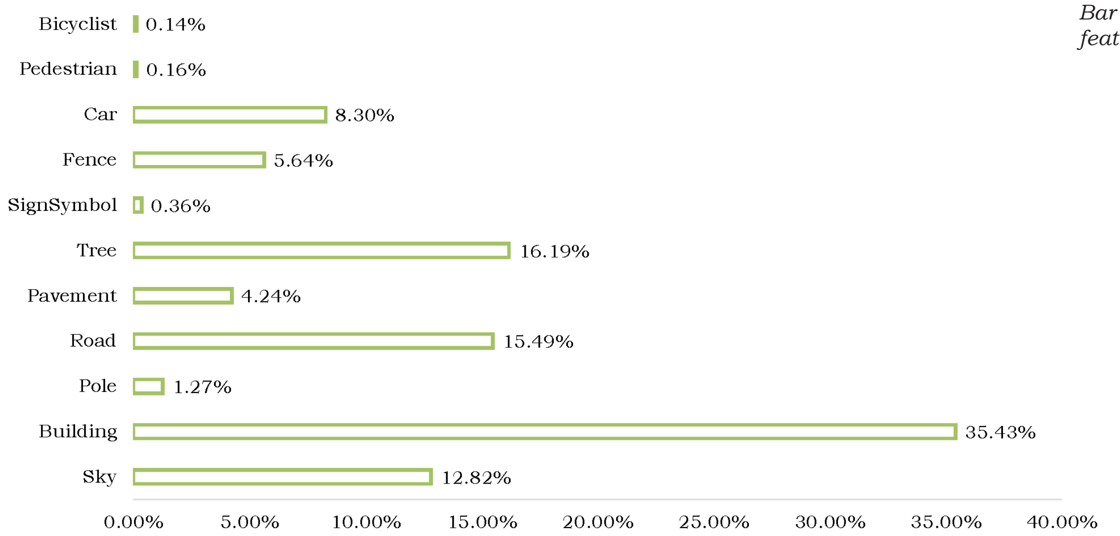
51.526179,	*-0.178543*
51.524613,	*-0.176633*
51.521963,	*-0.181955*
51.523565,	*-0.184015*

Area

103.275,01 m²

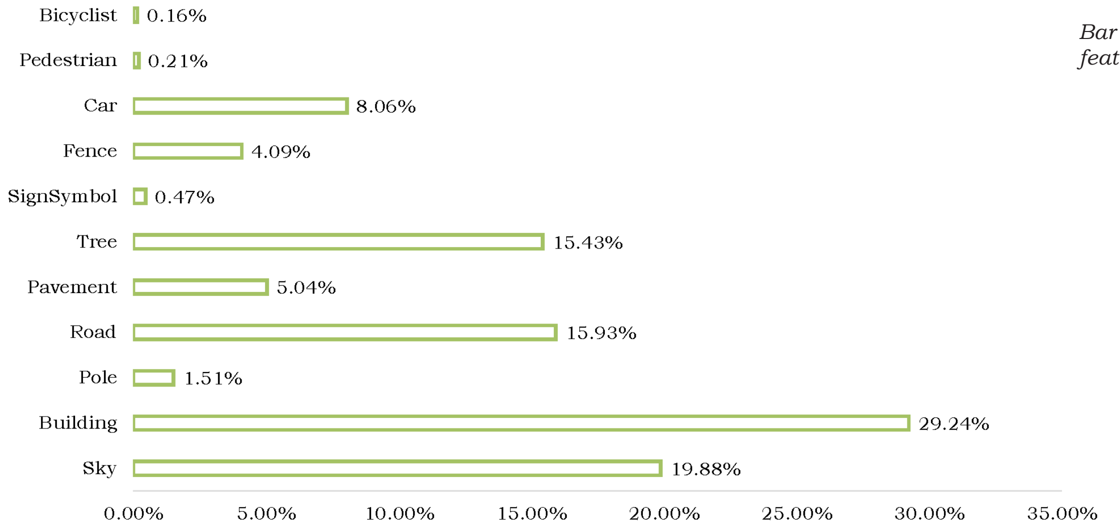
Number of image

Collecting:	*892*
Processing:	*812*



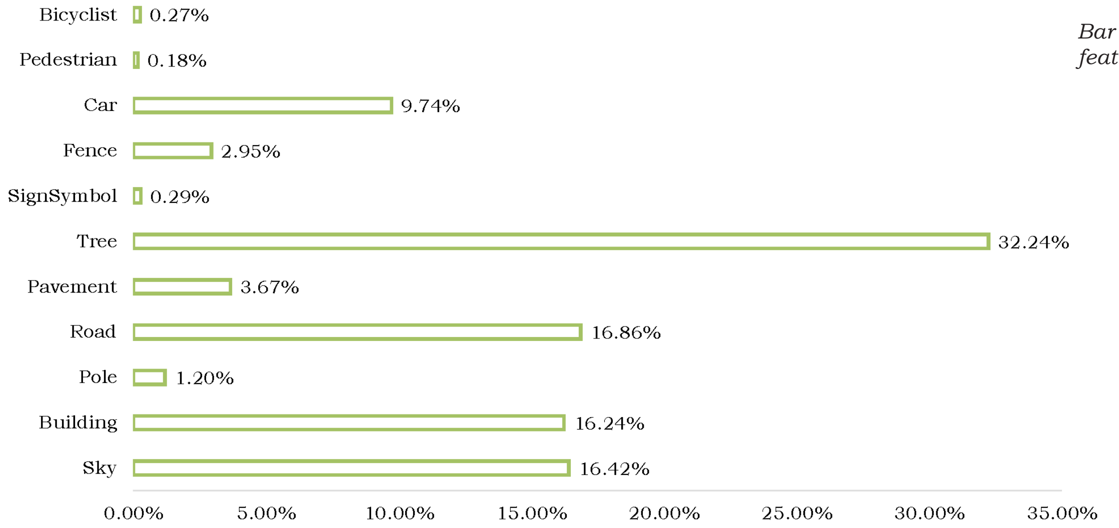
Bar chart 5.3.L1 Street features distribution in low housing price

L1



Bar chart 5.3.L2 Street features distribution in low housing price

L2



Bar chart 5.3.L3 Street features distribution in low housing price

L3

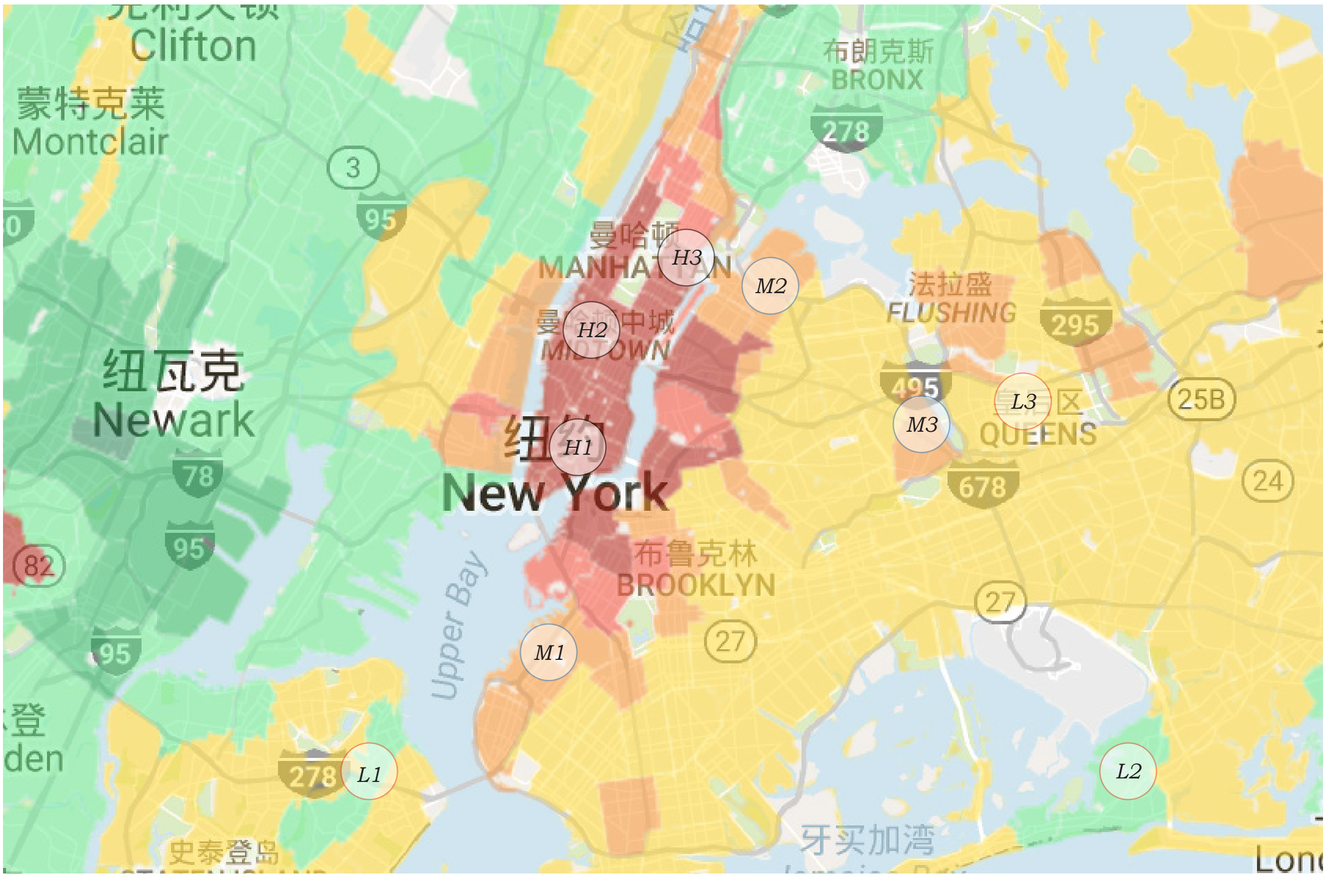


Figure 5.12 Heatmap of NewYork's property values

Souce: https://www.trulia.com/home_prices/New_York/New_York_County-heat_map/

○ Region of high housing price

○ Region of middle housing price

○ Region of low housing price

Median Sale Price per Sqft

< $50/sqft 🟩🟩🟨🟧🟥 $800+/sqft

5.3.2 New York Housing Price Distribution and UrSA

From Bar chart 5.4, we can see that H1 and H2 show the visual enclosure of high-priced regions in a high"building" ratio. However, H3 began to show a more balanced VCI ratio and the value of "tree" exceeded that of "building" and we saw it for the first time.

Relative to the consistency of the London housing price and street environment, the mid-price regions of New York shows three different situations: there are high “building” and “sky” value in M1, M2 has the similar “tree” and “building” values and the value of “tree” in M3 is particularly prominent. However, they all show the characteristics of low VCI.

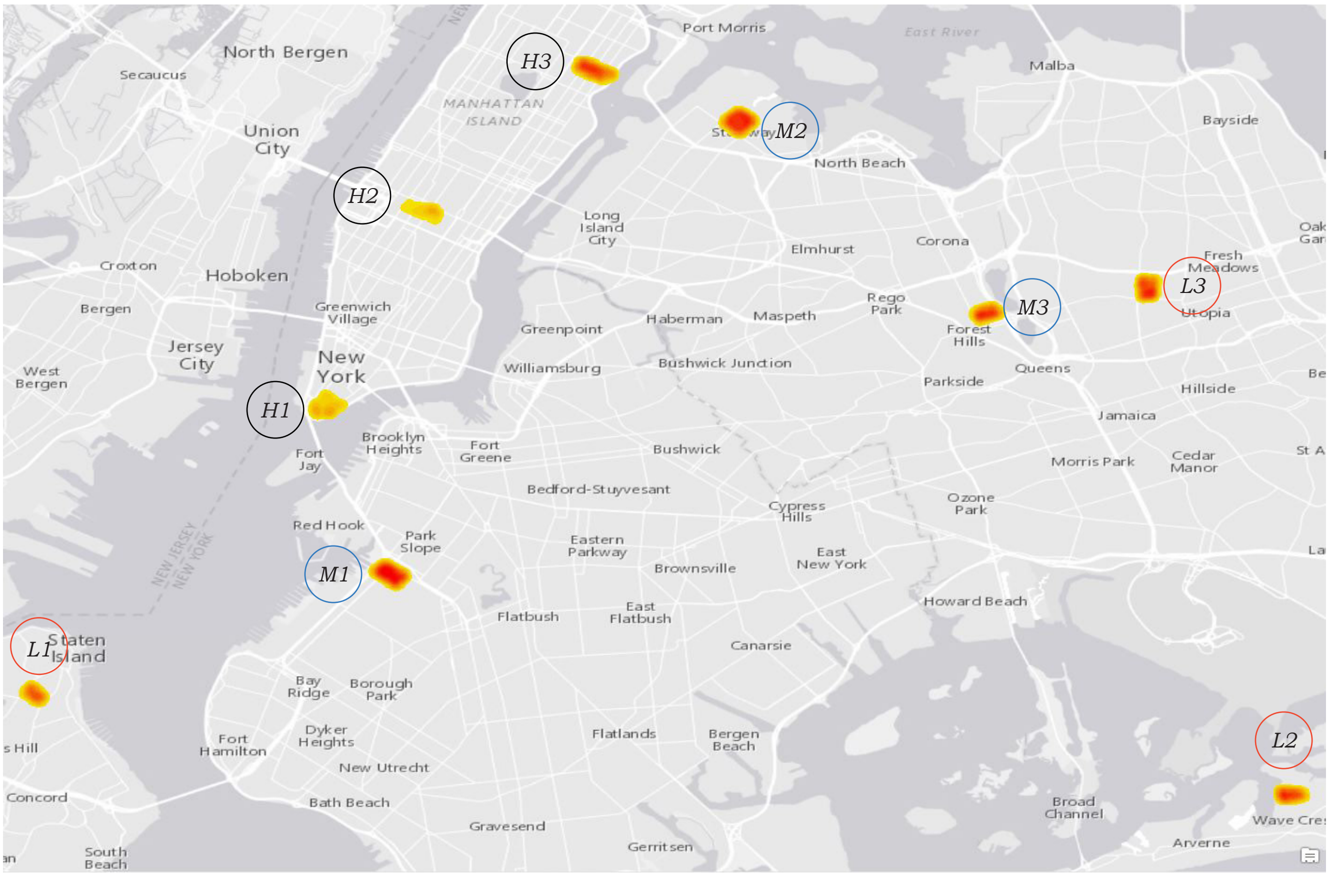


Figure 5.13 Heatmap of NewYork's VCI values

○ Region of high housing price

○ Region of middle housing price

○ Region of low housing price

High Medium Low

In New York's low-cost regions, we can see a significant increase in the "tree" value. "Tree" dominates L1, L2, and L3. In this case, they get a more intermediate VCI value, which means that the overall feeling of people in these regions is more comfortable, not overly visual enclosure and not excessively visual open.

Figure 5.18 shows the street features distribution in different housing price region of New York, it is an overall analysis of UrSA.

● ○ ○

UrSA in high housing price region

selected region H1

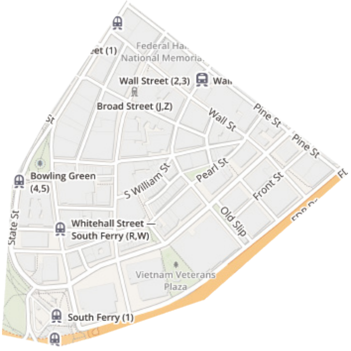


Figure 5.14.H1 Site of high housing price

Coordinate

40.708544, -74.011387
40.704620, -74.004979
40.701923, -74.009343
40.701240, -74.013477
40.703203, -74.014716
40.704887, -74.014287

Area

367,253.36 m^2

Number of image

Collecting: 2540
Processing: 1680

selected region H2

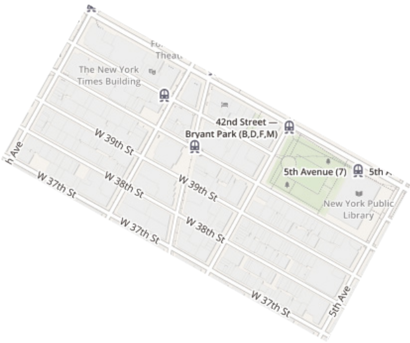


Figure 5.14.H2 Site of high housing price

Coordinate

40.757444, -73.989844
40.753564, -73.980616
40.750196, -73.983074
40.754125, -73.992299

Area

346,784.93 m^2

Number of image

Collecting: 2584
Processing: 1508

selected region H3

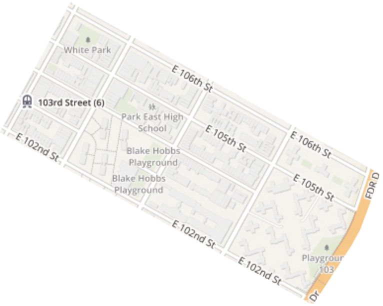


Figure 5.14.H3 Site of high housing price

Coordinate

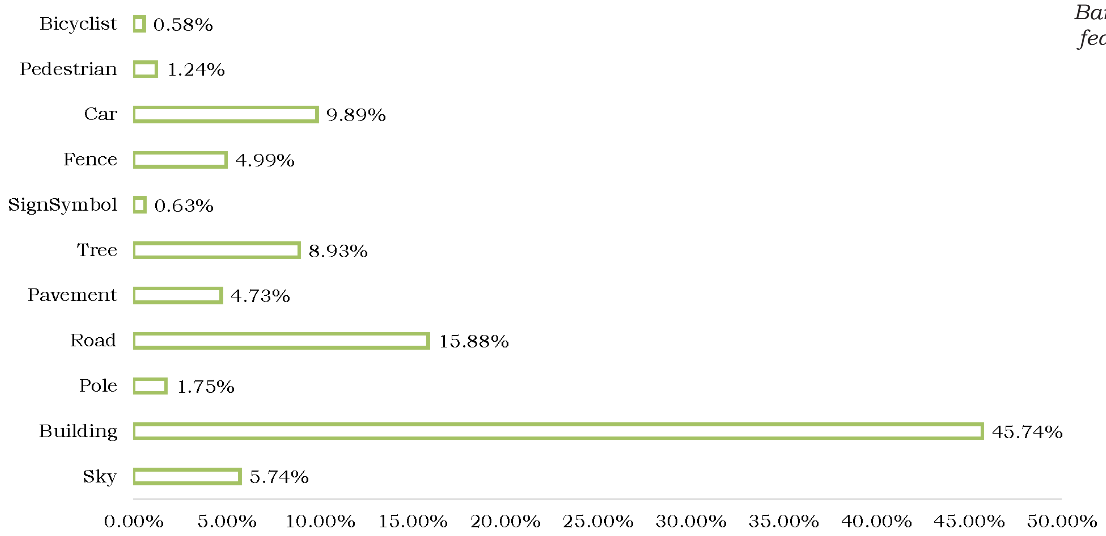
40.792524, -73.946389
40.788672, -73.937214
40.785710, -73.938999
40.789712, -73.948452

Area

302,459.23 m^2

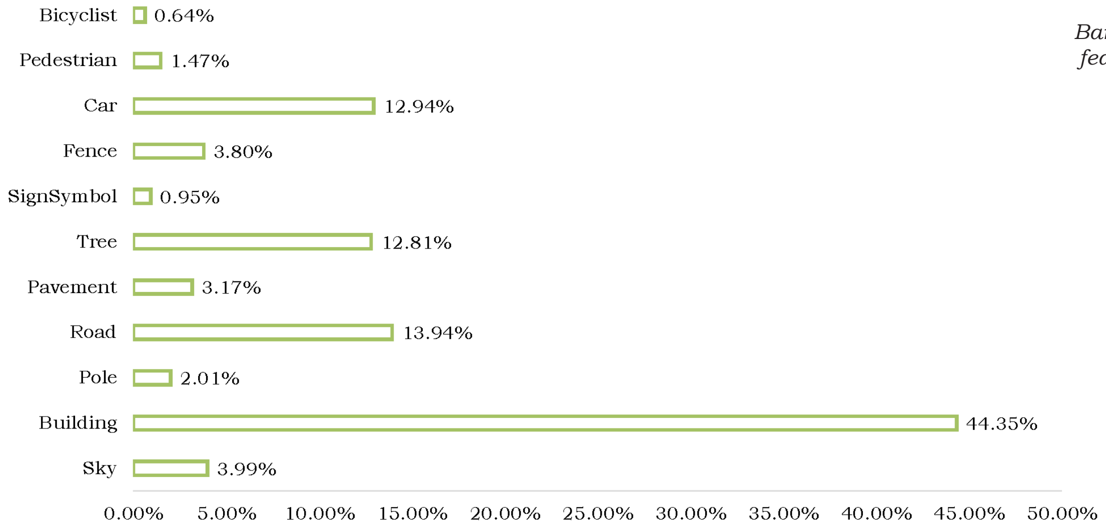
Number of image

Collecting: 2204
Processing: 1824



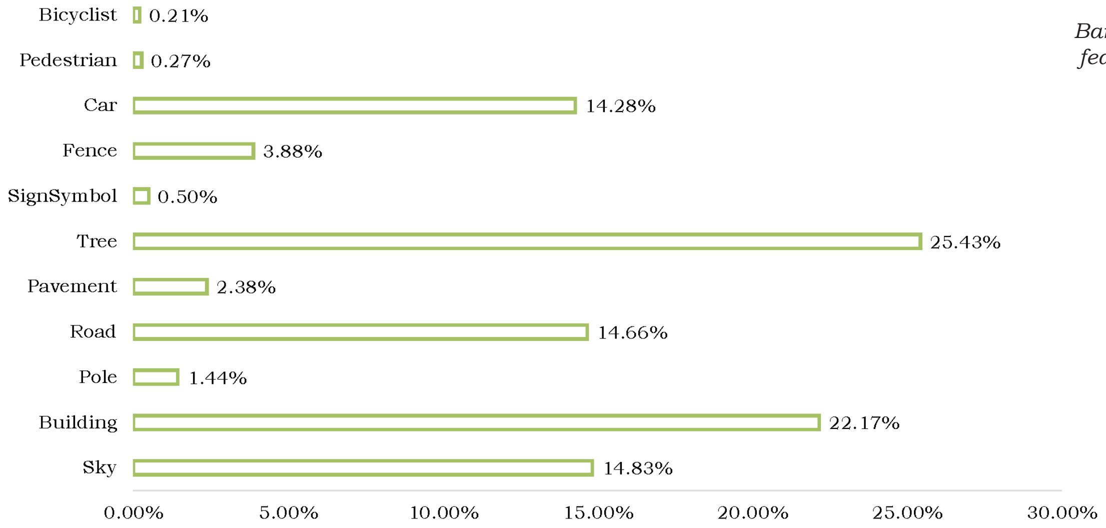
Bar chart 5.4.H1 Street features distribution in high housing price

H1



Bar chart 5.4.H2 Street features distribution in high housing price

H2



Bar chart 5.4.H3 Street features distribution in high housing price

H3

UrSA in middle housing price region

selected region M1

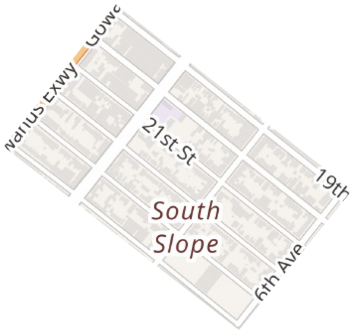


Figure 5.15.M1 Site of middle housing price

Coordinate

40.665211,	*-73.996466*
40.661190,	*-73.989830*
40.658246,	*-73.993018*
40.662162,	*-73.999591*

Area

277,979.42 m²

Number of image

Collecting: 2068
Processing: 1948

selected region M2

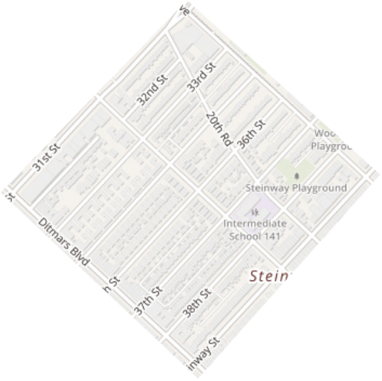


Figure 5.15.M2 Site of middle housing price

Coordinate

40.779982,	*-73.906379*
40.776526,	*-73.901422*
40.772663,	*-73.906061*
40.776123,	*-73.910915*

Area

294,192.67 m²

Number of image

Collecting: 2232
Processing: 1880

selected region M3

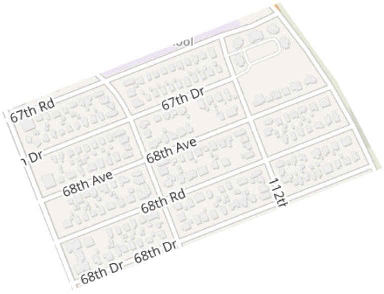


Figure 5.15.M3 Site of middle housing price

Coordinate

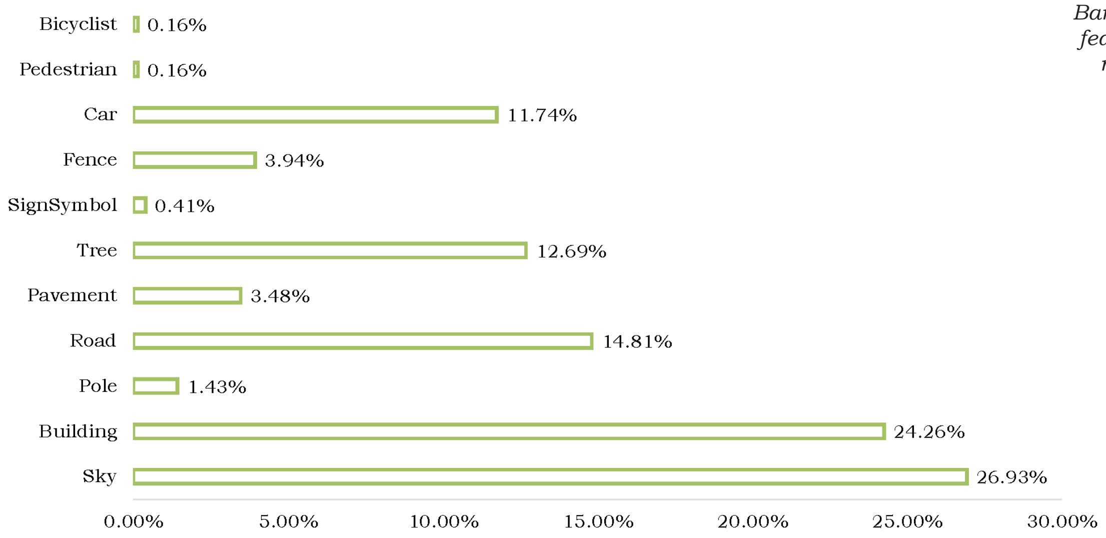
40.728363,	*-73.847404*
40.730110,	*-73.841434*
40.728842,	*-73.840253*
40.727527,	*-73.839359*
40.725510,	*-73.846005*

Area

179,784.01 m²

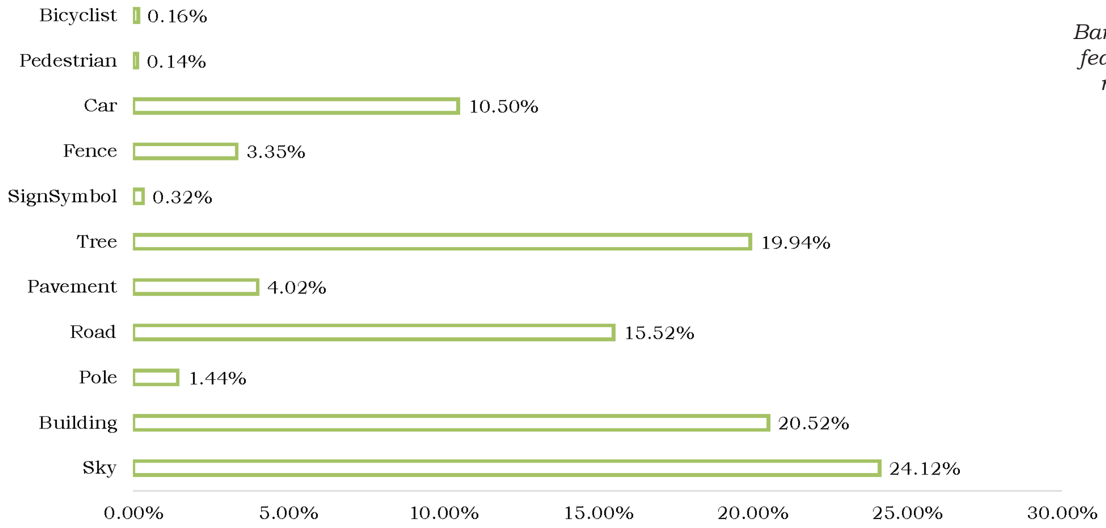
Number of image

Collecting: 1312
Processing: 1256



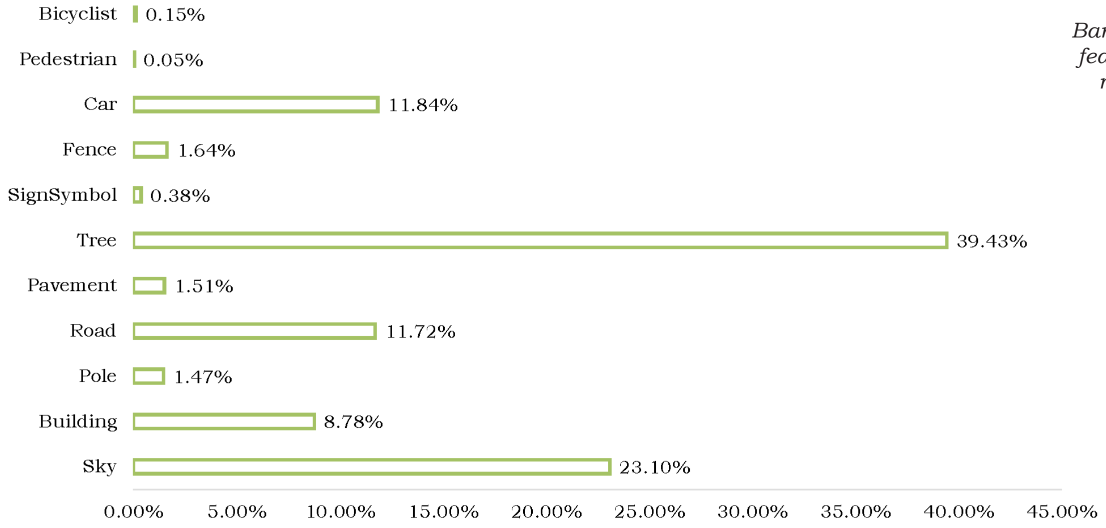
Bar chart 5.5.M1 Street features distribution in middle housing price

M1



Bar chart 5.5.M2 Street features distribution in middle housing price

M2



Bar chart 5.5.M3 Street features distribution in middle housing price

M3

○ ○ ● UrSA in low housing price region

selected region L1

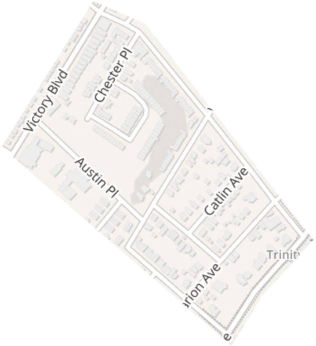


Figure 5.16.L1 Site of low housing price

Coordinate

40.634176, -74.085823
40.633749, -74.084753
40.630275, -74.081219
40.628398, -74.083041
40.631109, -74.088391

Area

184,923.83 m²

Number of image

Collecting: 1344
Processing: 1076

---

selected region L2

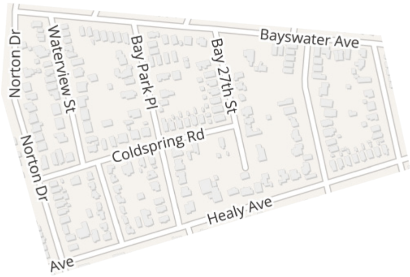


Figure 5.16.L2 Site of low housing price

Coordinate

40.607250, -73.769616
40.606711, -73.762607
40.604895, -73.762092
40.603390, -73.768676
40.605904, -73.769658

Area

188,944.21 m²

Number of image

Collecting: 1292
Processing: 1196

---

selected region L3

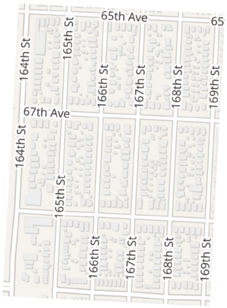


Figure 5.16.L3 Site of low housing price

Coordinate

40.736946, -73.804749
40.736706, -73.800014
40.731331, -73.800373
40.731578, -73.805104

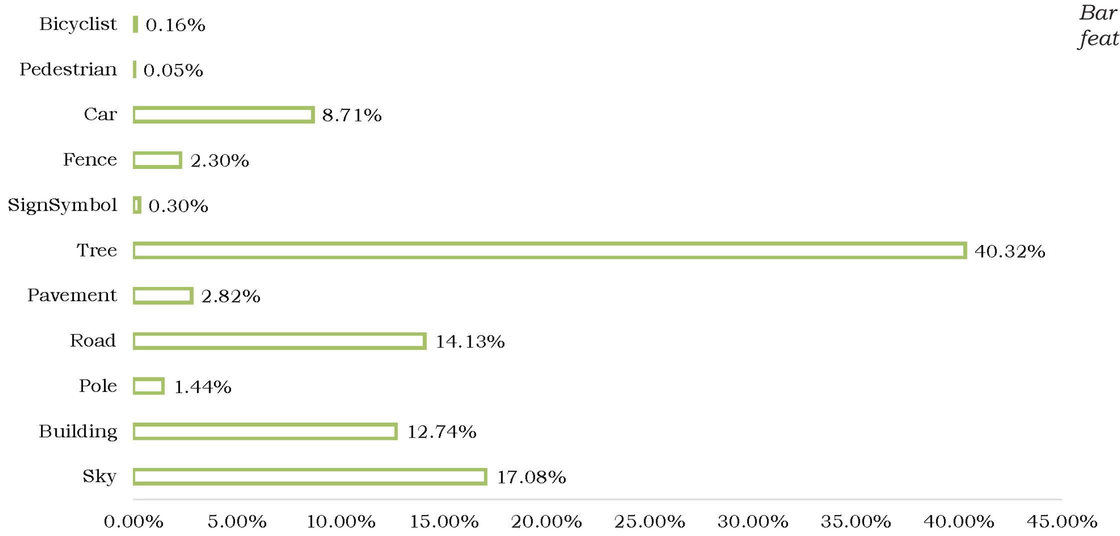
Area

219,701.59 m²

Number of image

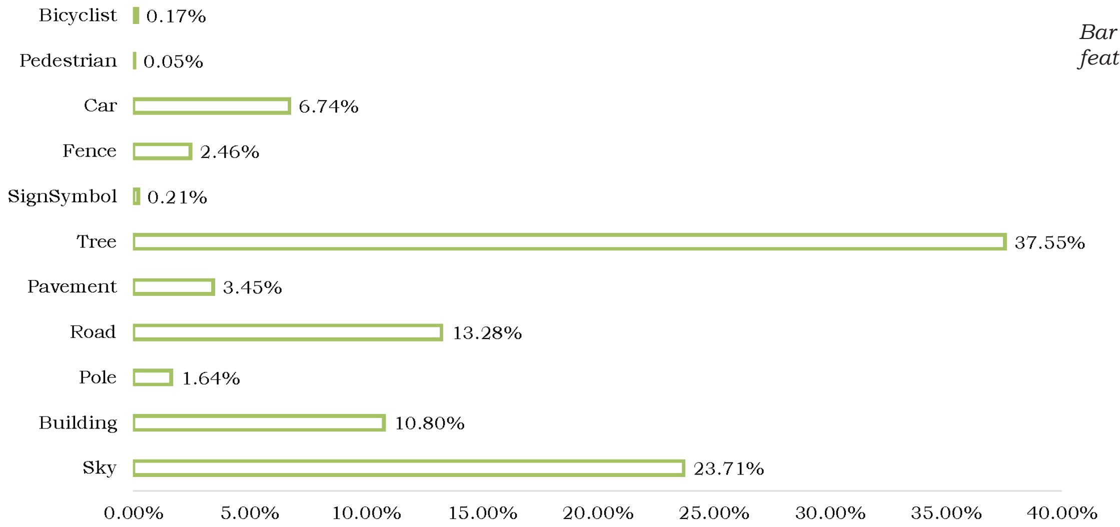
Collecting: 1628
Processing: 1540

Bar chart 5.6.L1 Street features distribution in low housing price



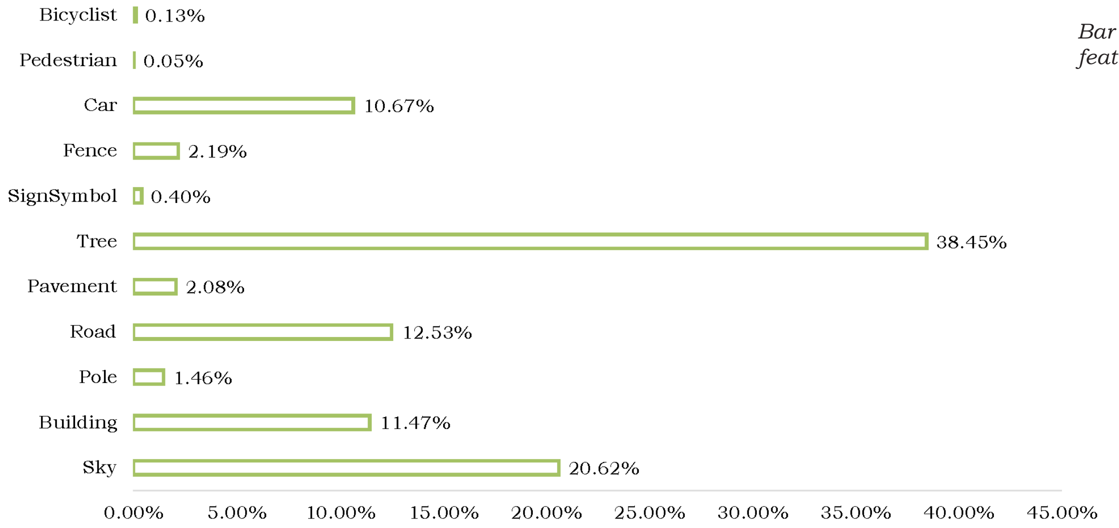
L1

Bar chart 5.6.L2 Street features distribution in low housing price



L2

Bar chart 5.6.L3 Street features distribution in low housing price



L3

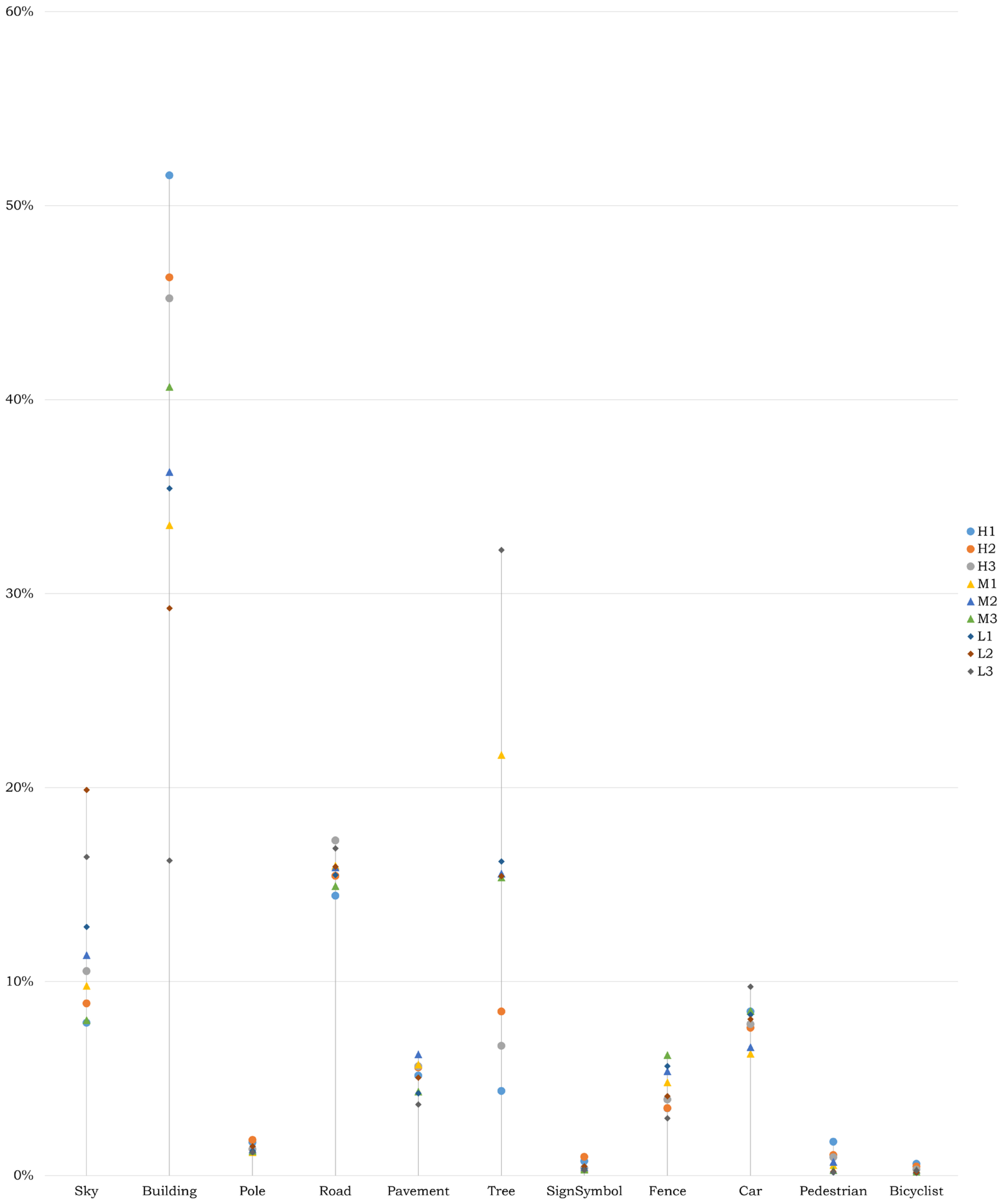


Figure 5.17 Street features distribution in different housing price region of London

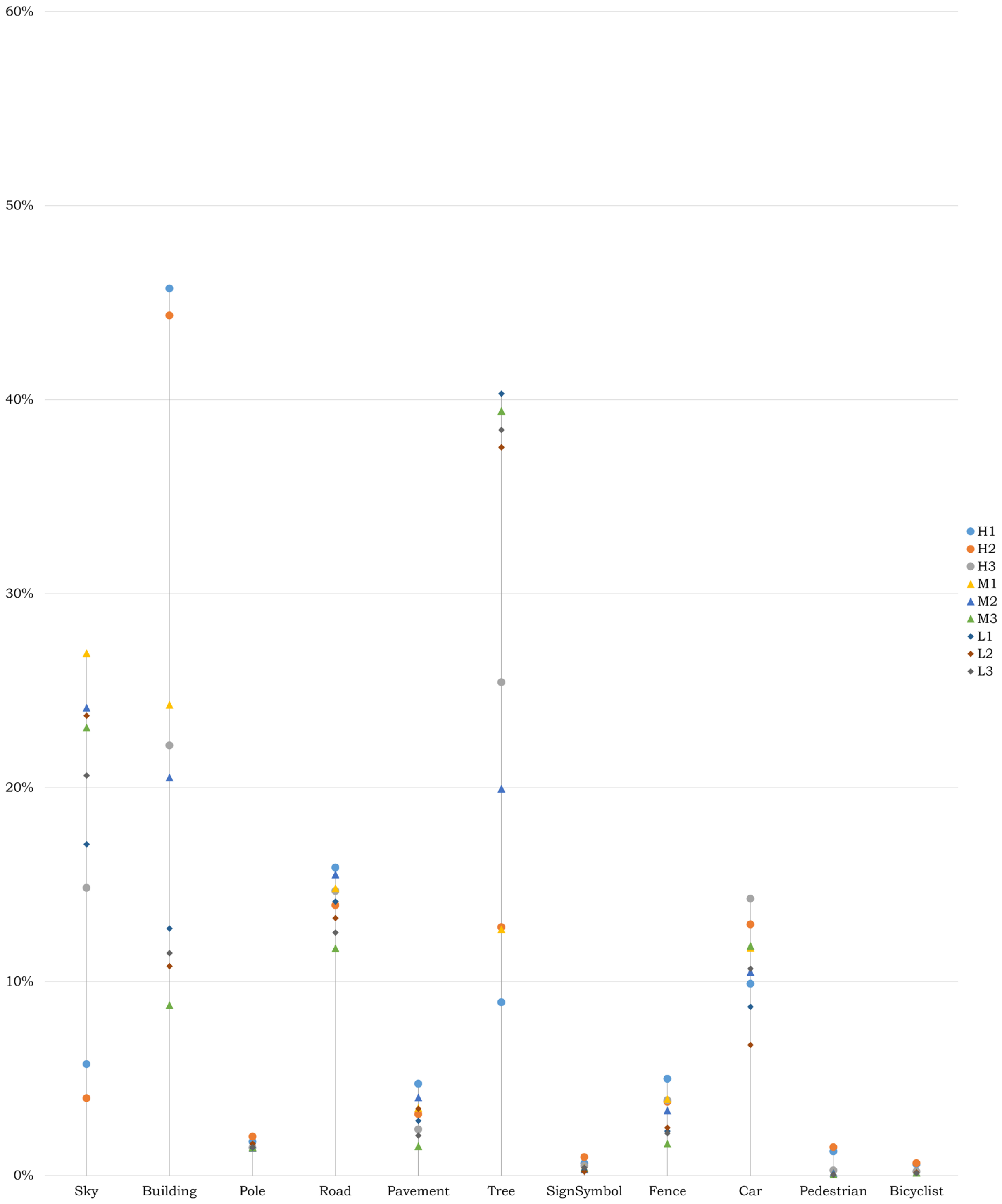


Figure 5.18 Street features distribution in different housing price region of NewYork

5.4 Application of Research 2 - Crime Rate and UrSA

5.4.1 Detroit's Crime Rate Distribution and UrSA

According to figures published on the "neighborhoodscout" website, Detroit is just safer than 4% of U.S. Cities. In Detroit, the violent crime rate is 17.57 per 1,000 residents and the property crime rate is 40.94 per 1,000 residents.

From Figure 5.20, we found that, with the exception of M1, all other regions showed low VCI values, indicating that Detroit is very open in terms of street space, which is also consistent with the scattered population reality in the United States. So we have some inaccuracies in explaining the relationship between urban street space and crime rates from the city's VCI value. Fortunately, UrSA provides more detailed data of street characterization and we can dig deeper into Detroit's street character.

Figure 5.19 Heatmap of Detroit's crime rate values

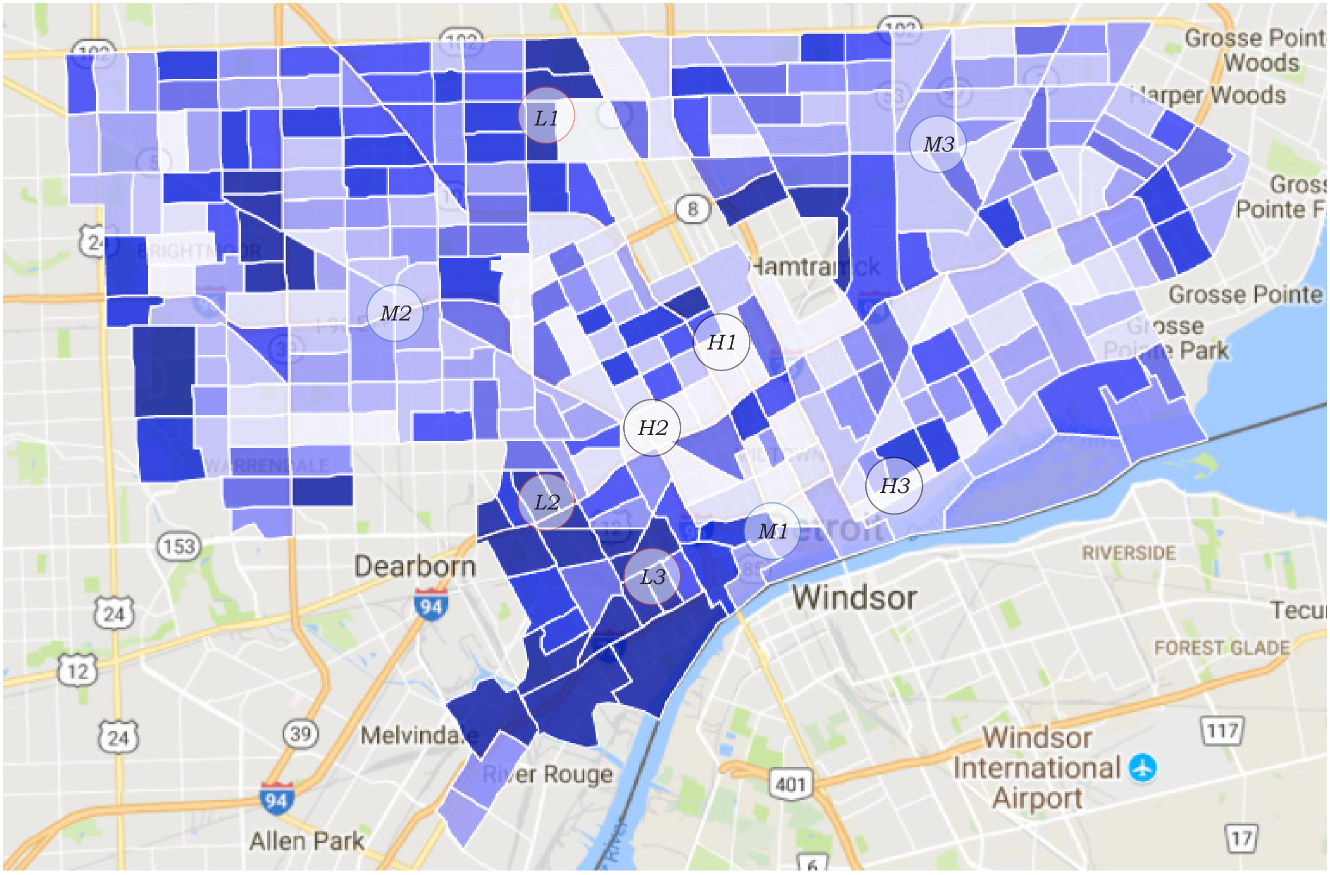
Source: https://www.neighborhoodscout.com/mi/detroit/crime

○ Region of high crime rate

○ Region of middle crime rate

○ Region of low crime rate

Safest High Crime rate



According to the results of UrSA (Bar chart 5.7-5.8), we found that Detroit is a city with a very high "tree" value and a low "building" value, regardless of the crime rate. However, the value of "sky" has a certain positive correlation with the crime rate. In regions with a high crime rate, the "sky" value is above 25% and the H3 reaches 30.36%. In relatively safe regions, there is a clear drop in the "sky" value of Detroit Street, which suggests that overly open spaces may be responsible for the rise in crime rates.

Figure 5.29 shows the street features distribution in different crime rate region of Detroit, it can help planner to do the overall analysis of UrSA's result.

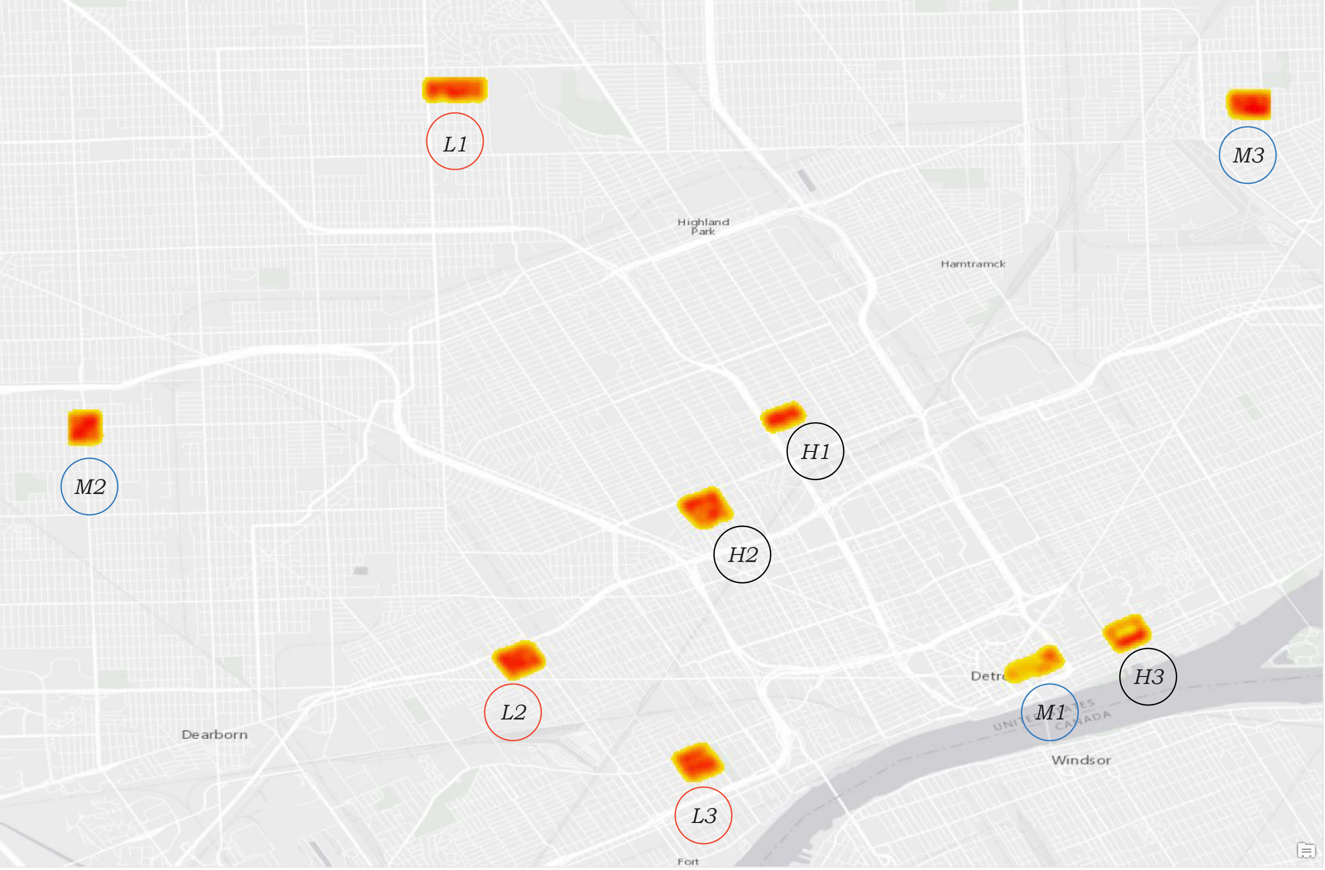
Figure 5.20 Heatmap of Detroit's VCI values

○ Region of high housing price

○ Region of middle housing price

○ Region of low housing price

High Medium Low



● ○ ○

UrSA in high crime rate region

selected region H1

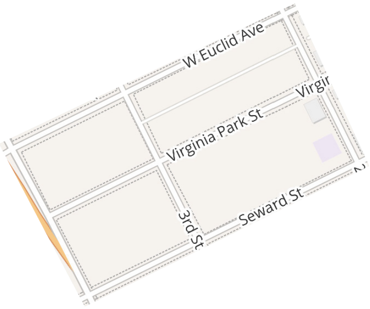


Figure 5.21.H1 Site of high crime rate

Coordinate

42.374190, -83.086172
42.376130, -83.080902
42.373976, -83.079438
42.372091, -83.084626

Area

135,622.25 m²

Number of image

Collecting: 888
Processing: 784

selected region H2

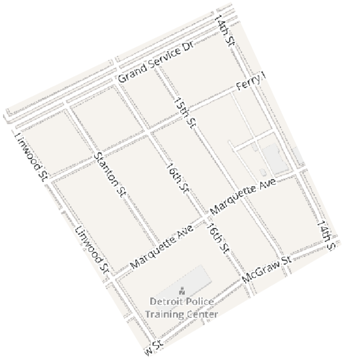


Figure 5.21.H2 Site of high crime rate

Coordinate

42.360058, -83.099945
42.362140, -83.094273
42.358043, -83.091532
42.357329, -83.093764
42.356006, -83.096213

Area

246,280.76 m²

Number of image

Collecting: 1740
Processing: 1612

selected region H3

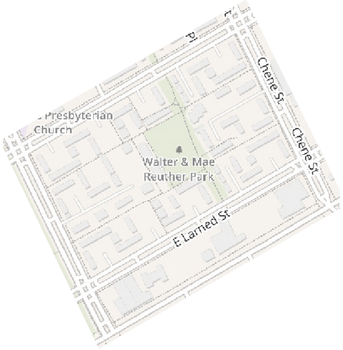


Figure 5.21.H3 Site of high crime rate

Coordinate

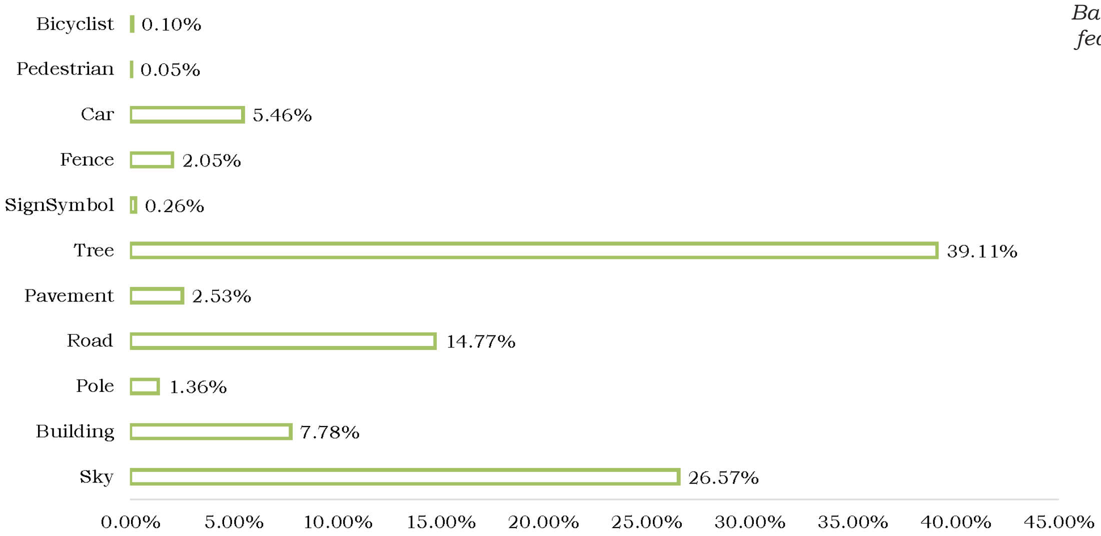
42.339152, -83.029880
42.341339, -83.024791
42.338163, -83.022675
42.335990, -83.027777

Area

197,635.59 m²

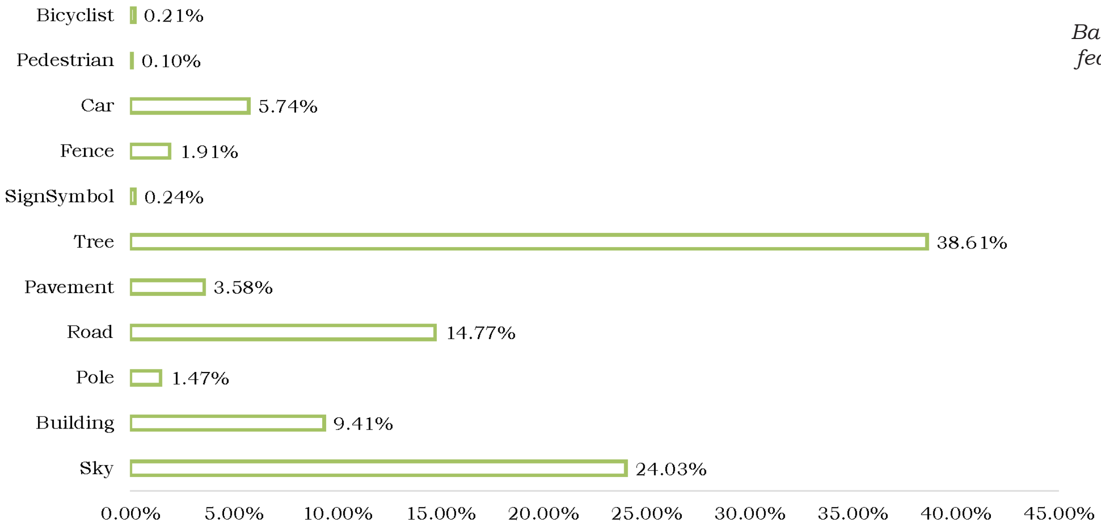
Number of image

Collecting: 1316
Processing: 868



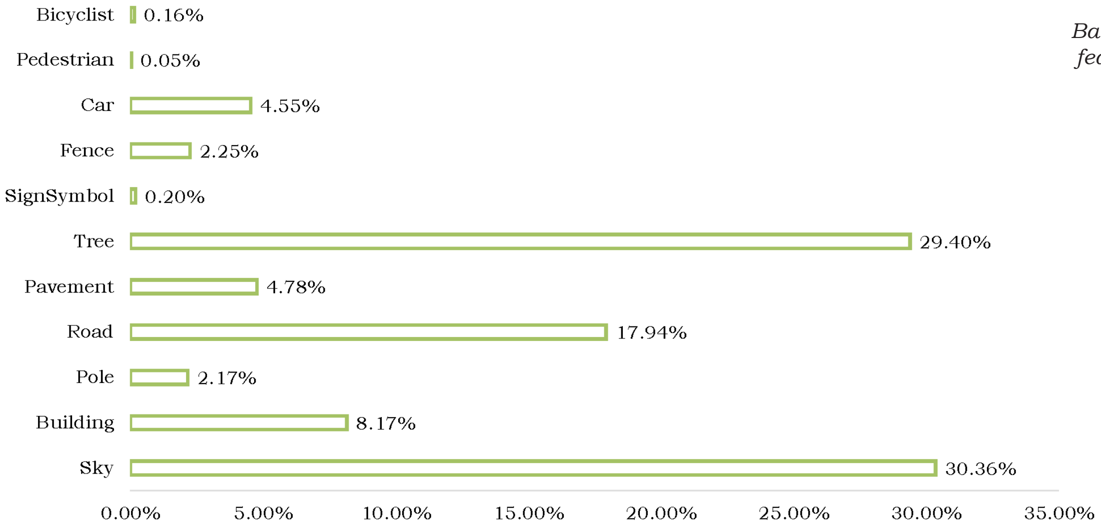
Bar chart 5.7.H1 Street features distribution in high crime rate

H1



Bar chart 5.7.H2 Street features distribution in high crime rate

H2



Bar chart 5.7.H3 Street features distribution in high crime rate

H3

○ ● ○ UrSA in middle crime rate region

selected region M1

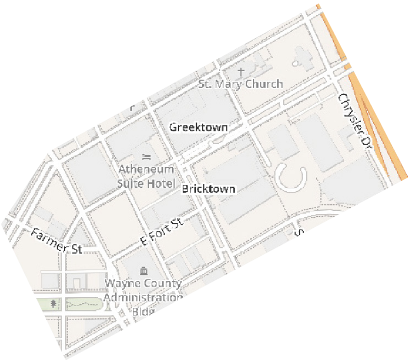


Figure 5.22.M1 Site of middle crime rate

Coordinate

42.332885, -83.046918
42.336598, -83.038939
42.333809, -83.037120
42.330300, -83.045510

Area

207,804.66 m²

Number of image

Collecting: 1776
Processing: 1460

selected region M2

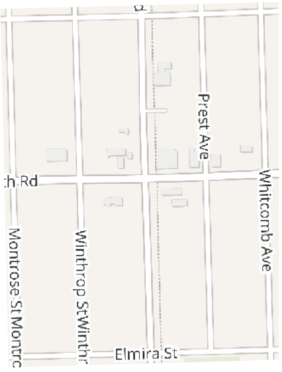


Figure 5.22.M2 Site of middle crime rate

Coordinate

42.374793, -83.200174
42.374862, -83.195363
42.370177, -83.195079
42.370116, -83.199936

Area

201,371.58 m²

Number of image

Collecting: 684
Processing: 684

selected region M3

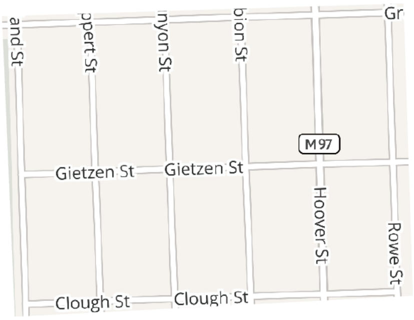


Figure 5.22.M3 Site of middle crime rate

Coordinate

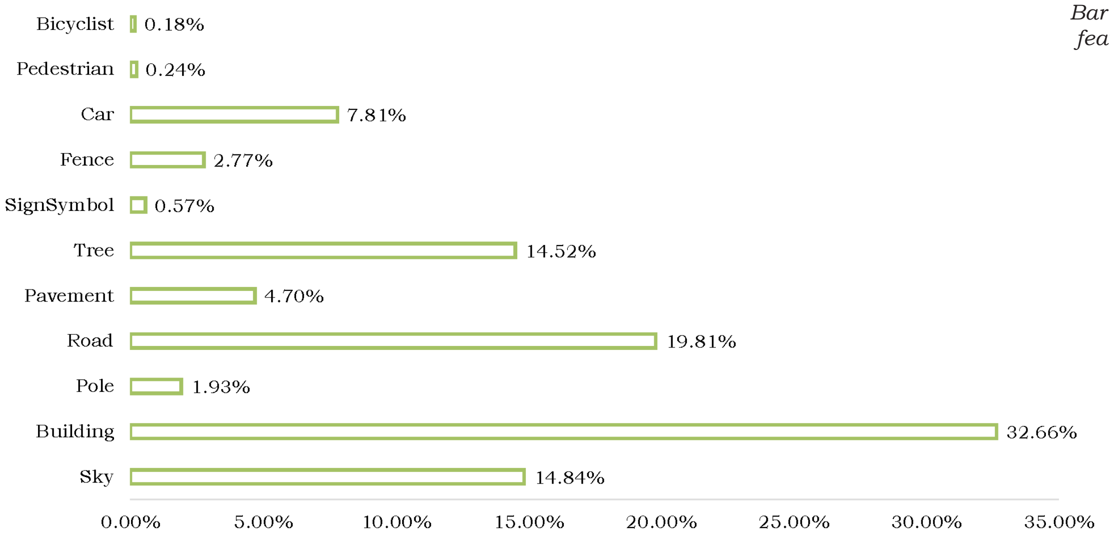
42.427176, -83.009353
42.427328, -83.003064
42.423818, -83.002918
42.423669, -83.009152

Area

203,129.79 m²

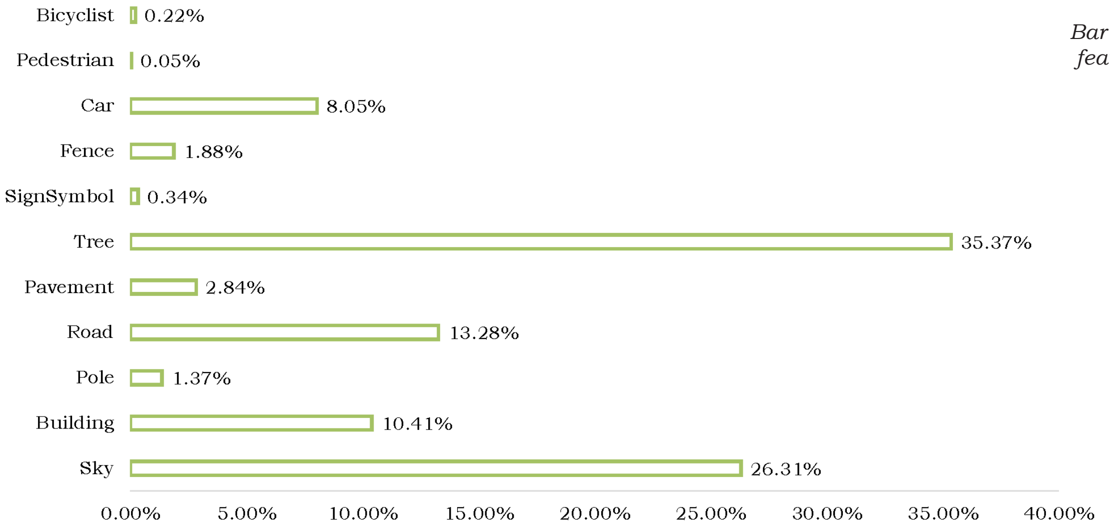
Number of image

Collecting: 1400
Processing: 1384



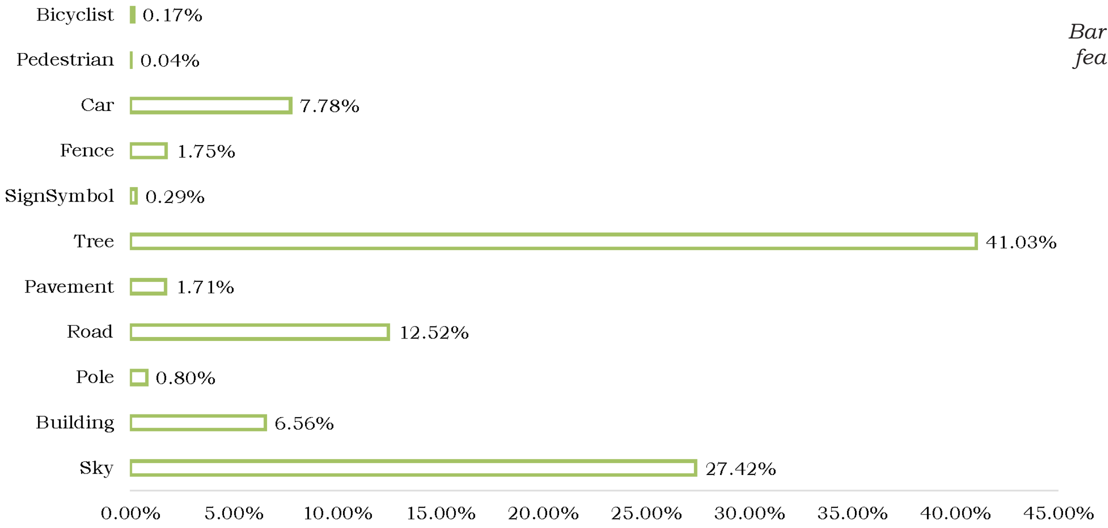
Bar chart 5.8.M1 Street features distribution in middle crime rate

M1



Bar chart 5.8.M2 Street features distribution in middle crime rate

M2



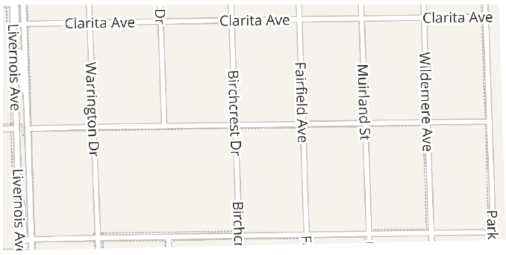
Bar chart 5.8.M3 Street features distribution in middle crime rate

M3

○ ○ ●

UrSA in low crime rate region

selected region L1



Coordinate

42.429597, -83.141425
42.429691, -83.132162
42.426469, -83.132016
42.426373, -83.141281

Area

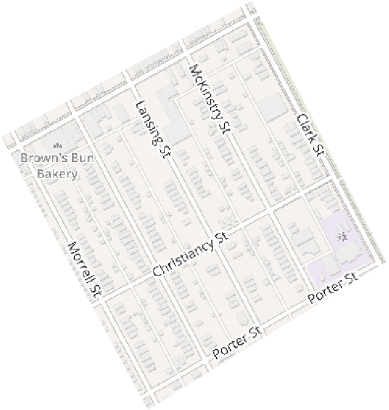
266,631.62 m²

Number of image

Collecting: 1884
Processing: 1808

Figure 5.23.L1 Site of low crime rate

selected region L2



Coordinate

42.318285, -83.100747
42.320288, -83.095783
42.316475, -83.093081
42.314452, -83.097929

Area

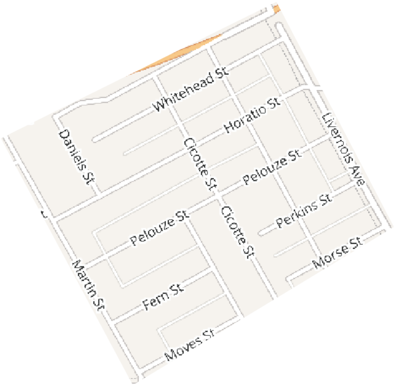
226,532.71 m²

Number of image

Collecting: 1556
Processing: 1460

Figure 5.23.L2 Site of low crime rate

selected region L3



Coordinate

42.335019, -83.130194
42.337002, -83.124583
42.333652, -83.122141
42.331346, -83.127447

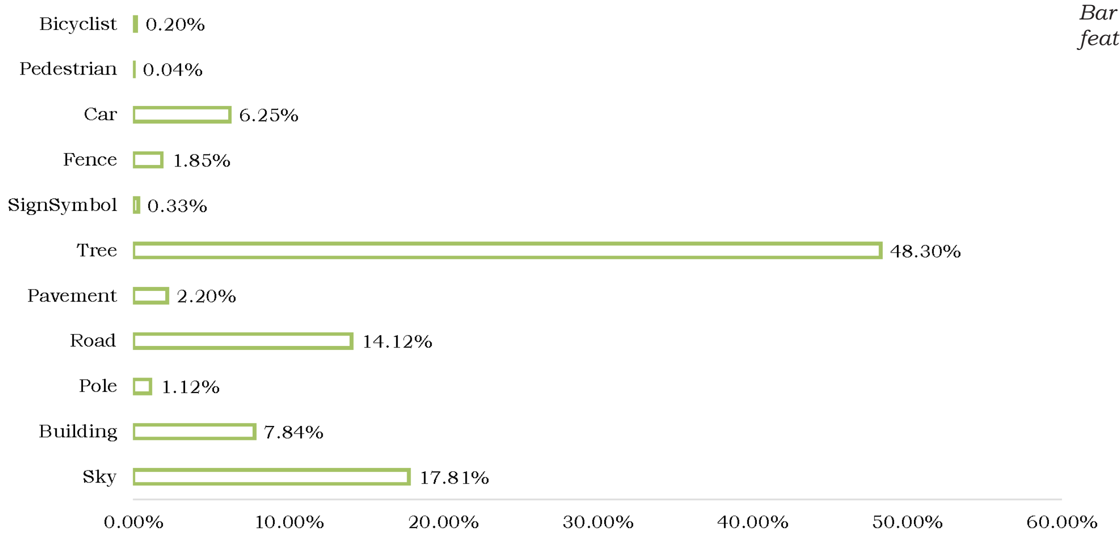
Area

218,468.57 m²

Number of image

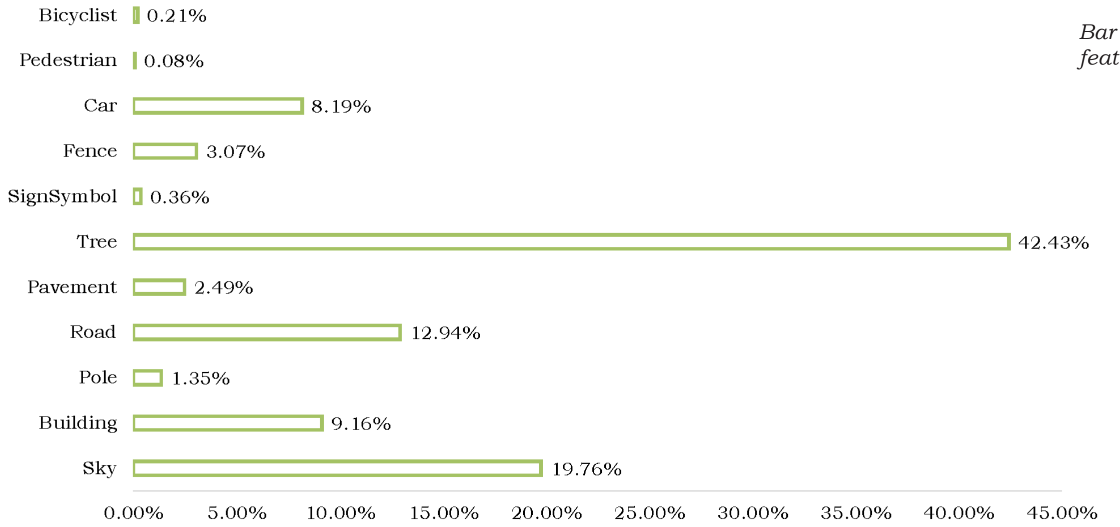
Collecting: 1580
Processing: 1572

Figure 5.23.L3 Site of low crime rate



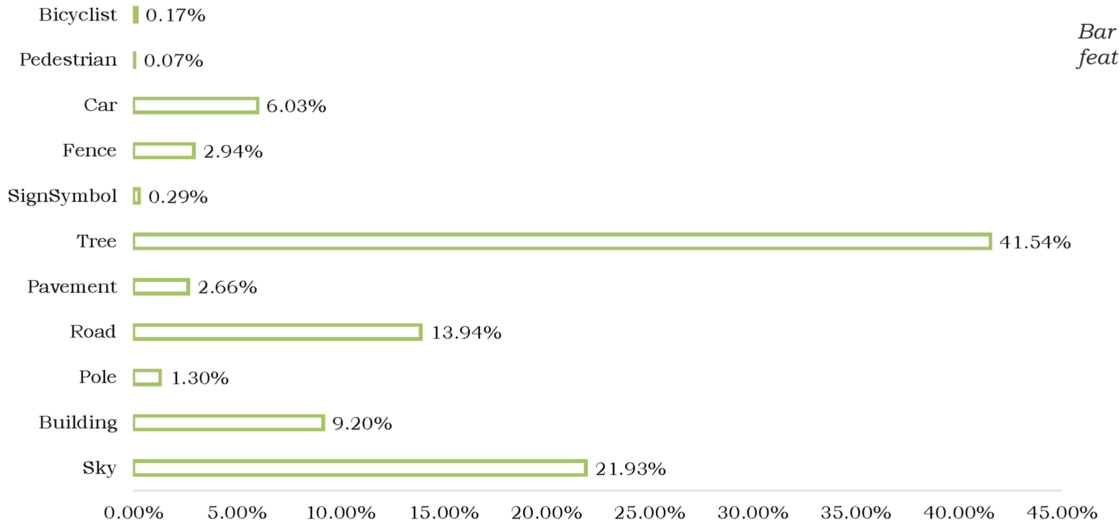
Bar chart 5.9.L1 Street features distribution in low crime rate

L1



Bar chart 5.9.L2 Street features distribution in low crime rate

L2



Bar chart 5.9.L3 Street features distribution in low crime rate

L3

5.4.2 Johannesburg's Crime Rate Distribution and UrSA

As we all know, the crime rate in South Africa is very high, and Johannesburg is one of the highest. According to the rankings of Crime Index 2017 Mid-Year provided by number, Johannesburg ranked sixth with a high crime index of 78.63%.

The Figure 5.25 shows that the VCI in the center of Johannesburg is relatively high. Crimes often occur in the region with lower VCI values, such as H1 and H3, but there is a clear rise in the value of VCI in areas with low crime rates,

Through an in-depth analysis of UrSA results, we found that the higher the "building" value, the safer the community, meaning that a high building visibility in the street can suppress the crime in a certain extent. Crime easily occurs in the open area of the street.

From Bar charts 5.10 and 5.11, it is clear that the higher "tree" and "sky" values for high-crime areas are the main reasons for

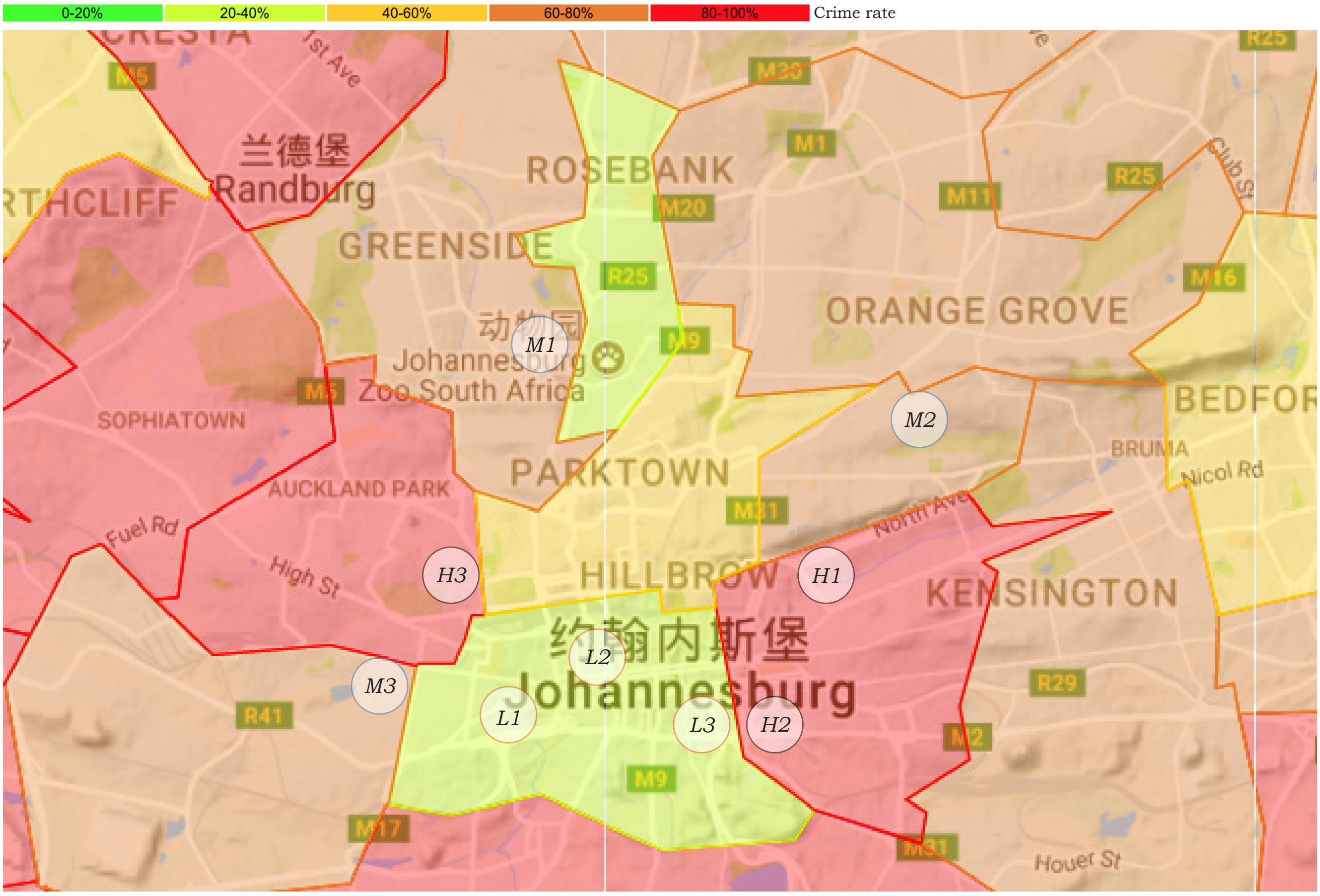
Figure 5.24 Heatmap of Johannesburg's crime rate values

Source: http://www.crimestatssa.com

○ Region of high crime rate

○ Region of middle crime rate

○ Region of low crime rate



the lower VCI value, implying an imbalance in street attributes. This imbalance makes people unwilling to appear in the streets, indirectly resulting in a very low "pedestrian" value of UrSA showed in the Bar chart.

In the low crime area, we can see that although the proportion of "building" is much higher than other parameters, no more than 40% in each area. Contrasts with New York and London's high housing price region, it does not result in a too high VCI value, that is to say, it has not formed a strong visual enclosure in these low crime rate region.

Figure 5.30 shows the street features distribution in different crime rate region of Johannesburg, it can help planner to do the overall analysis of UrSA's result.

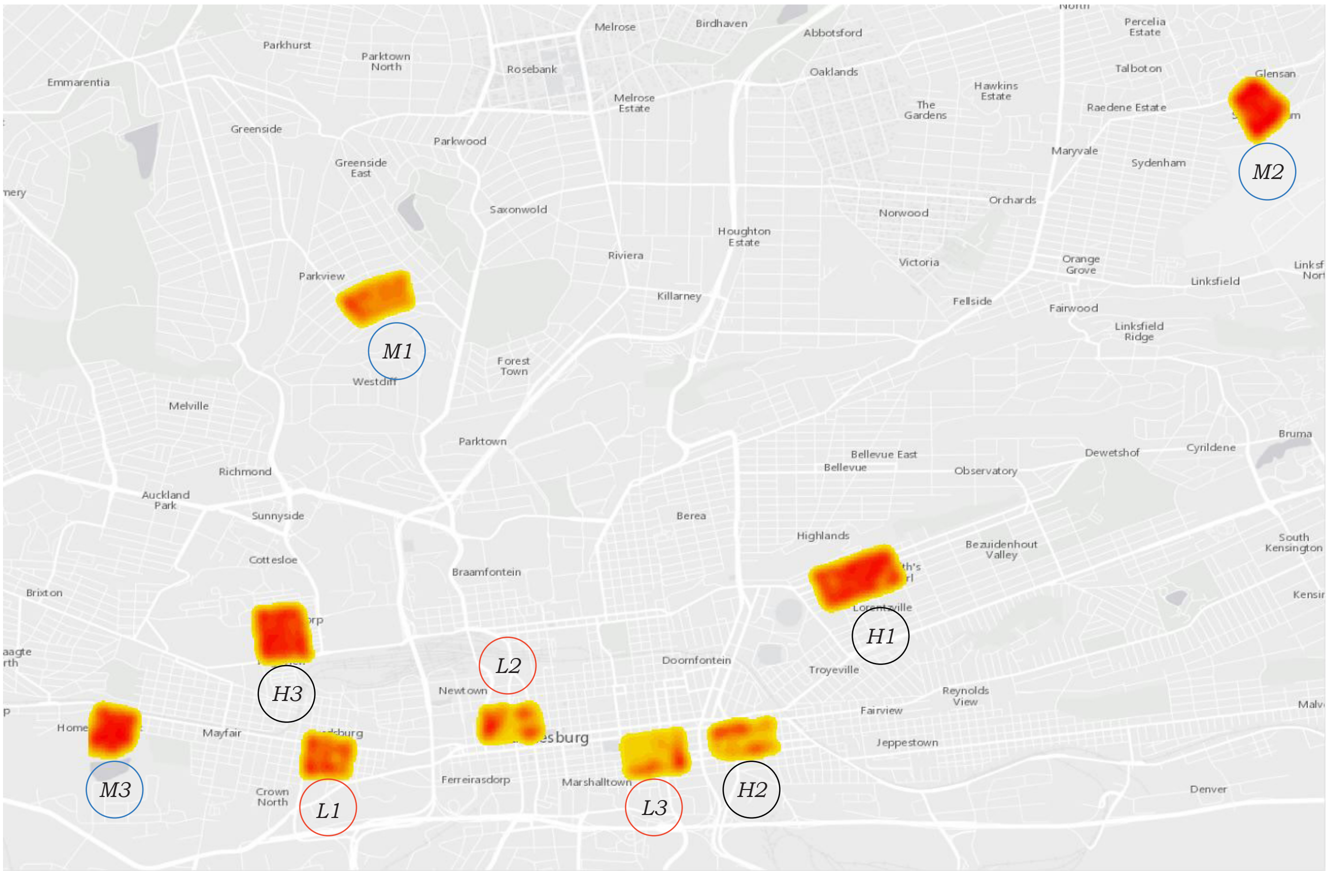
Figure 5.25 Heatmap of Johannesburg's VCI values

○ Region of high housing price

○ Region of middle housing price

○ Region of low housing price

High Medium Low



● ○ ○

UrSA in high crime rate region

selected region H1

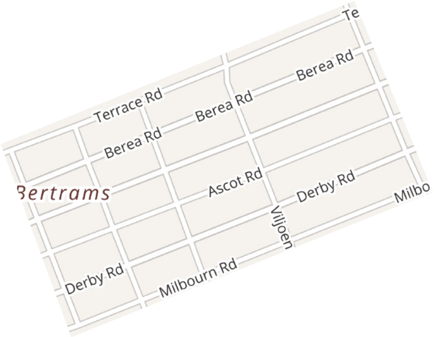


Figure 5.26.H1 Site of high crime rate

Coordinate

-26.190324, 28.064416
-26.187924, 28.071214
-26.191009, 28.072558
-26.193410, 28.065835

Area

252,744.47 m²

Number of image

Collecting: 1548
Processing: 1548

selected region H2

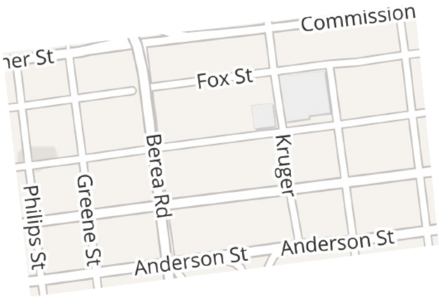


Figure 5.26.H2 Site of high crime rate

Coordinate

-26.203905, 28.055411
-26.203265, 28.060877
-26.206064, 28.061291
-26.206710, 28.055831

Area

163,357.60 m²

Number of image

Collecting: 992
Processing: 992

selected region H3

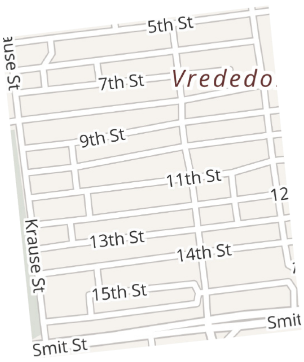


Figure 5.26.H3 Site of high crime rate

Coordinate

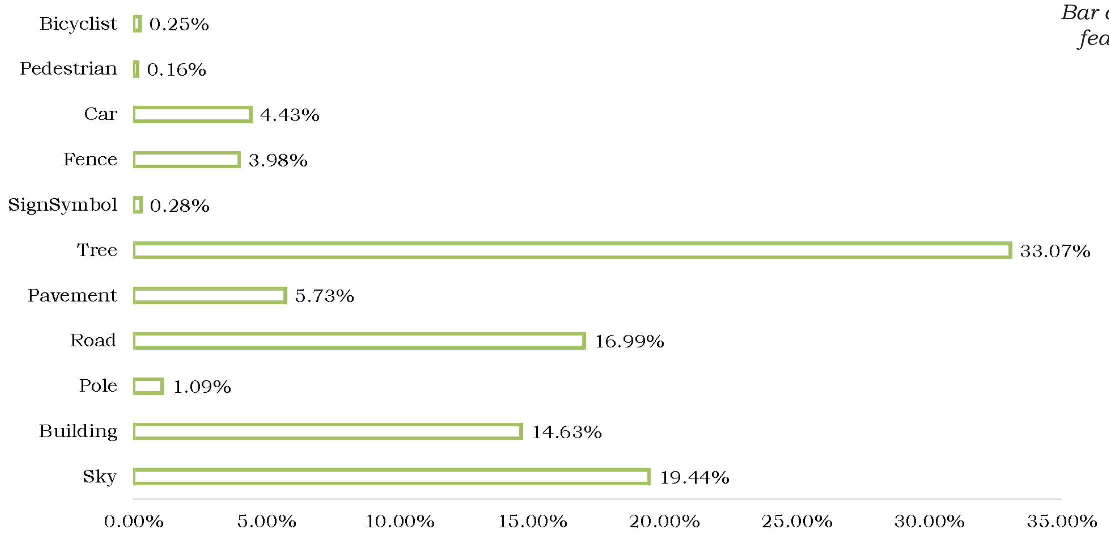
-26.193588, 28.014849
-26.193145, 28.018974
-26.197634, 28.019657
-26.198132, 28.015462

Area

203,818.92 m²

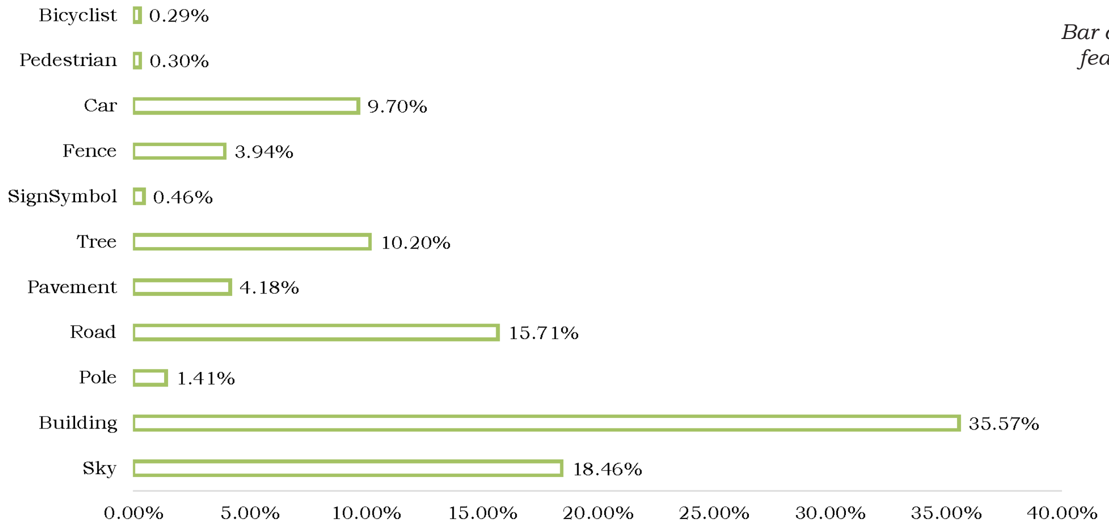
Number of image

Collecting: 1216
Processing: 1216



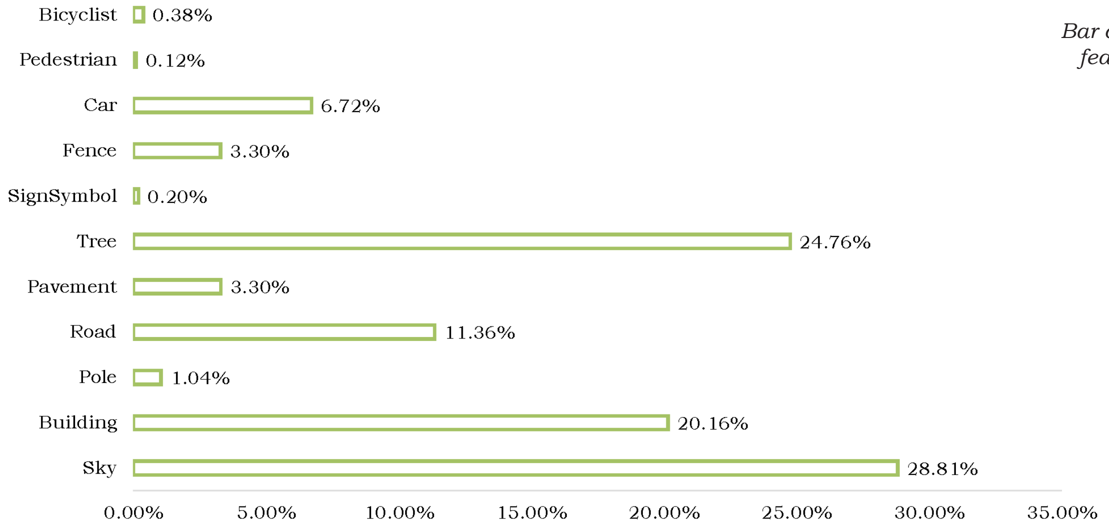
Bar chart 5.10.H1 Street features distribution in high crime rate

H1



Bar chart 5.10.H2 Street features distribution in high crime rate

H2



Bar chart 5.10.H3 Street features distribution in high crime rate

H3

○ ● ○ UrSA in middle crime rate region

selected region M1

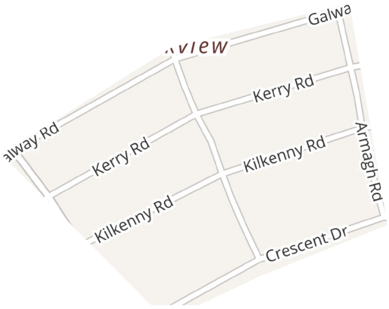


Figure 5.27.M1 Site of middle crime rate

Coordinate

-26.165735, 28.022204
-26.164253, 28.025178
-26.163502, 28.028159
-26.166399, 28.028859
-26.167038, 28.026484
-26.168154, 28.024202

Area

175,377.29 m^2

Number of image

Collecting: 1112
Processing: 1112

selected region M2

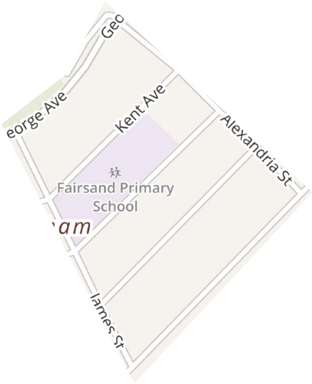


Figure 5.27.M2 Site of middle crime rate

Coordinate

-26.147879, 28.101852
-26.146060, 28.103785
-26.148918, 28.107171
-26.151804, 28.104001

Area

167,121.97 m^2

Number of image

Collecting: 964
Processing: 964

selected region M3

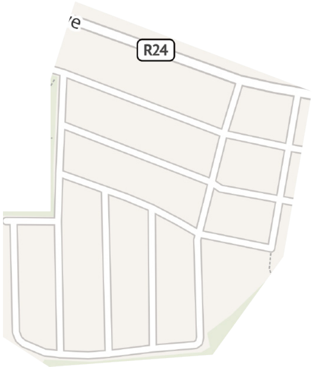


Figure 5.27.M3 Site of middle crime rate

Coordinate

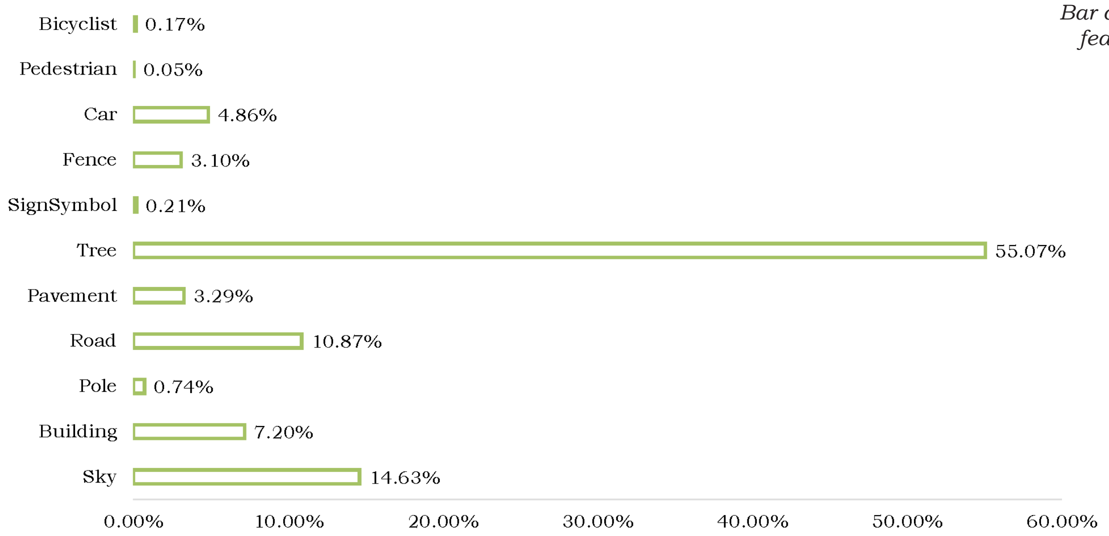
-26.201855, 28.001045
-26.202851, 28.004388
-26.205270, 28.003877
-26.206317, 28.002859
-26.206381, 27.999981
-26.204490, 27.999942

Area

151,979.30 m^2

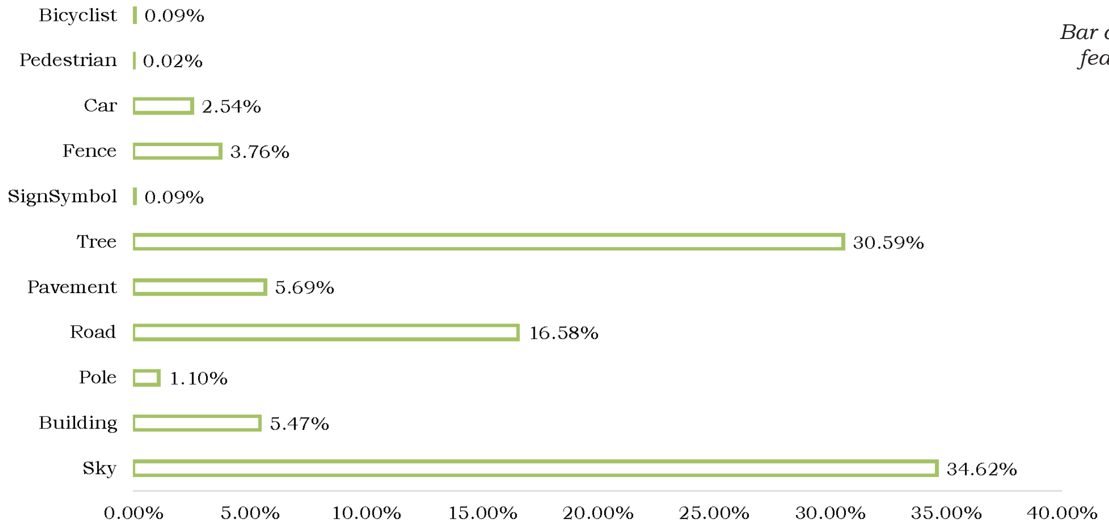
Number of image

Collecting: 936
Processing: 936



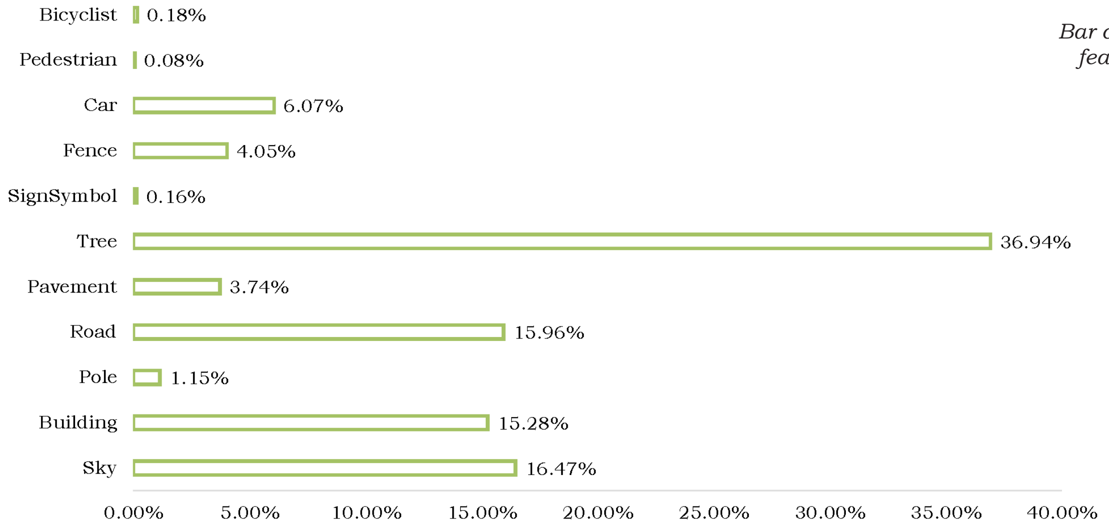
Bar chart 5.11.M1 Street features distribution in middle crime rate

M1



Bar chart 5.11.M2 Street features distribution in middle crime rate

M2



Bar chart 5.11.M3 Street features distribution in middle crime rate

M3

○ ○ ● UrSA in low crime rate region

selected region L1

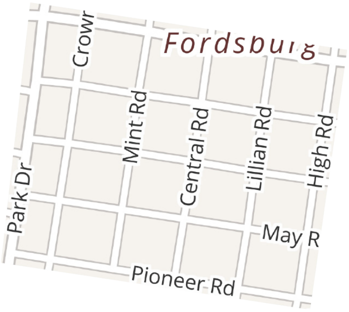


Figure 5.28.L1 Site of low crime rate

Coordinate

-26.204629, 28.019608
-26.205174, 28.023607
-26.208664, 28.023052
-26.208187, 28.019080

Area

151,827.94 m^2

Number of image

Collecting: 916
Processing: 916

selected region L2

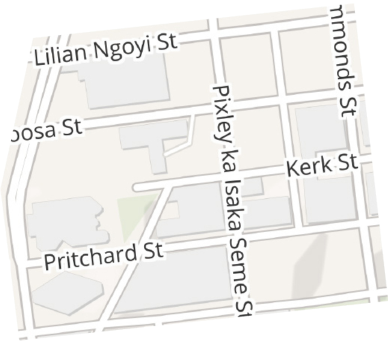


Figure 5.28.L2 Site of low crime rate

Coordinate

-26.202391, 28.035197
-26.201822, 28.040065
-26.204893, 28.040504
-26.205506, 28.035276
-26.203219, 28.034933

Area

176,177.72 m^2

Number of image

Collecting: 1028
Processing: 1028

selected region L3

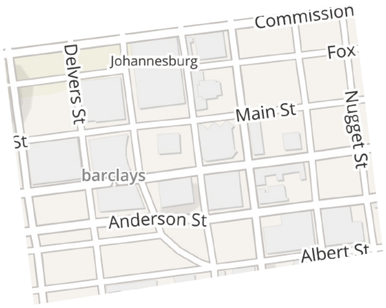


Figure 5.28.L3 Site of low crime rate

Coordinate

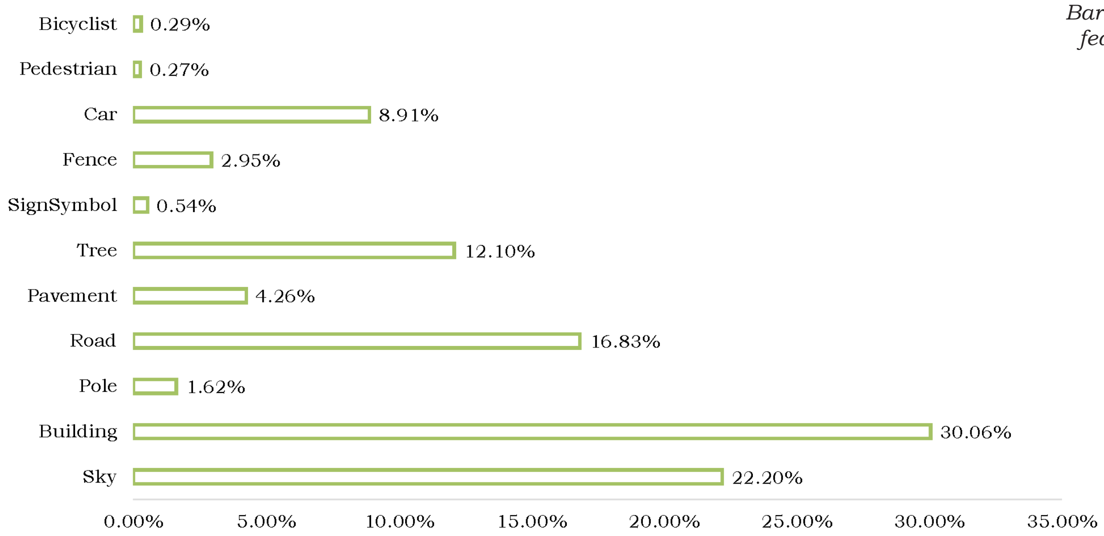
-26.204794, 28.047346
-26.204172, 28.052789
-26.207723, 28.053284
-26.208316, 28.047797

Area

207,489.77 m^2

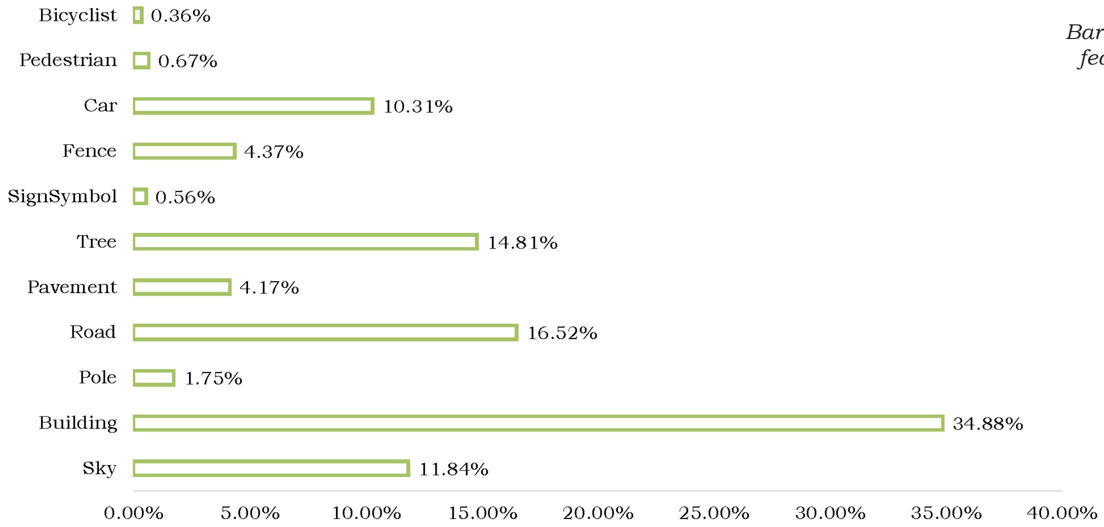
Number of image

Collecting: 1252
Processing: 1252



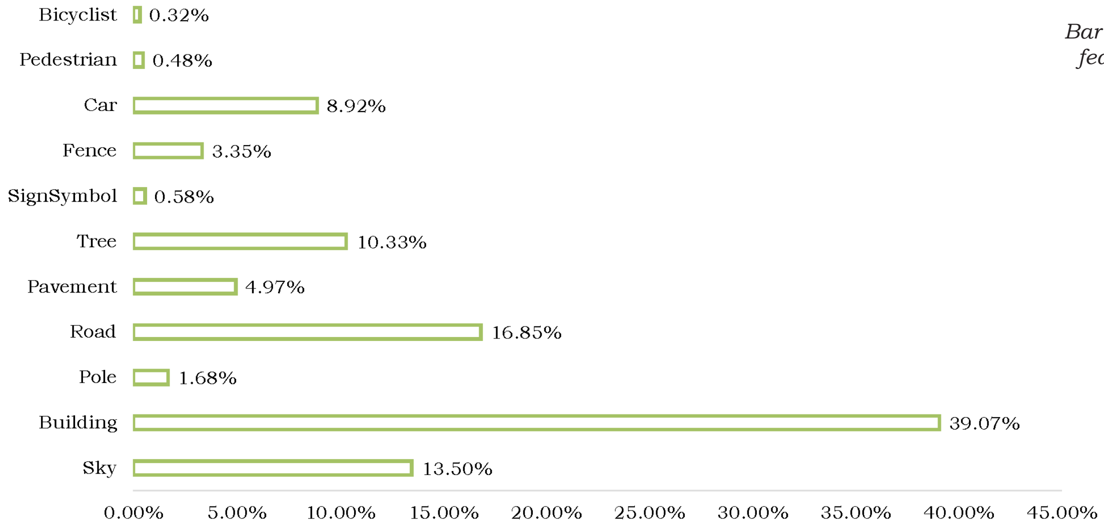
Bar chart 5.12.L1 Street features distribution in low crime rate

L1



Bar chart 5.12.L2 Street features distribution in low crime rate

L2



Bar chart 5.12.L3 Street features distribution in low crime rate

L3

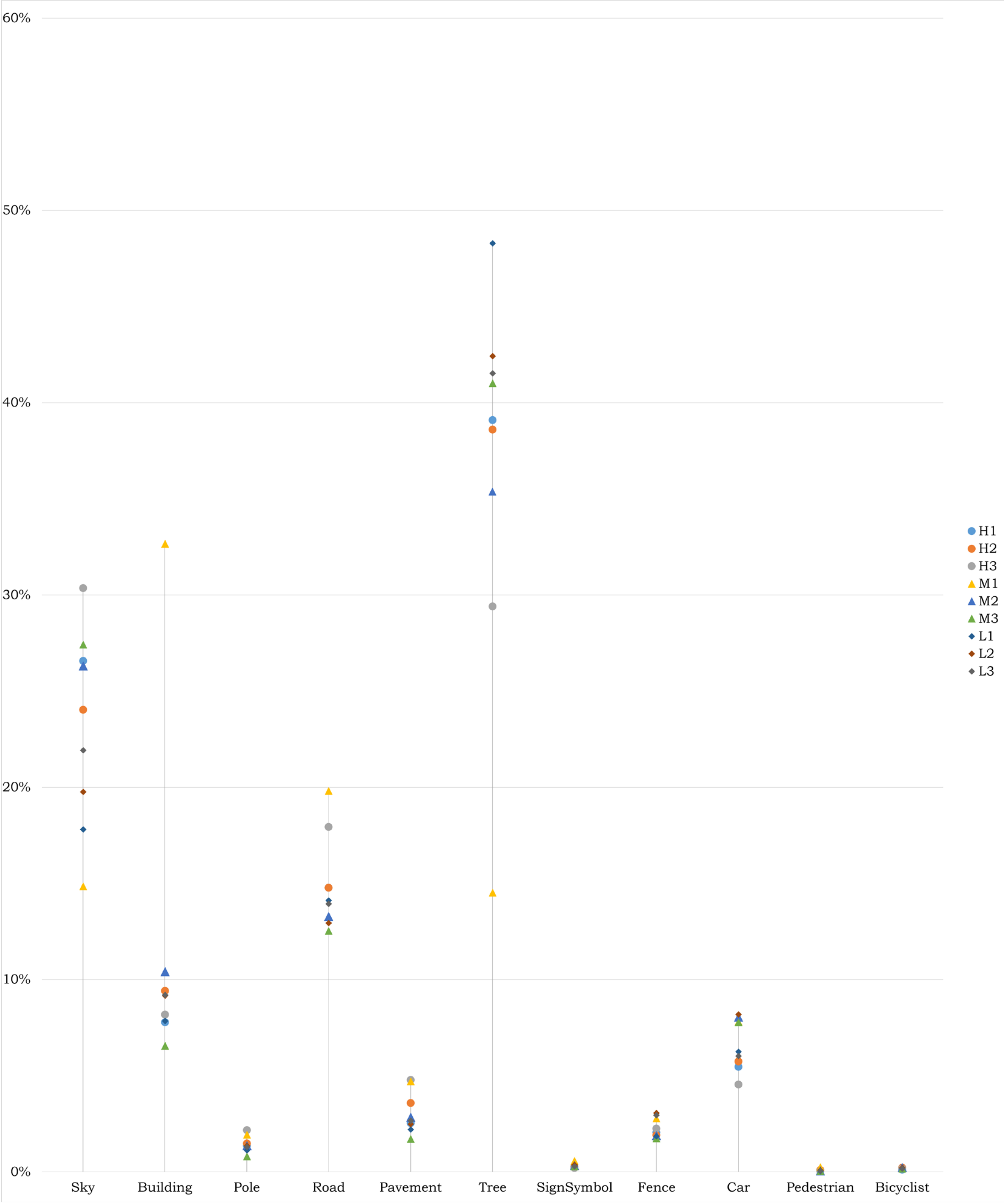


Figure 5.29 Street features distribution in different crime rate region of Detroit

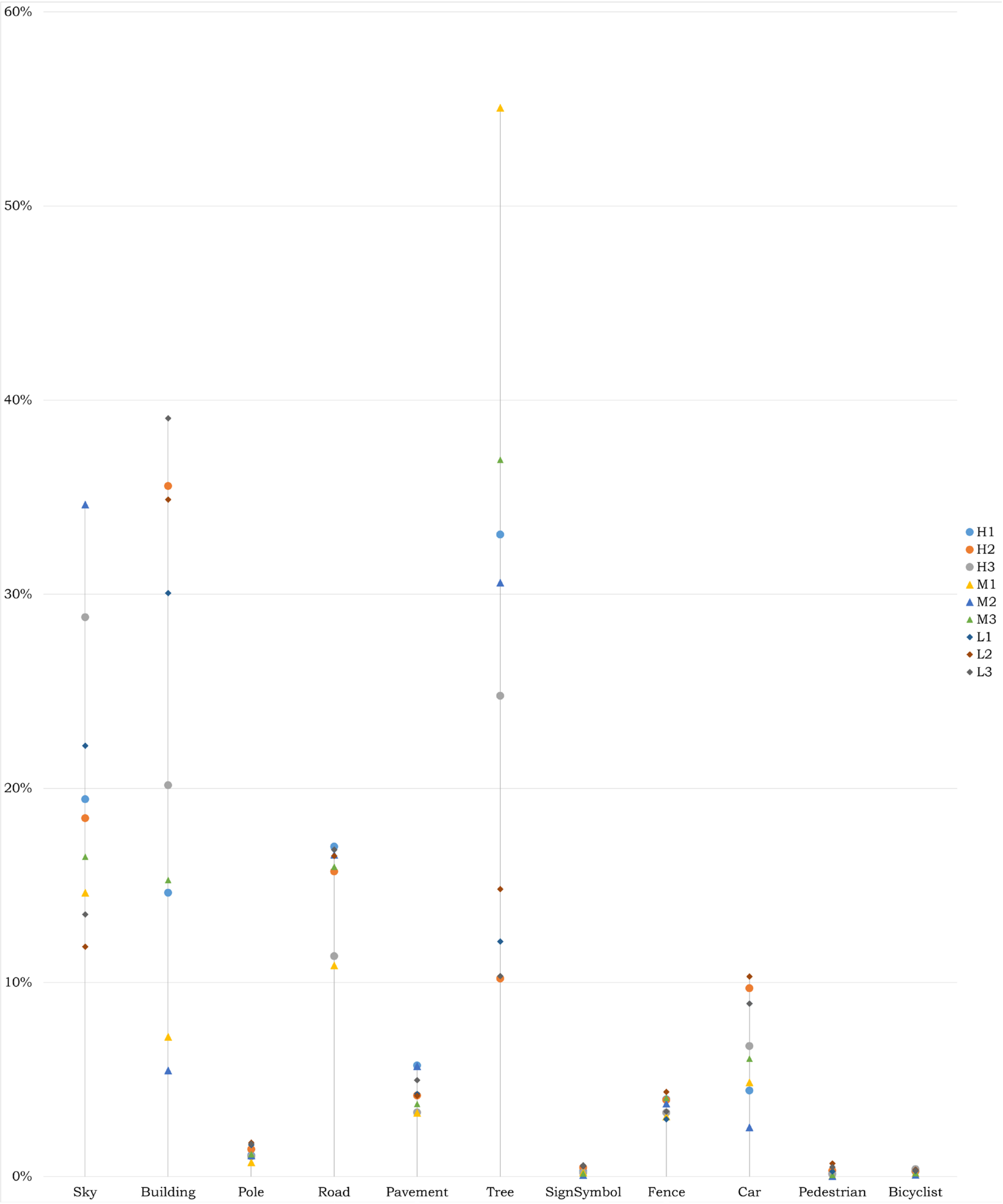


Figure 5.30 Street features distribution in different crime rate region of Johannesburg

References

Mumford, L. (1937). What is a city. Architectural record, 82(5), 59-62.

Tyrväinen, L., Mäkinen, K., & Schipperijn, J. (2007). Tools for mapping social values of urban woodlands and other green areas. Landscape and urban planning, 79(1), 5-19.

Beaney, K. (2009). Green spaces in the urban environment: uses, perceptions and experiences of Sheffield city centre residents (Doctoral dissertation, University of Sheffield).

Newman, O. (1996). Creating defensible space, us department of housing and urban development. Office of Policy Development and Research, Rutgers University.

Ashihara, Y. (1983). The aesthetic townscape. The MIT press.

Ashihara, Y. (1981). Exterior design in architecture. Van Nostrand Reinhold Company.

Lorè, I. (2016). The Development and Promotion of the Inland Areas of the Metropolitan City of Reggio Calabria through the Enhancement and Restoration of the Calabro-lucane Railway Line– The Greenway Project, Train-hotel and Valorization of Former Railway Stations. Procedia-Social and Behavioral Sciences, 223, 363-370.

Habitat, U. N. (2007). 3. Enhancing Urban Safety and Security: Global Report on Human settlements. Earth Scan.

Websites:

https://en.wikipedia.org/wiki/Street_canyon, retrieved on 03-12-2017

https://en.wikipedia.org/wiki/Crime_prevention_through_environmental_design, retrieved on 03-12-2017

Chapter Content

6 Conclusions

6 Conclusions

From the research, we summarized what we have done and also realized that it still has some limits and shortage because of various reasons, but in future, we will study on those problems, enlarge and perfect our application.

6.1 Main Contributions

Our research aims at the applicability of Computer Vision and Deep Learning in the field of urban planning and urban design. Kevin Lynch came up with the five types of elements that constituted the physical forms, in another word, city images, paths, edges, districts, nodes, and landmarks (The Image of the City, Kevin Lynch); which provides a method to deal with urban visual form in urban dimension, also some principles for urban design.

However, human visual perception is the most subject, different people have various feeling to the same urban scene, we expect to use Computer Vision technology for analyzing the city image, which makes the research of urban spatial features more rational and scientific. Therefore, in order to keep the targeted quality and operability, we concentrated on "path", which is the urban street in our research.

> *Paths are the channels along which the observer customarily, occasionally, or potentially moves. They may be streets, walkways, transit lines, canals, and railroads. For many people, these are the predominant elements in their image. People observe the city while moving through it, and along these paths the other environmental elements are arranged and related. (Lynch, K. (1960). The image of the city (Vol. 11). MIT press, 41)*

1) The exploration of the application of Computer Vision and Deep Learning in the urban design field. Recently, Deep Learning technology has been developed rapidly and has been become the most expected Computer Technology. At the same time, because of the industry barrier, the potential of cross-industry application of Deep Learning technology has not been explored completely. Currently, Deep Learning technology is the main approach for implementing Computer Vision technology.

In the urban design field, we found less document related to Deep Learning technology and the relationship between urban design and Deep Learning is prefer to stay in the strategic plan and there is the little case in practice. As we know much more about the Computer Vision and Deep Learning, we found the potential and possibility of application in urban design. More important is that fast learning ability and generalization ability of Deep Learning technology, which is implemented by a part of our research that includes computer learning, extracting features and quantification.

2) We proposed a method, UrSA. We found that the merging of Computer Vision technology and Deep Learning technology, which has potential and possibility to apply in urban design field, and then we invented a technical framework of classification of urban space in street level - UrSA. UrSA includes all the techniques of street image acquisition, image processing, image features learning and output of image features. Besides, we put the analysis of street feature from image and visualization into UrSA; therefore, it is a completely technical framework to do the research and analysis Google Street View (GSV) images.

3) We proposed VCI concept to measure the human perception of the urban street and describe the urban streets' features. UrSA as a technology for urban street analysis also has an ability to explore the primary outcomes and maximize the utility, also it is helpful for urban design and provides a forward-looking direction. Therefore, we proposed VCI concept as an exploratory research, and VCI is the refining and distillation of urban street spatial features that are summarized by UrSA.

4) Using UrSA to study in the relationship between social issues and urban street space. We took VCI to bridge the social issues (crime rate and housing price) and urban space, in order to have a better work of UrSA, which applied in a certain space in which social issued existed to explore the relationship.

5) Implementation of quantification of human visual perception. Quantification of urban streets' features by UrSA, it makes much more rational and objective in the analysis of human visual perception. UrSA could provide guidance for urban design pursuing preciseness and accuracy.

6.2 Limitation of the Research

The main limitation of the research is from technology and source.

1) Technical limitation: in our research, we did not use the latest Convolutional Neural Network (CNN) model, but the model we used is relatively more targeted, which limited the maximization of research effect to some extent.

2) Resource limitation: although Google opens the GSV image to the public, the limitation is a number of download image. Thus, we chose a small range of image collection in final, which had an influence in authority and generality.

3) Accuracy limitation: because we adopted the supervised learning method whose character is that the wider of data, the better of effect, in another word, the bigger of labeled datasets, the better of accuracy. Our research was based on the small sample, thus it is hard to maximize the accuracy.

4) Our research is more explorative than practical. The aim of the research is to implement Deep Learning and Computer Vision technology and apply in urban design, but currently, in terms of the final result, we still need to do the deep research then have an official practice of UrSA.

5) Platform limitation: we employed much software in our research, and it took a lot of time on transformative use, which lowered operational convenience. Now, the complicated operation of UrSA making a worse interactive experience for the user.

6.3 Further Work

We will continue to make the Computer Vision and Deep Learning technology into the application in urban planning field in the further work, so we plan to follow:

1) Modify the kernel of UrSA; try to employ the best CNN model into UrSA in order to improve the efficiency and accuracy;

2) Optimize the workflow of UrSA and create an integrative operation interface to improve the operational convenience;

3) Construct the Web service for UrSA to display the research outcome and serve for public;

4) Study much more on the social issues and its relationship

with urban space;

5) A comprehensive research on five elements and of city image, paths, edges, districts, nodes, and landmarks, and its relationship.

6) Explore the possibility of employing unsupervised learning in UrSA, because it can promote the intellectual level of UrSA;

7) Enlarge the research range, such as number of city and street;

8) To obtain the higher permission for deep research.

References

Lynch, K. (1960). The image of the city (Vol. 11). MIT press.

Appendix

Appendix.1 Semantic Segmentation Code

MATLAB code:

1- Load a pretrained version of SegNet

```
pretrainedFolder = fullfile(‘E:\’,’Test_matlab’,’pretrainedSegNet’);
pretrainedSegNet = fullfile(pretrainedFolder,’segnetVGG16.mat’);
```

2- Load Training Dataset

```
trainedimgFolder = fullfile(‘E:\’,’Test_matlab’,’Dataset’);
```

3- Load Training Images

```
imgDir = fullfile(trainedimgFolder,’images’);
imds = imageDatastore(imgDir);
```

4- Load Pixel-Labeled Images

```
classes = [
    “Sky”
    “Building”
    “Pole”
    “Road”
    “Pavement”
    “Tree”
    “SignSymbol”
    “Fence”
    “Car”
    “Pedestrian”
    “Bicyclist”
    ];
labelIDs = PixelLabelIDs();
labelDir = fullfile(trainedimgFolder,’labels’);
pxds = pixelLabelDatastore(labelDir,classes,labelIDs);
```

5- Analyze Training Dataset Statistics

```
tbl = countEachLabel(pxds);
```

6 -Count Number of Training Images

```
numImages = numel(imds.Files)
```

7- Create the Network

```
imageSize = [360 480 3];
numClasses = numel(classes);
lgraph = segnetLayers(imageSize,numClasses,’vgg16’);
```

8- Balance Classes Using Class Weighting

```
imageFreq = tbl.PixelCount ./ tbl.ImagePixelCount;
classWeights = median(imageFreq) ./ imageFreq;
pxLayer = pixelClassificationLayer(‘Name’,’labels’,’ClassNames’, tbl.
Name, ‘ClassWeights’, classWeights);
lgraph = removeLayers(lgraph, ‘pixelLabels’);
lgraph = addLayers(lgraph, pxLayer);
lgraph = connectLayers(lgraph, ‘softmax’ ,’labels’);
```

9- Select Training Options

```
options = trainingOptions(‘sgdm’, ...
    ‘Momentum’, 0.9, ...
    ‘InitialLearnRate’, 1e-3, ...
    ‘L2Regularization’, 0.0005, ...
    ‘MaxEpochs’, 100, ...
    ‘MiniBatchSize’, 64, ...
    ‘Shuffle’, ‘every-epoch’, ...
    ‘VerboseFrequency’, 2);
```

10- Data Augmentation

```
augmenter = imageDataAugmenter(‘RandXReflection’,true,...
    ‘RandXTranslation’, [-10 10], ‘RandYTranslation’,[-10 10]);
```

11- Load Target Dataset

```
TargetimgFolder = fullfile(‘E:\’,’Test_matlab’,’TargetDataset’);
```

12- Load Target Images

```
imgDir = fullfile(TargetimgFolder,’images’);
imds = imageDatastore(imgDir);
```

13- Start

```
datasource = pixelLabelImageSource(imds,pxds,...
    ‘DataAugmentation’,augmenter);
```

```
data = load(pretrainedSegNet);
net = data.net;
```

14- Analyze Target Dataset Statistics

```
pxdsResults = semanticseg(imds,net,”WriteLocation”,’forder of results’);
tbl = countEachLabel(pxdsResults)
frequency = tbl.PixelCount/ sum(tbl.PixelCount);
figure
bar(1:numel(classes),frequency)
xticks(1:numel(classes))
xticklabels(tbl1.Name)
xtickangle(45)
ylabel(‘Frequency’)
```

15- Transfer unit8 to rgb

```
all_LabImgs=dir(‘E:\ind2rgb_output_forder\*.png’);
LabDir = fullfile(‘E:\’,’ind2rgb_output_forder’,filesep);
all_Imgs=dir(‘forder of results\*.png’);
LabNum=length(all_LabImgs);
disp(LabNum)
for n=1:LabNum
    labsname = all_LabImgs(n).name;
    BW = imread([LabDir, labsname]);
    X = uint8(BW);
    csmap = ColorsMap;
    RGB = ind2rgb(X,cmap);
   dirpath = ‘E:\unit8toColor’;
    if ~exist(dirpath,’file’)
        mkdir(dirpath);
    end
    imgsname = all_Imgs(n).name;
    imgsname(end-3:end)=[];
    LabFolder = fullfile(dirpath,filesep);
    imwrite(RGB,[LabFolder,imgsname,’_L.png’],’WriteMode’,’append’);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
function labelIDs = PixelLabelIDs()
labelIDs = { ...
    % “Sky”
    [128 128 128;]
    % “Building”
    [128 000 000; ]
    % “Pole”
    [192 192 128;]
    % Road
    [28 064 128;]
    % “Pavement”
    [060 040 222;]
    % “Tree”
    [28 128 000; ]
    % “SignSymbol”
    [92 128 128; ]
   % “Fence”
    [064 064 128;]
    % “Car”
    [064 000 128;]
    % “Pedestrian”
    [064 064 000;]
    % “Bicyclist”
    [000 128 192;]
    };
end
% % % % % % % % % % % % % % % % % % % % % % % % % % %
function pixelLabelColorbar(cmap, classNames)
colormap(gca,cmap)
c = colorbar(‘peer’, gca);
c.TickLabels = classNames;
numClasses = size(cmap,1);
c.Ticks = 1/(numClasses*2):1/numClasses:1;
c.TickLength = 0;
end
% % % % % % % % % % % % % % % % % % % % % % % % % % %
```

```
function cmap = ColorMap()
cmap = [
    128 128 128   % Sky
    128 0 0       % Building
    192 192 192   % Pole
    128 64 128    % Road
    60 40 222     % Pavement
    128 128 0     % Tree
    192 128 128   % SignSymbol
    64 64 128     % Fence
    64 0 128      % Car
    64 64 0       % Pedestrian
    0 128 192     % Bicyclist
    ];
% Normalize between [0 1].
cmap = cmap ./ 255;
end
% % % % % % % % % % % % % % % % % % % % % % % % % %
function csmap = ColorsMap()
cmap = [
    0 0 0         % Others
    128 128 128   % Sky
    128 0 0       % Building
    192 192 192   % Pole
    128 64 128    % Road
    60 40 222     % Pavement
    128 128 0     % Tree
    192 128 128   % SignSymbol
    64 64 128     % Fence
    64 0 128      % Car
    64 64 0       % Pedestrian
    0 128 192     % Bicyclist
    ];
% Normalize between [0 1].
cmap = cmap ./ 255;
end
```

Appendix.2 Quantification of Image Information

```
import os
import random
import pandas as pd
import numpy as np
from PIL import Image

df = pd.DataFrame()
dataroot = csv file’s forder path
imgroot = imgae’s forder path
file = imgroot + ‘51.510444,-0.137694_0_L.png’
print (len(file))
image = Image.open(file)
counts = image.getcolors()
df1 = pd.DataFrame(counts, columns = [‘NUM’, ‘RGB’])
df1[‘file’] = file[60:]
df1.to_csv(file+’.csv’)
df2 = df1.pivot(index=’file’, columns=’RGB’, values=’NUM’)
df2.to_csv(dataroot+file[60:-4]+’.csv’)

def get_imlist(path):
    return [os.path.join(path,f) for f in os.listdir(path) if f.endswith(‘.png’)]
filelist = get_imlist(imgroot)
for c in range(len(filelist)):
  file = filelist[c]
  image = Image.open(file)
  counts = image.getcolors()
  df1 = pd.DataFrame(counts, columns = [‘NUM’, ‘RGB’])
  df1[‘file’] = file[60:]
  df1.to_csv(file+’.csv’)
  df2 = df1.pivot(index=’file’, columns=’RGB’, values=’NUM’)
  df2.to_csv(dataroot+file[60:-4]+’.csv’)
  with open(dataroot+’Labels.csv’, ‘a’) as f:
        df2.to_csv(f, header=True)
```

Appendix.3 Transfer Coordinate to Address

```
import os
import googlemaps
import pandas as pd
import numpy as np
from pandas import DataFrame

#######################################################
# # GitHub: https://github.com/googlemaps/google-maps-services-python
# reverse_geocode_result = gmaps.reverse_geocode(client, latlng, result_type=None, location_type=None, language=None)
# result_type - ‘country’,’street_address’,’postal_code’
# location_type - ‘ROOFTOP’,’RANGE_INTERPOLATED’,’GEOMETRIC_CENTER’,’APPROXIMATE’
#######################################################

# request imgs list from a csv_file
dataroot = ‘E:/csv_file forder/’
df = pd.read_csv(dataroot+’test.csv’).astype(str)
df2 = pd.DataFrame()
#get lat&lng list
lst = zip(df.lat,df.lng)
latlng = list(lst)
gmaps = googlemaps.Client(key=’geocoding API key’)

for c in range(len(latlng)):
    num = df.num[c]
    reverse_geocode_result = gmaps.reverse_geocode(latlng[c], result_type=’street_address’, location_type=’ROOFTOP’, language=’en’)
    if reverse_geocode_result:
        geocode_result = reverse_geocode_result[0]
        address = geocode_result[‘formatted_address’]
        location = geocode_result[‘geometry’][‘location’]
        lat= location[‘lat’]
        lng = location[‘lng’]
        address_dir = [num, address, lat, lng]
        address_dir = np.array([address_dir])

        df1 = pd.DataFrame(address_dir, columns=[‘num’,’address’,’Y’,’X’])
        df2 = df2.append(df1, ignore_index=True)
        df3 = pd.merge(df, df2, how=’left’,on=’num’)
        df3.to_csv(dataroot+’NY_kfinal_ttt.csv’)

    else:
        pass
```

Appendix.4 Get Road Coordinates

```
import os
import googlemaps
import pandas as pd
import numpy as np
from pandas import DataFrame

dataroot = ‘E:/Test_matlab/Get_road_location/’
gmaps = googlemaps.Client(key=’geocoding API key’)

path= [‘45.480510, 9.224529’, ‘45.481617, 9.227574’, ‘45.482726, 9.230894’, ‘45.483888, 9.234211’]
snap_result = gmaps.snap_to_roads(path, interpolate=True)
df = pd.DataFrame()

for c in range(len(snap_result)):
    list_location = snap_result[c]
    location = list_location[‘location’]
    lat= location[‘latitude’]
    lng = location[‘longitude’]
    location_dir = [lat, lng]
    location_dir = np.array([location_dir])
    df1 = pd.DataFrame(location_dir, columns=[‘Y’,’X’])
    df = df.append(df1, ignore_index=True)
    df.to_csv(dataroot+’Road.csv’)
```