

Context-rich Urban Analysis Using Machine Learning

A Case Study in Pittsburgh, PA

Jinmo Rhee

School of Architecture, College of Fine Arts,
Carnegie Mellon University,

Master of Science in Computational Design

2019

CARNEGIE MELLON UNIVERSITY

School of Architecture
College of Fine Arts

Thesis

Submitted in Partial Fulfillment of the requirements for the degree of

Master of Science in Computational Design

TITLE:

Context-rich Urban Analysis Using Machine Learning

A Case Study in Pittsburgh, PA

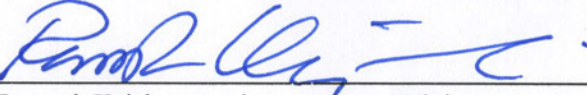
AUTHOR:

Jinmo Rhee

ACCEPTED BY ADVISORY COMMITTEE:


Daniel Cardoso Llach Principal Advisor

MAY 15, 2019
DATE

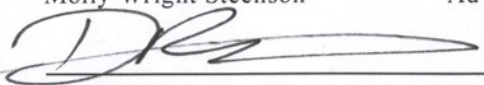

Ramesh Krishnamurti Advisor

5/13/2019
DATE



Molly Wright Steenson

Advisor



Daragh Byrne

Advisor



Eddy Man Kim

Advisor

13 May 2019

DATE

13th MAY '19

DATE

May 10, 2019

DATE

Table of Contents

1 Introduction	1
1.1 Intermediate Scale Between City and Building	1
1.2 Intended Contributions	3
2 Literature Review	7
2.1 City as Data	7
2.1.1 Urban Granularity: Multi-Elements	8
2.1.2 Urban Collectivity: Patterns	10
2.2 Context-ness and Contextless-ness	11
2.2.1 Contextless-ness cases	11
2.2.2 Context-ness cases	14
2.2.3 Contextuality in Architectural Design and AI	15
3 Contexts and AI in Architecture and Urbanism	19
3.1 Representation of Architectural Contexts for Machine Learning	20
3.1.1 Diagrams in Architecture	20
3.1.2 Diagrams for Machine	21
3.1.3 Advantages of the Diagrammatic Image	23
3.1.4 Diagrammatic Image Dataset (DID) Generation	24
3.2 Recognition of Context-rich Dataset	26
3.2.1 Overview	27
3.2.2 Setup for DID of Seoul	28
3.2.3 Models and Learning	30
3.2.4 Results & Discussion	31
3.3 Usability of Context-rich Dataset	33
3.3.1 Overview	34
3.3.2 Setup for DID of Shadyside	35
3.3.3 Models and Learning	37
3.3.4 Results & Discussion	37
4 Urban Morphological Types: Analysis Architectural Contexts	41
4.1 Dataset	43

4.1.1 Data Pre-Processing	44
4.1.2 DID of Pittsburgh (DID-PGH)	48
4.1.3 Characteristics of DID	49
4.2 Clustering Based on Similar Context Conditions	52
4.2.1 Dimensionality Reduction	52
4.2.2 Clustering by Similar Context Conditions	54
4.3 Data-driven Map of Urban Morphological Types (DMUMT)	54
4.3.1 Map of Building Clusters and Urban Types' Boundaries	55
4.3.2 Morphological Analysis by Clusters	57
4.4 Interpretation of DMUMT	61
4.4.1 Conflicts among Urban Fabrics	62
4.4.2 Historical Reasoning	65
5 Discussion and Applications	69
5.1 The Value of the Urban Analysis through Context-rich Dataset	69
5.2 The Applicable Scenarios	72
6 Conclusion	75
Appendix	79
Appendix 1	80
Appendix 2	82
Appendix 3	82
Appendix 4	83
Appendix 5	84
Appendix 6	85
Appendix 7	86
Appendix 8	87
Appendix 9	88
Appendix 10	89
Appendix 11	90
References	95
List of Figures	100
List of Tables	103

Context-rich Urban Analysis Using Machine Learning

A Case Study in Pittsburgh, PA

Jinmo Rhee

School of Architecture, College of Fine Arts,
Carnegie Mellon University,
Pittsburgh, PA 15213
jinmor@andrew.cmu.edu

Abstract. This thesis reports on the analytic and generative potential of using machine learning methods on context-rich urban datasets. Working with a dataset of Pittsburgh's urban data, this research proposes a novel method to extract information about each building in the city in ways that reflect its immediate urban context and analyze it in terms of urban homogeneity through machine learning algorithms. The method can be used for defining urban morphological types as the clusters of urban spaces sharing the similar architectural contexts conditions by classifying and clustering of all buildings in Pittsburgh based on their footprints and contexts conditions. Then, the results of the clustering can be remapped to the real-world map, generating a data-driven map of urban morphological types. The interpretations and application of this map and method will be discussed. From these discussions, I will address the value, meaning, and potential of this new method in terms of architectural design, urbanism, and computation.

Keywords: Urban Morphology, Machine Learning, Context-rich, Urban Analysis, GIS

1 Introduction

In contemporary architectural design, the relationship between architecture and the city has become a critical element of urban architecture. Within this aspect, how the architect's proposal is influenced by the existing urban space or influences them has become a key issue in architectural design. The existing spaces are the circumstances that form a setting for specific conditions of a site or building, called architectural contexts. As the definition of the urban morphology¹, it is hard to establish the concrete and clear definition of architectural contexts too. In this thesis, I define the architectural contexts as the surrounding physical conditions of urban artifacts within a certain range around the target building in terms of the aspects of urban morphology. Urban artifacts are the terminology that Aldo Rossi used to explain the city as a set of man-made objects², such as buildings, infrastructures, public space, etc. Thus, the physical conditions of urban artifacts refer to a situation of how these man-made objects form the city.

Several researchers have recently tried to use computational approaches to urban morphology to obtain more delicate results of an analysis with higher complexity. Various computation methods are employed to define and distinguish the urban morphological types of a city, such as the statistical model³, space syntax⁴, convolutional neural network⁵, etc. Their methods have suggested the new point of view to the urban space. However, their datasets only include the feature metrics of every single element that cannot include the architectural context representing the relationships among these elements.

In this thesis, I will report the development of a novel method to create an urban dataset that contains information about a building and its architectural context. Also, through two experiments, the recognition and generative potential of dataset to be a design basis will be tested (Chapter 3). Defining the urban morphological types as the clusters of urban spaces sharing the similar architectural contexts conditions, I will classify and cluster all buildings in a city by their footprints and contexts conditions using machine learning algorithms. Then, the results of the clustering can be remapped to the real-world map, generating a new type of map: data-driven map of urban morphological types (Chapter 4). The interpretations and application of this new map and its method will be discussed (Chapter 5). From these discussions, I will address the value, meaning, and potential of this new method in terms of architectural design, urbanism, and computation (Chapter 6).

1.1 Intermediate Scale Between City and Building

As today's city is becoming larger and more complex to interpret as a simple huge object, there have been attempts to analyze the morphological rules of the city. These morphological rules include specific arrangements and repetitive variations, called pattern or urban pattern. Patterning allows to group and differentiate individual information in

1. Vitor, *Urban Morphology*, p.2

2. Rossi, *The Architecture of the City*, p.24

3. Colaninno et al, "An Automatic Classification of Urban Texture: Form and Compactness of Morphological Homogeneous Structures in Barcelona."

4. Alexander, "A CITY IS NOT A TREE."

5. Albert et al, "Using Convolutional Networks and Satellite Imagery to Identify Patterns in Urban Environments at a Large Scale."

6. Vitor, *Urban Morphology*, p.3



Fig. 1. Example of Urban tissue shown at increasing levels of resolution

7. Kropf, “Urban Tissue and the Character of Towns.”, p.8

elements of urban form, such as streets, street blocks, plots, and buildings⁶. The results of patterning have an intermediate scale between the city and the building. It is smaller than the city scale but larger than the elements scale. Therefore, patterning of urban elements have been used a method to analyze urban space by investigating the morphological traits of intermediate scale urban forms, such as urban fabrics.

Karl Kropf described the relationship between patterning and urban fabric, or urban tissue in his paper (Fig. 1). In his viewpoint, patterns with complexity of urban forms can cover the character of a town or a town as whole. Urban form, in his definition, includes the seven primary components: materials, structures (i.e. encompassing details of construction), rooms, buildings, plots, streets and blocks, and urban fabrics. Particularly, the urban fabric is a synthesis of all components⁷. Therefore, the composition of urban form with complexity is the urban fabric that can be differentiated covering and illustrating the character of a part of city or a city as a whole.

The general method to use the concept of urban pattern in urban architecture is the figure-ground map (Fig. 2) which is a mapping method for the characteristics of city space by distinguishing buildings and land with black and white solids. The result of this method is a map that shows the spatial characteristics of the city by showing how public and private space exist in the city.

Another method for the concept of urban pattern is to distinguish monumental and mundane architecture (Fig. 2), so called the contexts of monuments⁸. This method shows not only how monumental buildings collide and coexist with mundane architecture in everyday space, but also how individual and repetitive types of buildings in the city are abstracted to create urban patterns.

But those representative methods for urban patterning are not enough to analyze the complex and diverse phenomena associated with contemporary urban spaces. Contemporary approaches to the matter of defining the urban morphological types and patterns with advanced computational methods, allows people to statistically calculate the similarity of each building to what typologies, and to express the urban organization

Fig. 2. The Nolli map of Rome, Giambattista Nolli (Left), Imperial Rome, views of model at the Museo della Civiltà Romana (Middle), The matrix of 17th century Rome – plan after Bufalini (Right)



in urban space through mapping. The use of computational methods in the study of urban morphology shows the potential of more scrutiny with urban data and spatial investigation of the phenomena in contemporary cities.

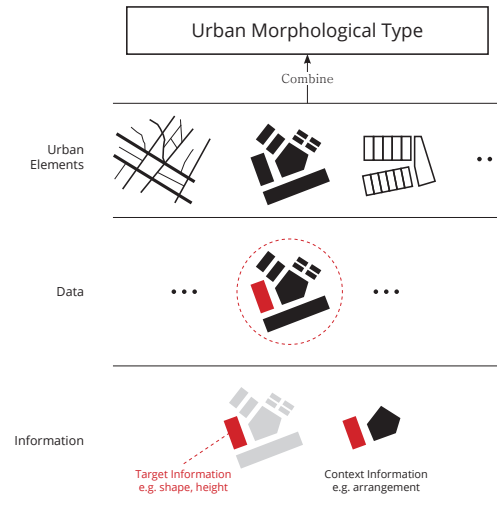


Fig. 3. A Concept of Urban Morphological Type with the Computational Aspect

Patterns with the complexity of urban forms can generate types which represent urban morphological traits such as repetitive arrangements buildings, similar hierarchical street system, regularly changing blocks, etc. These types are called ‘urban morphological types.’ Computational approaches to urban patterns can allow researchers to define urban morphological types with a large amount of urban data and high scrutiny. This thesis treats urban elements as urban data including the target or context information, and assumes that relationships among the data are reflected in urban morphological types (Fig. 3). For example, one building data has two different information: target and context. Shape and height of the building is the geometrical aspects of this information, which can be target information representing the feature of internal aspect of the object. Arrangement or locational information with other buildings, even still geometrical, can be context information representing the relationships. While Karl accentuate the notion of resolution in urban fabrics due to complexity of urban data, resolution does not the key component for this definition due to the advantages of the machine’s ability to calculate the high degree of complexity.

1.2 Intended Contributions

With the viewpoint of the urban patterning as the intermediate scale between building and city, this thesis foots on the studies of urban morphology and proposes the urban morphological types with advanced computational method, machine learning. This research will use machine learning techniques for the analysis of urban elements and

their relations to reveal the distinctive urban morphological types of a city with the research question: “How can AI technologies and context-rich urban data improve the analytical and generative understanding of city space?” I expect three contributions from this urban analysis research with context-rich dataset through advanced computational techniques: *context-rich data format to represent the city structures and their traits, urban morphological types sharing similar architectural context conditions, a bottom-up analysis method for the whole city.*

In terms of advanced computational techniques, this thesis will suggest a novel type of image-based data format to represent a city by each building in the city and its architectural context conditions for AI models. The change of the data format does not simply mean the new types of dataset, but it means new objects to learning for AI techniques. In this thesis, as the AI model should be trained to capture the features of the architectural context conditions, it needs the new data format including the new representation for the increased possibility of the successful learning.

Regarding urban analysis for architectural design, this thesis will identify urban morphological types derived from the homogeneity of architectural contexts conditions and provide a new boundary map of these types. These boundaries can be the new concept of the urban fabrics’ boundaries. This map will illustrate the boundaries which are different from conventional neighborhood boundaries and the more urban spatial characteristics. In urbanism, unprecedented way to define urban fabrics will open the new potential that diverse urban phenomena can be interpreted in terms of urban space.

Finally, in terms of the methodology of urban spatial research, this thesis will provide the new concept of analysis in urban scale: bottom-up analysis. Geographic information systems (GIS) is a tool to investigate urban spatial characteristics with various phenomena in a city. Conventional GIS usually employs top-down methods with individual data. For example, conventional GIS tool can statistically calculate the density of buildings in Shadyside, Pittsburgh, because the data has tags which are pre-defined categories by the administrative view. However, this data cannot be used to discover how an individual geographical data is related to the other data or overall the city. Compared to the conventional geographic information systems (GIS), the method in this thesis can individually investigate the characteristics of each data and collectively handle the relationships with other data in the viewpoint of the overall dataset at the same time.

2 Literature Review

This chapter reviews precedents and literatures in two aspects: city as data and contextuality. The first section examines how the characteristics of a city are interpreted in terms of data with computational approaches. I premise the city as the collective form of the individuality. The number of various individualities, called ‘urban granularity’, reflects the urban properties of the multi-elements that architect cannot precisely catch. Through the precedents of the urban studies and computational approaches, the first section will show how the computational methods can be used to handle the urban granularity. In addition, this section will qualitatively evaluate the various computational approaches handling the collectivity through the idea of patterns as the media to understand the urban granularity. These evaluations lead to the viewpoint about the urban morphological types as the medium between the city and buildings in the intermediate scale of them.

The second section of this chapter will discuss the aspect of the contextuality from precedents and referenced researches using AI for architectural design. I analyze the task types, algorithms, and research year of each study. Then, after setting the proper definition of contextuality for classifying the ‘contextless-ness’ and ‘context-ness’, I distinguish and indicate all references whether contextless-ness or context-ness. Based on individual classification and analysis of each reference, the tendency of the research types, history, and tasks will be statistically analyzed.

2.1 City as Data

This section reviews several precedents dealing with the individuality of cities, named ‘urban granularity’ and how patterns are used to represent ‘urban collectivity’ through this individuality. Specifically, from the aspect of computation in urbanism and architectural design, I investigate what data types represent this individuality of cities, what information and features of individuality the data contains, and how it can be a dataset forming the urban granularity. Besides, I explore the patterning methods from precedents in computational aspects, focusing on what kinds of computational process or algorithms they used for the data of urban granularity.

In addition to the analysis of urban granularity with computational approaches, this section demonstrates how the definitions of urban fabrics from precedents are newly interpreted in terms of individuality and collectivity. Through these analyses mentioned above, at the end, this section discusses the potential that the notion of urban fabrics can be perceived as an intermediate medium between individuality and collectivity. Urban fabrics as intermediate medium will become a material for a tool to newly interpret urban space by their definitions and view of city as data.

9. Sevtsuk, “Path and Place.”

10. Jiang et al, “Discovering Urban Spatial-Temporal Structure from Human Activity Patterns.”

11. Colaninno et al, “An Automatic Classification of Urban Texture: Form and Compactness of Morphological Homogeneous Structures in Barcelona.”

12. Zhou et al, “Recognizing City Identity via Attribute Analysis of Geo-Tagged Images.”

13. Albert et al, “Using Convolutional Networks and Satellite Imagery to Identify Patterns in Urban Environments at a Large Scale.”

14. Brandão and Ricardo Correia, “Rhythms of Renewal of the City.”

Table. 1. Analysis of precedents' Granularity and Collectivity

2.1.1 Urban Granularity: Multi-Elements

This small section and the following sub-section analyze six different precedents about the urban analyses that create urban patterns through computation (Table 1): Path and place : a study of urban geometry and retail activity in Cambridge and Somerville, MA⁹(Research A), Discovering urban spatial-temporal structure from human activity patterns¹⁰(Research B), An automatic classification of urban texture: form and compactness of morphological homogeneous structures in Barcelona¹¹(Research C), Recognizing City Identity via Attribute Analysis of Geo-tagged Images¹²(Research D), Using Convolutional Networks and Satellite Imagery to Identify Patterns in Urban Environments at a Large Scale¹³(Research E), and Rhythms of Renewal of the City¹⁴

	Research A	Research B	Research C	Research D	Research E	Research F
Data Type	A Matrix for Their Variables	A Matrix for Their Variables	A Matrix for Their Variables	Geo-tagged Images	Satellite Images	Georeferenced Building Shape File
Representation	Graph	Own Statistical Variable	Spatial Cells	Spatial Cells	Unit Square	Geometry and Tagged Information
Patterning Objective	Retail Location	Activity Type	Compactness	Attributes Detection	Land Use Detection	Renewal Processing Type
Patterning Algorithm	Linear Probability Model	K-Means Clustering via PCA	PCA	SVM	CNN, t-SNE	A Discriminant for Renewal Types
Patterning Results	Building Map with Assessment for Retailers	Map of Activity by Time and Programs o Space	Map of Urban Fabrics	Map by Different Urban Attributes	Prediction Map of Land Use	Map of Building Permit Approval Time of Renovation and Construction

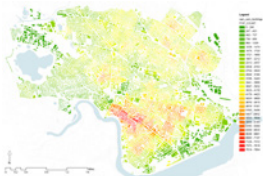


Fig. 4. Research A, Reach to built volume within a 2,000ft network radius from each building in Cambridge & Somerville, MA

15. Sevtsuk, p.3

(Research F).

Sevtsuk's works (Fig. 4) is the research about the retail location choice by using graph model to investigate the pattern of retail location in urban space. It proposes a novel graph-analysis framework in which retail location patterns can be represented under realistic constraints of urban geometry, land use distribution, and travel¹⁵. After creating the metrics of quantified variables, such as mode of travel, access to building, and remoteness, it uses linear probability model to score the buildings for location choices of different retail categories. The results of this research were the map of buildings colored with the assessment of retailers by retail types.

Jiang's research (Fig. 5) is the research about the relationship between urban space and activity types by time and users. After creating the metrics of activity attributes, such as time stamps of individuals' activities, activity locations, activity types, etc., it used its own criteria, called 'Urban Spatial-temporal Structure, USTS', to calculate and distinguish activity on the map. This employed principle components analysis (PCA) and K-Means clustering to detect clusters of individuals by daily activity patterns,

integrated with their usage of space and time, and show that daily routines can be highly predictable, with clear differences depending on the group (e.g. students vs. part time workers)¹⁶.

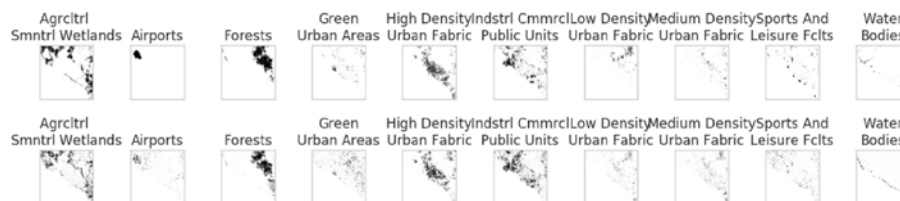
Colaninno (Fig. 6) worked on the automated classification method for distinguish the urban fabrics based on morphological homogeneity of urban contexts and the level of compactness of urban fabrics¹⁷. In order to quantify the morphological homogeneity, it proposed the concept of their own developed criteria, ‘compactness’, from building form characters and building distances relation. In the results, it provides newly defined the urban fabric map of buildings clustered by similar urban contexts.

Zhou (Fig. 7) used geo-tagged images to detect the attributes of several cities for discovering the spatial identity of them. After sampling seven city attributes representing the identity of a city, the research team conducted the city identity recognition experiment using support vector machine (SVM) for analyzing the visual similarity among different cities. As the results of the experiments, it provided the ‘city perception maps’ of several cities, which illustrated the various situations composed of the seven attributes.

Albert (Fig. 8) used the convolutional neural networks (CNN) with the dataset of satellite imagery to identify land use in urban scale. After collecting the satellite images of a target city with several urban environment classes, the researcher trained CNN model to classify the satellite images’ land use. Then he clustered the images vector with the classified labels using t-Distributed Stochastic Neighbor Embedding (t-SNE) algorithms to identify the characteristics of land uses in different cities.

Brandão (Fig. 9) proposed a computational method to generate maps that correlate building permits georeferenced data and their processing time. It used a georeferenced building data with shape file format (.shp) including the tags of building information, such as building types, permission types, construction year, etc. After statistically defining a discriminant for renewal types, it mapped the types of permission by approval time.

Data type and representations are keys to analyze how these five precedents interpret the granularity of city. There are three different types of data: metrices (vector), images (raster), shapes (geometry). In the cases of matrices (A, B, C), the researches handled an individual building as an object that includes features with numeric representation.



16. Jiang et al, p.1

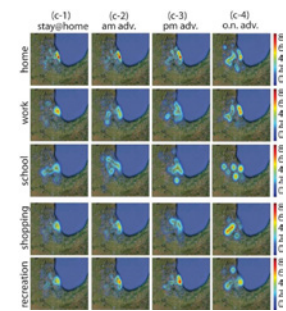


Fig. 5. Research B, Time-cumulative spatial densities of shopping/errand & recreation/entertainment activities in the Chicago sub-region

17. Colaninno et al, p.2

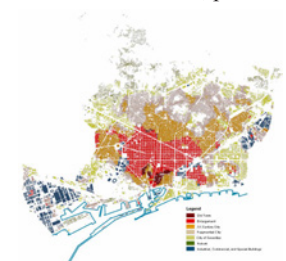


Fig. 6. Research C, Result of Cluster analysis for seven urban typologies

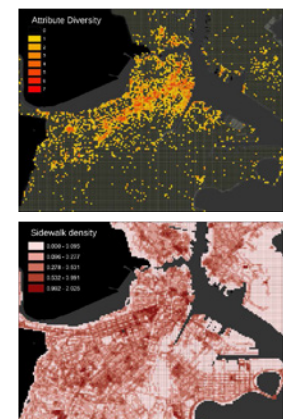


Fig. 7. Research D, The map of attribute diversity and the map of sidewalk density in Boston

Fig. 8. Research E, Barcelona: ground truth (top) and predicted probabilities (bottom)

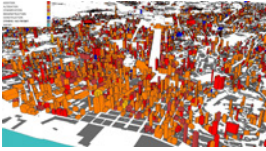


Fig. 9. Research F, Map of distribution and length of permit approval process in the center of Lisbon

The main purpose of the image-based data (D, E) is detecting the implicitly internalized information of a unit-sized image. In the cases of image-based data, researchers assumed that there are patterns which are composed of complex information and can be detected by learning system. The cases of georeferenced shape file type data (F) handled that an individual building as an object that include features like the cases of matrices, but they used geometrical representation. While vector data types are used directly in operations and calculations, and raster data types are used to find complex and implicit relationships.

Through this review, I realized the potential of a new data type that can represent the urban granularity better for calculating and learning systems by combining these two data types. If the new data will be a raster format data generated by the vector information, it will be used to catch the features of individual objects and the relational information at the same time.

2.1.2 Urban Collectivity: Patterns

The six research examples described above illustrate the individual elements of the city as data and tried to seek out specific rules and patterns of the data relating to the urban space. In other words, they used different patterning algorithms by their data types and patterning objectives. In the cases of vector data types, such as matrices (A, B, C) and georeferenced shape file (F), the researches employed statistical algorithms or their own formulas to extract the feature of data without learning systems. Otherwise, in the cases of raster data types, such as images (D, E), the research used learning algorithms to extract the feature of data.

They made the patterns of the data through these extracted features from whole data. With these features, they can distinguish each data and differentiate the data by generating the gradients. This is the process of generating the patterns from numerous individualities. Furthermore, if the patterns can be visualized through remapping on the real-world 2D space, it could provide another layer of information from the pattern, such as relationship between urban space and its change speed. For example, in Brandão's Map of distribution and length of permit approval process in the center of Lisbon (Research F), all buildings can be labeled by the types of renewal. Assuming there are rules or patterns in the data space, without proper visualizing information of the types of renewal on the map, it is be difficult to recognize these rules regarding to urban space. Conversely, by indicating these collective information or patterns on the map, the map will clearly show what speed the area is being renewal.

In examining the six example works above, it can be said that representing the elements of the city as data is a method to express individuality of city and computing the patterns of the data is a method to discover collectivity of city by the data type. This idea triggers the potential that a research can explore the relationship of urban fabrics

based on a target building and its architectural context conditions, unless there is a proper data type containing the relationship between building information and its contexts and proper algorithms to investigate the patterns of this data.

2.2 Context-ness and Contextless-ness

I have included contextual information as a constituent element of urban morphological types, as well as target information that specifies objects. One step further, this section defines the architectural context information, or architectural contextuality, in a broader sense based on the urban morphological types. Besides, this section examines how this architectural contextuality appears in previous researches on architecture and urbanism using AI techniques, and analyzes the tendency of researches in various aspects.

I chose 43 architectural and urban design projects that used AI techniques from 1960s to 2017. The criteria for the use of AI techniques were set more widely than the general criteria. If there is an AI algorithm that generates a design decision for a given problem, rather than a simple automation, it is considered as the case of the AI technology-integrated architectural projects. Then, I investigate several attributes in this precedent: task, algorithm, contextuality, year. 'Task' means the types of architectural problems that the AI techniques are used for. 'Algorithm' means the main computational logics to solve the given problems. 'Contextuality' means whether the algorithm(s) considered the context information or data to make design decisions, such as relationship with other buildings, environmental elements, terrain conditions, etc. If an algorithm handled this context information, the attribute of contextuality is marked as 'Context'. If not, it is marked as 'Contextless'. 'Year' means the year the project was published.

In the following two sub-sections, several example works of 'Context' and 'Contextless' will be introduced.

2.2.1 Contextless-ness cases

CARD¹⁸: CARD is the experimental algorithm devised by N.Cross for the comparison of human and machine performances in solving design problems¹⁹. The main design problem of this algorithm is the room layout. The reason why he set the experiment using room layouts is its properties, showing the limits of the integration of AI techniques with architecture at 1960s. At that time, there is almost no known algorithm for the construction of an optimized solution to such a problem. Instead of searching optimized solution, machine should exhaustively evaluate every possible layout - a procedure which is impossibly lengthy with anything other than a trivially small number of components in the problem²⁰. He set up the scientific environment to this experiment, by defining the comparison and control group, and 'cost' or 'distance' as an optimization parameter.

18. Cross, *The Automated Architect*.

19. Cross, p.74

20. Cross, p.72

21. Stiny and Mitchell,
“The Palladian Grammar.”

22. Stiny and Mitchell, p.1

Fig. 10. The Palladian Grammar, Rule for the generation of exterior walls (Left), The underlying wall pattern for the Villa Malcontenta (Right)

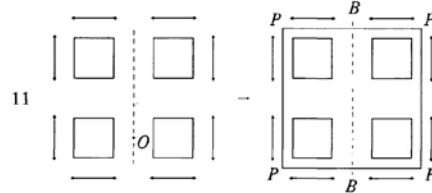


Figure 5. Rule for the generation of exterior walls.

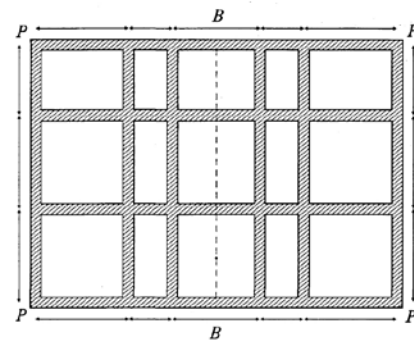


Figure 6. The underlying wall pattern for the Villa Malcontenta.

sites.

The Palladio's architectural system has the great potential for the object of the shape grammar in that Palladio already set up several rules and relationships for his architecture style. However, unlike modern and contemporary architectural design, Palladian architecture did not focus on the relationship with the contexts of a given site. The architects' archetype at the time was Rome style classic which had mainly focused on articulation and proportionality issues, called the order relation. Palladio founded a system for a proportion and order of pillar, the most important elements in classic architecture. His system was so successful to define the relationship among inner spaces, but there is little considerations and interactions to the contexts. Therefore, regarding the context, applying the shape grammar to the Palladian style architectural design does not interact with the contexts.

23. Michalek, et al,
“Architectural Layout
Design Optimization.”

24. Michalek, p.461

Architectural Layout Design Optimization²³: This is example of room layout design using computational optimization algorithms and presents a method for integrating mathematical optimization and subjective decision making during conceptual design.²⁴

It developed own optimization models based on mathematical optimization concepts, such as gradient-based search algorithms, simulated annealing, genetic Algorithms, with the graph and cell representation to make the design decision for the room layout. By scoring the ‘current status’ of the room layout through constraints and objectives of the optimization, it developed a floor plan of given programs of a house.

However, the constraints and objectives are about the internal aspects of buildings. Distance and connection among the rooms are the main information for the optimization algorithms. Although this research did not consider the aspects of contexts in the process of the optimization, it shows the effort to make design decisions directly by combining AI with architectural design in early 2000.

Computer Generated Residential Building Layouts²⁵: This was a relatively recent work on optimizing the floor plan or room layout using a machine learning technique to construct the plan. Particularly, it differs from precedent researches in that it not only presented a graphical representation of the structure of a plan but also generated a stochastic result by learning the graph structure. The meaningful points of this example are not only the usage of learning system to search the proper layout of floor plans but also the 3D relations of the rooms through graph representation. Moreover, in design aspects, the exterior lines of the houses are set by the results of the room layout, while



Fig. 11. Architectural Layout Design Optimization, Sample design topology and final geometry generated by the automated design tool

25. Merrell et al, “Computer-Generated Residential Building Layouts.”



Fig. 12. Computer-generated Residential Building Layout, architectural programs illustrated by a bubble diagram (Left) and a set of floor plans (Right)

most room layout studies, without using multi-agent system, has rectangular shape for exterior because their algorithms work like ‘partitioning’ algorithms.

There are no contextual considerations in this work too. Had they created a cost function for the context as well as an internal relation, they could create a system that responded to contexts such as the site terrain, next door, etc. However, in this work, it is reaffirmed that architectural problems can be relatively clearly resolved with the appropriate representation and AI models for the problem.

26. Negroponte, *The Architecture Machine*

27. Negroponte, *The Architecture Machine*, p.90

2.2.2 Context-ness cases

URBAN5²⁶: URBAN5 is an initial model of space design program-based AI. It was designed to be a self-teaching system. At first it was assumed that the architect-user would have had no previous programming experience. Later, it was further assumed that he had not even read an instruction manual. Thus, URBAN5 would have to teach its own language; learn through teaching, change from learning, and adapt from changing²⁷. The goal of URBAN5 was to study the desirability and feasibility of conversing with a machine about an environmental design project. There are two things that stand out in this project.

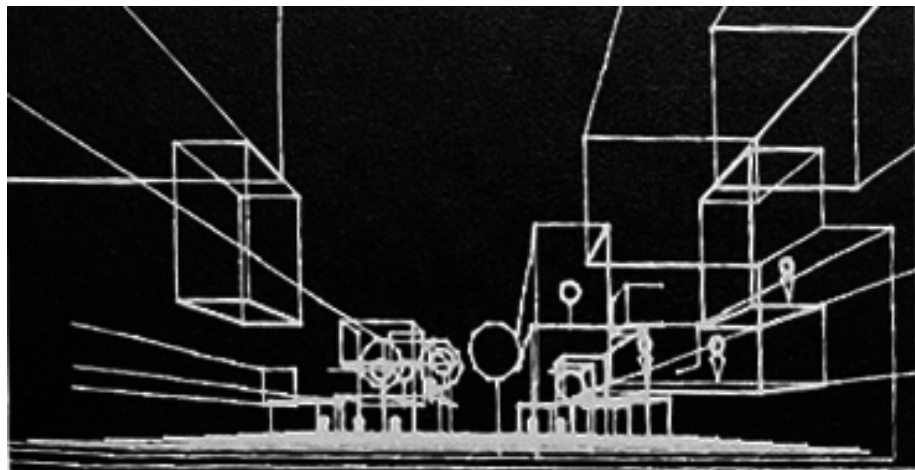
One is to make the computer understand the context of the site. From the interface, considerations on the architectural contexts conspicuously appear. URBAN5 can represent a site plan on a city scale rather than a building scale, and it can also show a representation of the exterior space between buildings. Although the configuration is made up of simple and regular boxes, it shows the possibility to design a space by inputting the relationship between spaces.

The other is how to deal with the quality of space. There were two qualities in URBAN5: implicit and explicit. Implicit qualities refer to the quality of the space, and explicit qualities refers to the quality that the designer intended²⁸. Implicit is an environmental factor such as aspects of sunlight, outdoor access, visual privacy, acoustical privacy, usability, direct access, climate control, natural light, flexibility, structural feasibility²⁹. All these factors are the context information. In URBAN5, implicit qualities are entered into a cube to form a relation with the surrounding cube.

28. Negroponte, *The Architecture Machine*, p.81

29. Negroponte, *The Architecture Machine*, p.81

Fig. 13. URBAN5, A moment of a sequence of frames that depict travel through an environment constructed jointly by the architect and the machine



30. Cross, *The Automated Architect*.

In explicit, the symbols are used to clarify the intention. It can be seen that the quality of the space was transformed into a data model, reflecting the context and computing the relations.

STUNI³⁰: STUNI was a program for computer synthesis of more complex plans in 1960s. Unlike other spatial allocation or configuration programs at that time, STUNI's input data consisted of a relationship values for each pair of departments: it scores plus values of 'internal' relationships for the elements within each department. Therefore, each department is assumed to consist of several elements, all of a standard plan area. This is an attempt to interpret the internal spatial arrangement as a relation and to input the relation into the computer. Furthermore, STUNI had site values are now also added into the computation for the 'cost' of each possible location of each new element. This was a significant attempt to include external relationships as well as internal relationships at computer synthesis, and STUNI was successful to recognize a much greater complexity. Although the interpretation of the quality of space may be insufficient by applying the concept of 'cost', it can be seen that the computer can simultaneously recognize the internal and external relations through a value that the STUNI attempt.

Evolution-Based Generative Design System³¹: This works is about the use of genetic algorithm (GA) to generate the layout drawings of floor plans optimizing to the environmental parameters. L. Caldas tried to apply GA to architecture design process, especially not just parts of the design process, but overall design parts to result feasible design proposals in the realm of an actual piece of high-quality architecture. In term of methodology, she experimented simple GA to building layout, then expanded it to real project, Alvaro Siza's School of Architecture at Oporto, which has clear but complex composition rules. In this experiment, she showed successful implementation of GA to architectural design process, but also the successful accommodation by different geographical locations which has different climatic conditions.

While both 'Architectural Layout Design Optimization' (see 2.2.1) and this research used GA, there are differences in using AI techniques for architectural design considering the context information. L.Caldas investigated for several relationships using environmental factors as determinants in the search phase. This research shows that the environmental factors can be integrated with the AI techniques in the process of architectural design decision as the crucial and parameters for objective functions. But relying too heavily on environmental factors can hinder the considerations on integration AI techniques with the spatial quality from the various relationships.

2.2.3 Contextuality in Architectural Design and AI

A point graph is generated to statistically investigate the research trend in architectural design and AI (Fig. 16). This graph is based on the table in Appendix 1. The table contains several researches that used AI for design process. It shows the researches' tasks and what kinds of algorithms were employed to execute the tasks. In addition, it has a category to distinguish the researches properties as context-ness or contextless-

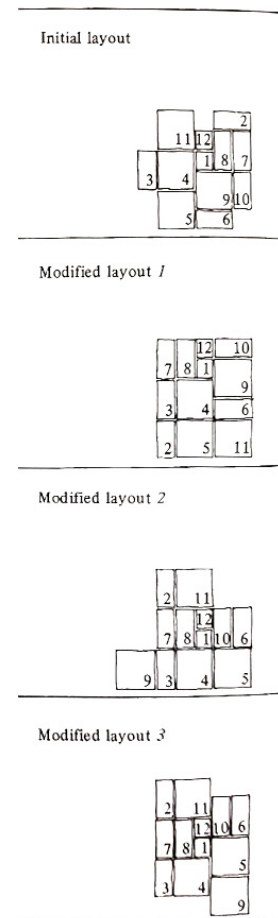


Fig. 14. STUNI, The Sequence of layout modifications

31. Caldas, "An Evolution-Based Generative Design System."

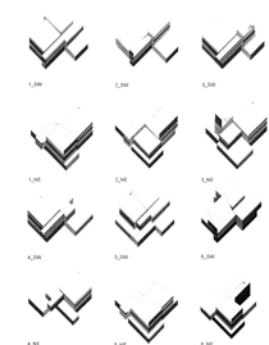
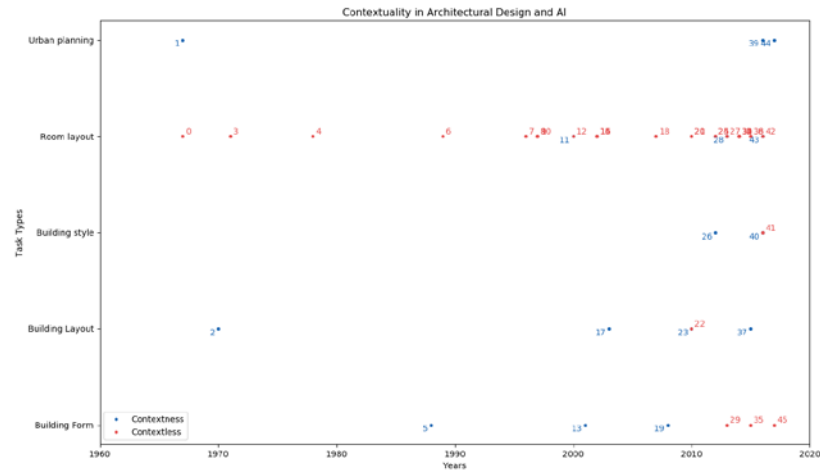


Fig. 15. Evolution-Based Generative Design System, Best solutions for Oporto using EUI as fitness function

Fig. 16. Contextuality of Architectural Design with AI techniques



ness. If design projects in research paper contain the aspect of urbanism, environments, geography, etc., they are marked as 'context' with blue points. While design projects only focused internal aspects of architecture, such as room layout, and no interaction with external aspects, they are treated as 'contextless' with red points.

While the major tasks of 'Context' projects are relatively various, the major task of most 'Contextless' projects is 'Room Layout'. Most projects with the tasks like building form, building layout, building style, and urban planning consider the aspects of contexts and reflect them into design decision. Furthermore, in terms of the research trend, architectural design research with AI techniques has been actively conducted since the mid-2000s.

In this figure, 67.4% of whole projects are contextless cases, mostly focusing on room layout tasks. In other words, most researches used AI for solving room layout without consideration of response to the contexts in architecture.

Through the precedent analysis so far, it has been revealed that using AI with consideration of context in architecture design increases the potential of solving various architectural problems. Using AI with the considerations on architectural contexts can address a broader range of architectural problems beyond room layout issues. In a society with a rapidly advancing AI technologies, if the architects employed AI techs for the perception of architectural contextuality, their relationship will become more intimate.

3 Contexts and AI in Architecture and Urbanism

So far, the city has been interpreted as the data from the viewpoint of urban architecture, and examined the potential that AI can make design decisions through context information. This chapter will cover *diagrammatic representation* which is one of the ways to generate urban or architectural data containing contexts information. Diagrammatic representation in architecture, or diagrams of architecture, is the spatialization of a selective abstraction and/or reduction of a concept or phenomenon³². For the long time, diagrams of architecture have played an important role as a tool to convey and develop architect's idea. There are strong claims have repeatedly been made for the ameliorating and emancipating effects of diagrams, and recent architectural theory and practice in this field continues. Today diagrams are valuable and used in seemingly every aspect of architecture³³. In other words, diagrams of architecture are a familiar tool for architects.

32. Garcia, *The Diagrams of Architecture*, p.18

33. Garcia, p.18

The question then arises as to whether a computer, or machine, can recognize diagrams and use them in architectural design like architects. This question is not about a level of recognizing or decoding diagrams, but about the possibility whether the computer can learn rules explaining complex phenomena through abstract information called diagrams. From this question, the development of new data types begins. Until now, architects have used computers to generate diagrams or to aid the process of drawing diagrams (Fig. 18). Architects have mostly used the computer as an auxiliary tool for drawing diagrams, but they have scarcely used it as a tool for interpreting diagrams. The computer's recognition of diagrams implies the discovery of new knowledge from abstracted images and computer's new interpretation.

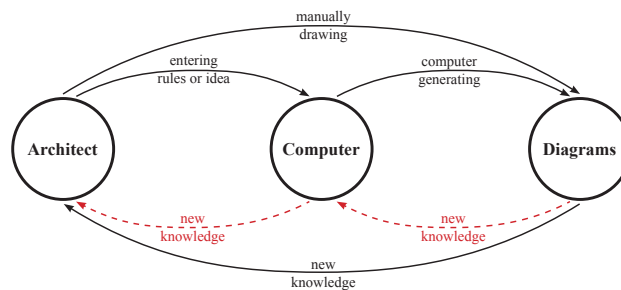


Fig. 17. Relationships among Architect, Computer, and Diagrams

The recognition of a diagram by a computer means that the computer and the architect share the architectural design language and ultimately the new interaction with the machine. This chapter proposes a machine-recognizable urban data format that includes the information of urban elements and their contexts at the same time, and experiments the recognition and usage of the data format in architecture and urbanism realms.

3.1 Representation of Architectural Contexts for Machine Learning

The data of urban elements mentioned in previous chapter includes the target information and context information. This thesis proposes a diagrammatic representation as one of various representation method to express the information. Diagrammatic representation can be employed to use AI technology to learn the situation of the city and use it in the design and analysis process. This section discusses why it should be a diagram or diagrammatic representation as the data format for the computer's recognition of urban situations.

First of all, this section will provide an overview of what the diagrams are, why they are needed, how they have developed historically, and how they are used in contemporary architectural design (3.1.1). Besides, this section will discuss the expected relationships that can be formed from computer's learning of diagrams that architects have used in architectural design aspects (3.1.2). Based on the advantages of diagrammatic representation unlike other data type, it proposes a new data type, called Diagrammatic Image Dataset (DID), and describes how to construct the information of each data (3.1.3). Finally, a general method to generate various DIDs will be introduced, configuring them with a plug-in of a specific application (3.1.4).

3.1.1 Diagrams in Architecture

Attitudes to diagrams in architecture varies. D.G. Shane explains a diagram is a compression of a lot of information about a proposition or argument into a small space, and it is a visual compression that illustrates the essential insight in a single picture, in a single instant³⁴. H. Pai explains a viewpoint of diagrams from scientific management. He said the diagram emerged as a necessary mechanism for the subject to control its object of knowledge in the gap between conception and execution³⁵. Vidler points out the key qualities of the architectural diagram: abstraction, reduction, and geometrical simplicity³⁶. P. Schumacher explains that a diagram is a type of representation that is characterized by a high level of abstraction which works well if what has been abstracted from can be taken for granted³⁷.

P. Eisenman has wider boundaries for diagrams in architecture. As he explained a diagram is a graphic shorthand. However, he insisted that a diagram is not necessarily

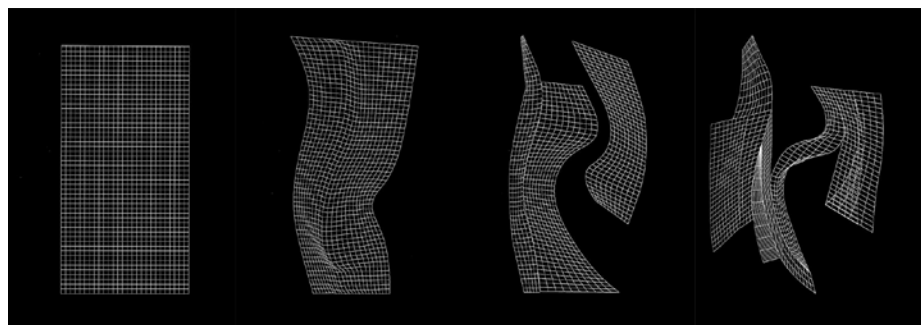
34. Shane, *Recombinant Urbanism*, p.80

35. Pai, *The Portfolio and the Diagram*, p.65

36. Pai, *The Portfolio and the Diagram*, p.54

37. Garcia, *The Diagrams of Architecture*, p.22

Fig. 18. Diagrams as a Generative Device, examples of Staten Island diagrams, Eisenman Architects



an abstraction but representation of something in that it is not the thing itself. Besides, he focused on two ways of understanding the diagrams in history: an analytical device, and a generative device³⁸.

38. Eisenman, *Diagram Diaries*. p.94

The origins of architectural diagrams can be viewed as scientific management, or in prehistoric sites like Stonehenge³⁹ or other visual representations of human history can be the origins of architectural diagrams: constellation maps, earth maps, geometry, drawings, etc. As Eisenman's point of view, the diagram as a graphic that generates information beyond a graphic that simply conveys and delivers information is a modern idea. This is relevant to the debate on images triggered by the development of photography in Western philosophy society. One of the premodern functions of painting, 'record', was weakened by the development and spread of the camera, and painting faced a new situation. At that time, paintings with abstracted graphic style appear. It had not only new graphic styles, but also new function. G. Deleuze described this new function as overcoming the optical bias⁴⁰. When the existing perceived visual information is recorded in different systems and representations, new emotions or information can be generated.

39. Garcia, *The Diagrams of Architecture*, p.19

40. Deleuze, *The Deleuze Reader*.

In this context, two important characteristics can be found among the various angles of architectural diagrams: increasing efficiency of information transmission through compressibility and generating new dimensions of information through abstraction. The new data types for machine to be presented in this thesis should also satisfy these two characteristics.

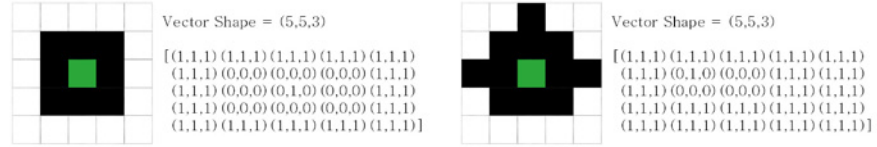
3.1.2 Diagrams for Machine

In modern architecture, various architects have begun to use the diagrams in architectural design. Various creative and diverse diagrams appeared in order to overcome the visual bias from the representative architectural diagram called 'drawing'. In modern times, architectural diagrams have evolved to more accurately abstract many and detailed information. The use of the computer is critical, especially when creating diagrams based on parametric or evolutionary logics.

Architects use computers to draw or create diagrams, reassemble information through the diagrams, and create new knowledge and logic. In other words, the computer was a tool for drawing diagrams, and the architect was the one who provided and understood ideas. Could computers be used more than generative tools in this era when computers become more design intensive? If the computer can recognize and analyze the diagrams, architects can extend the use of the computer as a tool of analysis. Computers as tools for interpreting diagrams will not only explore the knowledge that architects have not found, but will also trigger another layer of knowledges for architects (Fig. 17). If the diagram is to overcome visual bias as G. Deleuze said, computers' recognition of the diagram could overcome cognitive bias.

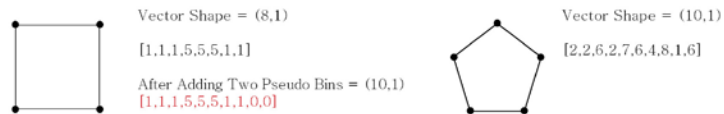
Therefore, this thesis suggests a diagram that the computer can recognize for the new urban analysis. This is a dataset for the machine learning, one of the AI techniques, and computers will be able to learn this type of data. Before generating the computer-recognizable dataset in architecture and urbanism, the type of dataset must be set and decided. There are two main types of the dataset: raster and vector. Raster or raster graphics means images. Otherwise, vector or vector graphics means the metrics of the information.

Fig. 19. Example of Vectorized Value from Raster Images



While both types have their own benefits for statistical approaches of urban analysis, in this research I chose the raster graphics. This is because raster graphics is better to standardize the format of information. When image size is set, an image can represent all containing information by specific modes (e.g. grayscale, RGB, CMYK), since these modes have their own format to represent the value of pixels in the image. For example, a pixel in RGB mode image can be represented by three-dimensional vector like (0-1, 0-1, 0-1). If an image size is 5×5, all pixels can be converted as a vector shaped like (5,5,3), ultimately all different shapes information in images can be converted by the same shapes of the vectors (Fig. 19).

Fig. 20. Example of Vectorized Value from Vector Geometries



However, since the representation of the vector graphics depends on the geometry types, it is not suitable to standardize and express various types of information. For example, there are two geometries (Fig. 20). A geometry is a rectangle in 2-D space which can be represented as four points coordinates and eight-dimension vectors. The other geometry in 2-D space is a pentagon which can be represented as five points coordinates and ten-dimension vectors. If the vector shapes are different for statistical approaches of the urban fabric analysis, there is a limit to accurate data processing, even if the shape of the vector is adjusted through additional processing, such as 'setting empty (pseudo) bins'. But increasing the empty bins of a vector can be an inefficient

representation of the data. For instance, there are 100 polygons in the dataset, and only one polygon is a hexagon and other 99 polygons are squares. Due to only one hexagon other 99 polygons should be represented by a vector with 12 dimensions, instead of 8 dimensions.

A city has a variety of situations, and the situations appears in relation to diverse geometric factors in urban morphology. Therefore, it is more beneficial to create raster images to represent diverse geometric information in a consistent way.

3.1.3 Advantages of the Diagrammatic Image

Diagrammatic Image is an image containing information as a form of diagrams. It can be used as an illustration that shows relationship in architecture and urbanism, as well as in other fields. Using this data set, the patterns or features of relationships that appear in diagrams can be investigated, extracted, and analyzed through statistical methods (e.g. deep learning)

There are two main ways to create the images of urban information: satellite images and map images. However, both methods contain more information than the researchers would expect to see. This can cause more noise than necessary information.

Compared to these methods, diagrammatic image, as a format of data, has two advantages: low noise in the image and synthesis of information.

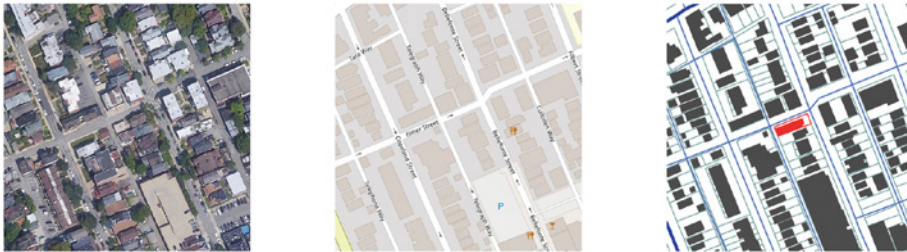


Fig. 21. Three Types of Urban Information Image: Satellite Images, Map Images, and Diagrammatic Images

Low Noise in the Image: diagrammatic images can have lower noise than satellite images. Since the satellite image is the image of the city at a specific point in time, various information such as trees, cars, lanes, materials and shadows are mixed. It is likely to contain unnecessary information. On the other hand, since the diagram image is composed of specific data into an image, it is unlikely to contain unnecessary information.

Synthesis of Information: this is the most valuable advantages of diagrammatic images. Unlike other types of urban information image, users can choose the information to include in the image. Furthermore, when synthesizing the information into an image, users can control the importance of information by making a difference in the graphical expression method to variously represent information. Specifically, it

is possible to create an image reflecting delicate difference by using graphic elements such as color, line thickness, and degree of overlapping.

The composition of information in a diagrammatic image varies depending on the definition of the analysis objects, such urban fabrics. Because the main goal of this thesis is the new analysis through the dataset including relational and context information of urban space and fabrics, the experiments and researches based on the information of physical elements that affect the shape of the urban space, such as building footprints, building heights, parcels, thoroughfares, and neighborhoods. Other non-physical information can be integrated with the diagrammatic images as the format of values, tags, or classes.

41. Jinmo Rhee, Pedro
Alves Louis Veloso, Eddy
Man Kim in Carnegie
Mellon University

In the diagrammatic images, we⁴¹ invented a specific setting for configuring the information and distinguishing the target and context information. A diagrammatic image has a target information placed on the center of the image, such as a target building footprint and a target parcel which is the lot boundary of the target building. The context information, such as context buildings, context parcels, and streets with a hierarchy is surrounding the target information. Due to this configuration, computer can detect the morphological relationship between a target and context information. For example, the **Fig. 22** shows an example of diagrammatic images. It contains information about specific urban patterns surrounding the target building. Neighborhoods can be a label of an image, showing which neighborhoods the target building belongs to. One diagrammatic image dataset (DID) can include the number of diagrammatic images as many as given buildings.

Fig. 22. An Example of
Diagrammatic Image



Diagrammatic images with these two main advantages are suitable for urban fabric analysis in that it can include only the designated morphological elements of city and demonstrate the relationships between the elements constructing the urban structure around the buildings.

3.1.4 Diagrammatic Image Dataset (DID) Generation

Custom components are required to generate the DID with an assigned diagram drawing style in Rhinoceros and Grasshopper. All custom components are written and

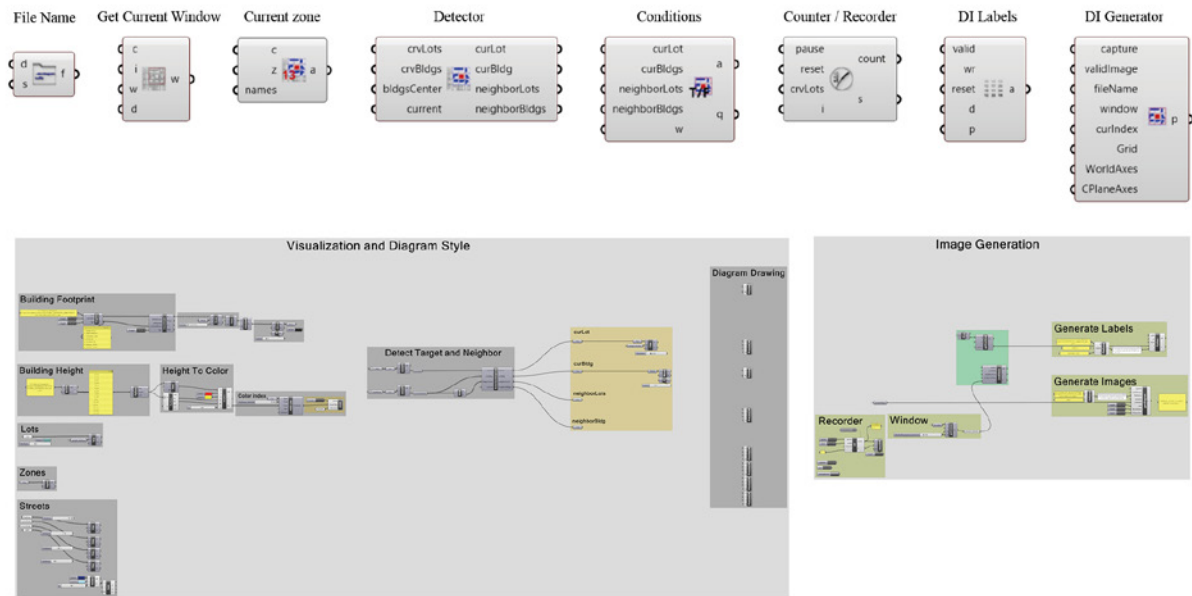


Fig. 23. Custom Components for Generating DID (Top) and The Grasshopper Algorithm for DID Generation (Bottom)

scripted by Python. There are several main custom components: 'File Name', 'Get Current Window', 'Current Zone', 'Detector', 'Conditions', 'Counter/Recorder', 'DI Labels', 'DI Generator'. The 'File Name' component returns the file name given the input directory and suffix. The 'Get Current Window' component returns the square window geometry centered on target building given the input width and height. The 'Current Zone' component detects which neighborhood the target building belongs to and returns the index number of the neighborhood. The 'Detector' component distinguishes the target building's footprint curves and parcel curves from all building and parcel curves. The 'Conditions' component determines whether the target building and parcel can be recorded as an image by given conditions. For example, if a building footprint is too large and extrapolates the image boundaries, it is not considered 'suitable for DID', returning False value. The 'Counter/Recorder' component generates the integer value which could be used as index number of the building or parcel list. It helps to automatically generate DID through whole data of building footprint or parcels. The 'DI Labels' component is a recorder to write a label vector of each image into a .csv file. Users can set their own labels to write. Labels can vary depending on the purpose of the dataset (See the following paragraphs for more information on labels.) The 'DI Generator' component is core components to generate DID. It aligns the center of the current view to the target building or parcel, create a screenshot within the range of the window, and save the screenshot image to given file path. The default image size is 512 × 512 pixels, but users can change the image size.

In addition to generating diagrammatic images, labeling is one of the most important factors in DID generation. Labeling is determined according to the experimental objectives and the machine learning model. Even the unsupervised learning cases

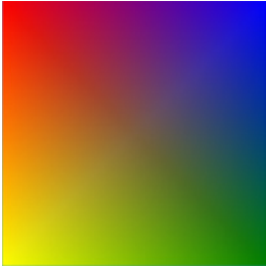


Fig. 24. Four Color Gradients for representing location information

which do not require the labels, users can label them to distinguish and filter data. DIDs from three experiments in this thesis have different labels, such as price, location, footprint geometries, and neighborhoods. One of the good strategies for selecting labeling architecture and urbanism data is to pick data that is difficult to be contained in diagram images and is essential for learning.

For example, location information including the coordinate of longitude and latitude can be considered as a critical factor to the shape of building footprints and the formation of the urban fabrics. The colors of building footprints can represent the location information (Fig. 24). Four corner of the gradient rectangle means the min/max values of longitude and latitude. In the Northern hemisphere, the min longitude and the max latitude of the building footprint have solid red color. The min longitude and the min latitude of it have solid yellow color. The max longitude and the max latitude of it have solid blue color. The max longitude and the min latitude of it have solid green color. Although color can be used to distinguish location information, it is difficult to expect a high learning success rate and accuracy. By specifying the location information as a vector of two values, [longitude, latitude], and labeling each diagram image, accurate vector label data can be provided to improve learning success rate and accuracy. Likewise, rather than simply coloring the building footprint, users can achieve higher accuracy by labeling the RGB (e.g. [255,0,0]) of the color as a vector to represent the location information.

3.2 Recognition of Context-rich Dataset



Fig. 25. Apartment complexes in Seoul,

In this section, I introduce an experiment to investigate implicit building layout principles of South Korean apartment complexes by their price. An apartment complex in South Korea (Fig. 25) is a group of apartment buildings in the form of collective housing. To prevent the negative impact of large and dense occupations on many sectors of the city, the South Korean government provides rigid zoning regulations for these complexes. Designing a complex is similar to applying variations on a prototype in that the designers should address the limits imposed by the regulations and respond to the site conditions to add value. Over the last half-century, the city of Seoul has been the site for prototyping and developing variations of the complexes. In other words, Seoul has a wealth of data that has implicit relationships with the economic value and various physical factors of the complexes.

42. Britanica, “Price | Economics.”

According to the definition of the price⁴², we can translate the price of an apartment building in a complex as a co-positive phenomenon of various factors based on economic, social, and cultural aspects. Therefore, despite of the wealth of data, it is difficult to investigate the effect of design alone on apartment prices and few explicit relationships between their design and prices by specific design rules. However, in order to improve the design of new apartment complexes, designers need to learn these embedded principles

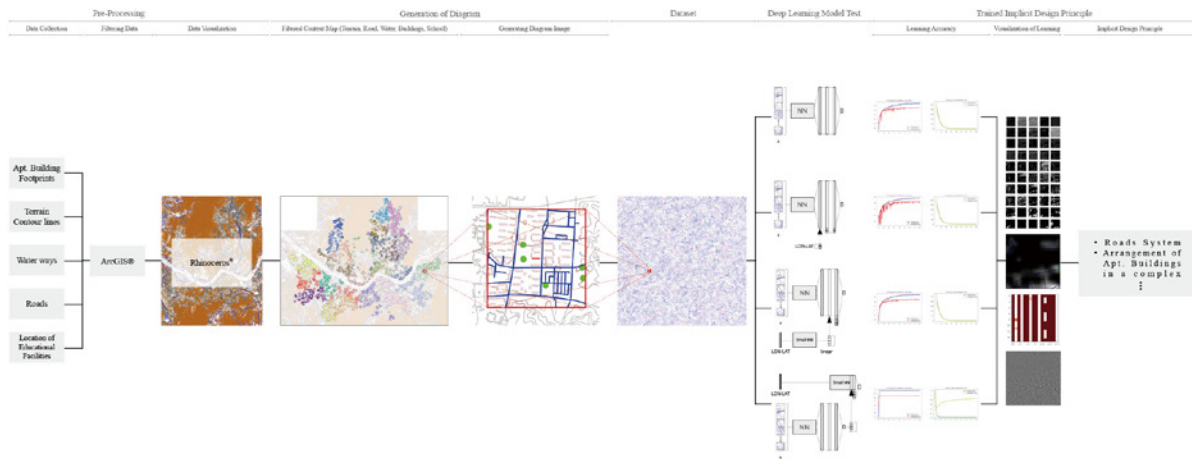


Fig. 26. Overall Process: Recognition of DID Seoul

and how they affect the value of the respective complexes. In our research, we opted for a learning-based system to investigate these implicit and unformalized design principles of the apartment complex layout and its price using diagrammatic image dataset that includes apartment layout information.

In this experiment, through the diagrammatic representation, we expect the computer to find the principles of phenomena that are difficult to formulate by specific rules. This illustrates the use of computers as a new way to discover design principles as well as the discovery of the principles themselves. Furthermore, this experiment will show whether the computer can recognize the complex and implicit relationship through DID.

3.2.1 Overview

As a designer, we can analyze visual media and recognize the recurring design patterns of the complexes. However, due to a large amount of information and the great level of complexity, the relationship between apartment layout patterns and price is hard to be captured by designers. In contrast, deep learning techniques are well-suited for an inquiry into large databases and to learn about these relationships meaningfully. For learning the design principles of apartment complexes' layout with machine learning techniques, two tasks must be preceded: producing a proper dataset and framing the machine learning task.

We opted to synthesize our DID instead of using satellite images for this experiment. The object and scope of this DID is apartment buildings in Seoul for two reasons: an abundance of various examples; and secondly, as the dominant architectural typology of the city (See 3.2.2 for details on how to generate DIDs.) In addition to the diagram image information, building location is needed as supplementary information since both global and local aspects are necessary to investigate the relationship between price and design principles. To distinguish global and local aspects we used location information,

longitude and latitude, and normalized values of the buildings in the same complex. According to the location in the city, the price range of the buildings belonging to an apartment complex is determined. However, in the same complex, prices are differentiated according to local factors such as location and size of the building.

To find the principles of apartment complex design using machine learning, we framed this task as a multiclass classification. The image of the building and its context conditions in our diagrammatic dataset is the input, and a model tries to classify the respective local and global price in a specific number of bins. After training, the model can predict the price bin for new images.

3.2.2 Setup for DID of Seoul

The process of generating DID of Seoul consists of five steps: gathering the data, pre-processing the data, extracting the necessary information, setting the diagram style, generating a diagrammatic image (Fig. 30).

To generate diagrammatic image dataset, digital map file is required, which contains necessary information affecting the price. The digital map provided by the Korean government from the website⁴³ includes all data that we want to use. In our case, we filtered out the apartment buildings and houses that do not form a complex on the map and kept the information of terrain, water, roads, apartment buildings' footprints, and schools on the map⁴⁴.

Regarding the price data, we used '2017 posted prices of Korea Appraisal Board'. This data specifies the posted prices of all buildings in Seoul. The reason for adopting the posted price instead of the actual market price is stability. Since the posted pricing calculated through specific rules, which are legally undisclosed, changes once a year, it is more stable than market prices. The posted price is more appropriate in that this experiment focuses on relative price differences rather than predicting absolute prices.

This data consists of the address, the name of the complexes, the name of the building, the area, the floor, and the posted price. Due to the diagram image extracted from the map without the address name composed of strings, we can use the location information to tag the price to the image. In other words, the location information plays a role like medium linking the price of the building to the building. In order to obtain one of the location information, geocoding with GRS coordinate, which both image and price data can share as common features, we used a geocoding program that converts addresses into specific coordinates⁴⁵.

A reference measure is required to compare prices due to various the number of units and the floor, and the area of an apartment building. We created 'average price per square meter' representing the value of a building in complex. Average price P is the average value of the unit price of one household divided by the area. In the

43. <http://www.nsdi.go.kr/lxmap/index.do>

44. Lee and Ko, "An Analysis of the Factors Affecting Apartment Rent/Price Ratio in Seoul Area.", p.118-120

45. <http://www.iz-gis.com/XRayMap/>

equation below (Fig. 27), p is the price of the i th household, and a is the area of the i th household. n is the total number of households in a building. Besides, because the unit of this average price is very large in billion units, normalization can reduce units. Through the process of normalization, the price domain is changed: the most expensive price is remapped into 1.0, the lowest price is remapped into 0.0. Normalizing not only the price but also the location information can improve the learning success rate through normalization. In this case, the maximum latitude is 1, the lowest latitude is 0, and the maximum hardness is 1, and the lowest hardness is 0.

$$P = \frac{\sum_{i=1}^n \frac{p_i}{a_i}}{n}$$

Fig. 27. Average Price per square meter of a Building in an Apartment Complex

As multiclass classification problems, normalized prices are converted into values of several class. We divide the price intervals, called ‘bins’ and assigning a rate to each interval. For instance, if the values range is divided into the four price intervals, buildings are assigned with the rating of 0,1,2,3. A lower number of rating means lower prices. The rate value can be converted into a vector using one-hot-encoding. We tested 4 bins case, and 8bins case. For better learning, we need to distinguish the global and local price by location: global price is the price differentiated by the location within a city, and local price is the price differentiated by the location within a complex. Training only the surrounding conditions and building price to machine would limit machine’s understanding of apartment complex layout design principles, if building’s location has hugely affected the price. To avoid this, both global and local prices are needed. If the price is divided into eight sections, one building can be represented as a vector with sixteen bins (e.g. [0,1,0,0,0,0,0,0,0,0,0,0,1,0,0]) The first eight bins represent the global price and the last eight bins represent the local price (Fig. 28).

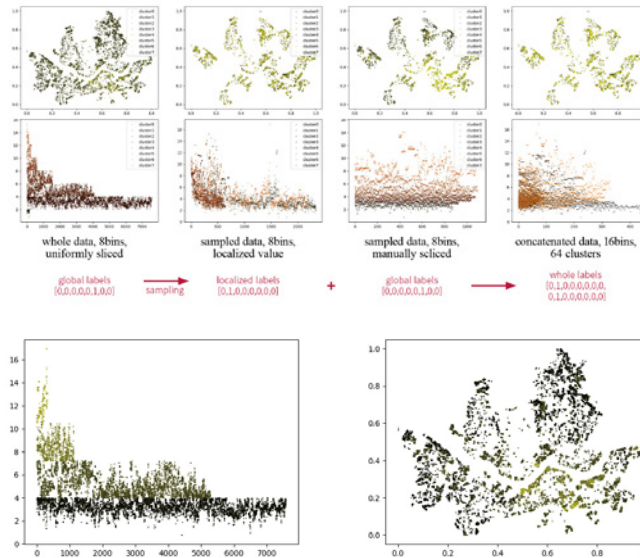


Fig. 28. Example Process of Sampling the data and Global-Local Prices

Fig. 29. Price classification with ‘k-means-clustering’

When preparing the global price vector, the most important thing is sampling the price within the intervals of the price. Each price interval could be divided to have the same



Fig. 30. The process of Generating DID of Seoul

amount of data points, but this would result in different price levels in the adjacent two buildings. Therefore, we set the minimum m as the amount of data points that each class can learn. The price classification is based on the k value satisfying m using ‘k-means-clustering’ (Fig. 29).

After generating the labels, next step is to set the representation style of each element of information. This step is about how to draw a diagram. For example, in our case, a target building, which will be located on the center of the image, is filled with solid red with white background and we fill the neighbor building’s boundaries as pink. We set roads as 10px wide blue lines, water paths as 5px wide light blue lines, terrain contour lines as 2px wide orange lines, and schools as 30px radius green circles.

After setting the representation definition, we set the image size as 512 x 512 px and generated an image file. One building and its context information are synthesized and stored as one image information. Looking through all apartment buildings in the Seoul, and repeating the generation, total 15,284 diagrammatic images are created in this way (See Appendix 2.).

3.2.3 Models and Learning

As the base for our models, we adopt Visual Geometry Group 19 (VGG19)⁴⁶ which is a deep convolutional network for object recognition developed and trained by University of Oxford. We use five different models (Table 2) discretized the predicted values in two different classification settings: four and eight bins.

46. Simonyan and Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition.”

Table. 2. Five Different Types of Models

Model Type	Specification
model A	Multiclass classification with fine-tuned VGG19 model
model B	VGG19 with location information
model C	Ensemble learning of two different neural networks for location and images
model D	Neural network with auxiliary input of price prediction
model E	Multilabel and multiclass classification VGG19

‘model A’ is a customized model to train the diagrammatic image dataset with given number of classification bins. It includes some layers to prevent overfitting, such as batch normalization and dropout. ‘model B’ is the model that concatenate location data as an auxiliary input with the results of the convolution in the fully connected layer.

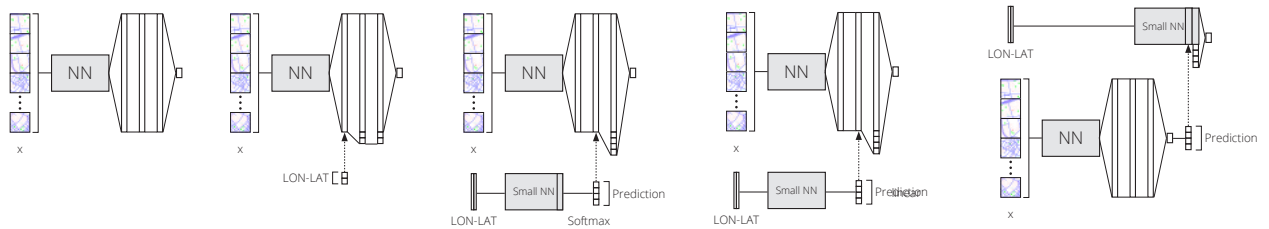


Fig. 31. Structures of the five different models

‘model C’ is a kind of ensemble learning of two different neural networks for location and images respectively. The neural network for images is same with the ‘model A’. Small neural network is a simple neural network which was trained to predict the price based on the location data. The results of those two neural networks are predictions of price. The prediction from the small neural network will be concatenated with the result of the convolution layer in the VGG19. ‘model D’ is a model combining the prediction of VGG 19 into small neural network, as opposed to ‘model C’.

‘model E’ is multilabel and multiclass classification based on VGG19. Basic structure of ‘model E’ is the similar with ‘model A’ except for the number of bins, which contains the global price vector and local price vector. For example, in the 4 bins case, it has 8 bins after one-hot encoding, such as [1,0,0,0,0,0,0,1]. The first four binary number means the rank in global, and the next four binary number means the rank in local. The encoded list says the target building is in the most expensive apartment complex group and the building is in the cheapest group within the buildings in a complex.

3.2.4 Results & Discussion

The final learning accuracy of each model is shown in **Table 3**. The learning accuracy of each model by the epoch is shown in **Fig. 32**.

Model Type	Number of Bins	Train Accuracy (%)	Validation Accuracy (%)
model A	4	94.98	84.07
model A	8	99.25	55.68
model B	4	95.72	80.76
model C	4	95.52	84.73
model D	4	97.98	84.88
model E	8	97.92	87.98

Table 3. Learning Accuracy by Models

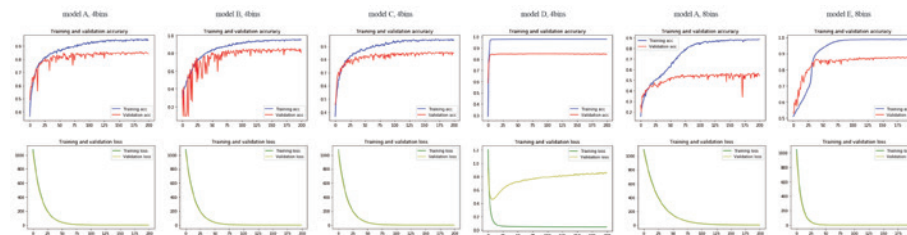


Fig. 32. Learning Accuracy of different models

Fig. 33. Sample image, focus, heatmap, and gradients

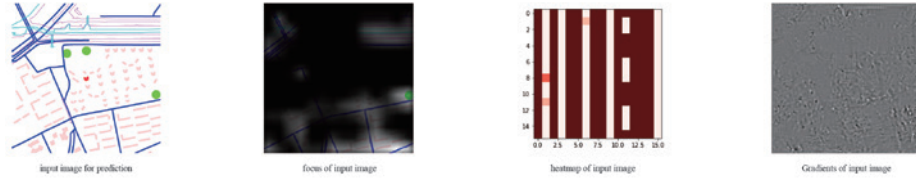
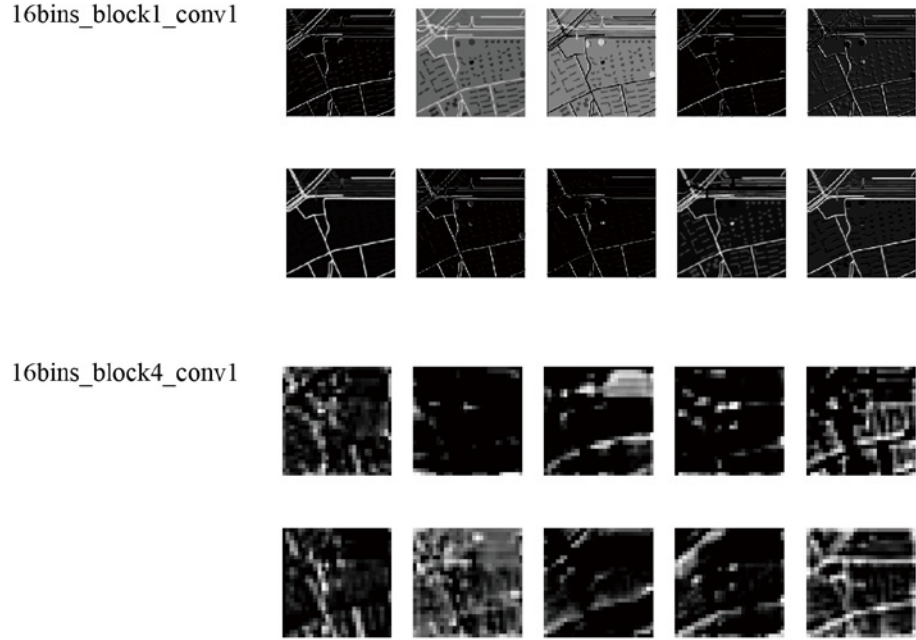


Fig. 34. Visualization of Weights in Convolution Layers



‘Model A’, ‘Model C’, ‘Model E’ are all classified successfully without any huge difference, but their accuracy is not high. The location information slightly increases the success of classification. However, if we increase the number of price interval of ‘model A’, the classification fails with accuracy dropping to about 55%. Also, in ‘model B’, we can notice that the classification accuracy is lowered if the location information is concatenated with result of convolution layers as a feature. The global price and the local price are expressed as a single vector in learning dataset label (model E), the price prediction success rate increases by about 4% as compared with the global price alone (model A).

After successfully training the model, test images are given to the model in order to visualize focus (or self-attention), heatmap, filters of convolution layers, and gradients (Fig. 33). Especially, by visualizing and analyzing the filters of convolution layers, we find that the neural networks keep tracing the pattern of the road systems and the arrangement of buildings. Moreover, the visualized focus of the neural networks shows the bright regions near the road’s geometries on the test images. Furthermore, The visualization of weight in convolution layers (Fig. 34) can show the process how the learning model handled the diagrammatic information. The first convolution layer in the first block of the learning model detects the difference of color and geometries among the included information. Especially, the road system was captured in every step of the

convolution layer. It can be said that road system is distinct information in a diagram. After few processing through more deep convolution layers, the first convolution layer in the fourth block shows more generalized conditions of the given image. It can be confirmed that the relationships between price labels and road arrangement, school location, and complex layout are trained. We can translate that as they are crucial design factors to determine the apartment price.

The results show that the machine succeeded in learning patterns from the diagrammatic image dataset. However, the lack of information in the sample image led the model to be confused after the increase in the number of bins. If more information is added to the image, it would increase in accuracy and prevent overfitting. Furthermore, the influence of color and shape remains as ambiguous. Our hypothesis is that the representation style and composition of the diagram would affect learning accuracy based on the focus image.

For further work, we plan to proceed more experiments on the diagrammatic image composition, color, weight, and shape to investigate how the image structure affects learning. Learning accuracy can be improved by developing certain types of input that combine global and local values. In addition, we would have to generate different image sets from other cities and test their applicability of our method. Although our model succeeded in the classification setting with diagrammatic images, ultimately this research should be framed as a regression problem for more precise relationships between implicit design principles and price.

This experiment is the first experiment that succeeded in learning using the diagram image. This experiment illustrates that diagrammatic image can not only convey the compressed and abstracted information, but also find relationships that were not resolved by several specific rules or formulas through inductive reasoning. The successful recognition of DID using the deep learning model implies that architects and computers can share the same design language. Now, the computer can be a tool for recognizing architectural situations and contexts and giving interpretations to given diagrams. This advance can be one further step to the relationship that architects have admired for several decades, called ‘cooperative partners’ beyond the ‘perfect slaves’⁴⁷.

47. Cardoso Llach, *Builders of the Vision*, p.49-72

3.3 Usability of Context-rich Dataset

In the previous experiment, we have shown that DID can recognize not only target information but also context information to be considered in architectural design. This section shows another experiment to see if these context-rich datasets can be used as design basis. The goal of this experiment is to use context-rich dataset for a simple design decision in a different way from the conventional computer-aided architectural design (CAAD) through the DID as a new medium of information and the learning system as a new frame

that processes information. In this experiment, we are developing an alternative to rule-based expert systems: a model based on learning for investigating the relationship at the urban scale between a site and its architectural context. We use the building footprint as an exemplar of the urban and architectural elements to represent a relationship between design and architectural context. As footprints collectively form a certain pattern in the urban context and directly relate to an individual building design at the same time, they can be a design indicator corresponding to architectural and urban context in a site.

In CAAD, the relation between has been partially addressed in two fronts: space planning and rule-based models. Space planning is concerned with the allocation of activities and definition of building layout. It focuses on well-defined building requirements, such as adjacencies, areas, shapes and minimization of the circulation flows, generally considering the building parcel and site information as constraints^{48 49}. This emphasis on the internal building goals can be a major bottleneck in architectural design cases that rely heavily on architectural and urban context information for generation.

Rule-based models, such as Shape Grammar, L-systems and Cellular Automata, are used to capture the formation logic of pre-existing or desired urban types and settlements^{50 51}. In this case, the elements of the context for design decisions not only can be generated but also embedded in the logic of the generative process, influencing the formation of buildings. However, the bottleneck in these studies is that the rules for the formation are interpreted, formalized and evaluated by an expert, relying exclusively on her access to information and personal biases.

Instead of suggesting a design solution with defined rules or rules that match only specific conditions, we propose a learning-based design system that can adapt to various situations by learning the data that include the target information and the context information extracting the rules and conditions under which the rule applies. In the conventional rule-based design, the computer did not generate rules but was execute the rules. In the system we propose, a computer not only executes rules but also generates rules and applicable conditions. The development of this new system means that the expansion of the computer's role in the realm of architectural design. Through this system the computer can broaden intervention in architectural design. Recalling the architectural milestones in which innovative architectural design trends have occurred in connection with the development of computers in the meantime, another innovative change in architectural design is expected to be brought about by the expansion of computer role.

3.3.1 Overview

Currently, with the availability of GIS data for cities and new learning techniques, it becomes possible to investigate the influence of the context in architecture beyond the limits of previous CAAD approaches. In this sense, our research team⁵³ developed an 'Occupancy Model' to represent the two-dimensional geometrical relationship between

48. William, *Computer-Aided Architectural Design*, p.425-429

49. S Liggett, *Automated Facilities Layout: Past, Present and Future*

50. Ena, "De-Coding Rio de Janeiro's Favelas - Shape Grammar Application as a Contribution to the Debate over the Regularisation of Favelas. The Case of Parque Royal."

51. Parish and Müller, "Procedural Modeling of Cities."

52. Hillier and Hanson, *The Social Logic of Space*.

53. Jinmo Rhee, Pedro Louis Alves Veloso, Ramesh Krishnamurti

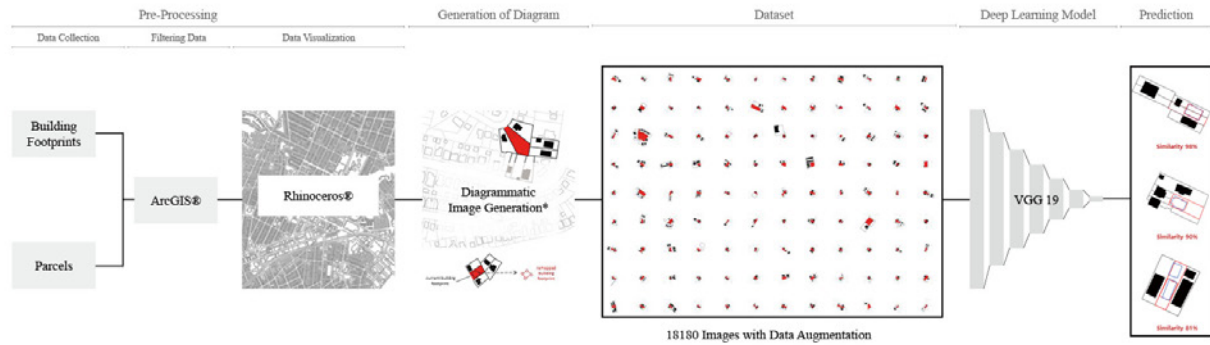


Fig. 35. Overall Process:
Usability of DID Shadyside

a building and its adjacent buildings with their parcels. We use a deep-neural network to predict building footprints, based on a dataset of urban parcels of Pittsburgh.

Fig. 35 show the overall process of this experiment consisting of three main steps: generating dataset, learning, and evaluating the generated footprints. The composition of the dataset is simpler than the previous experiment. A diagrammatic image includes only two types of information: building footprints and parcels. Each building shape information and adjacent building's morphological information in near Shadyside⁵⁴ neighborhood is represented by its own diagrammatic image and labels. At the center of a diagrammatic image, there is an empty parcel filled with solid red which means target parcel information. Around the target parcel, there are geometries representing the information of adjacent buildings and parcels. The footprint information of the target building is tagged to the diagrammatic images as the label. While labeling footprints information to images in DID, footprints geometry information of each building is approximated into rectangle shapes in order to be represented with a constant size of label vectors. Diagrammatic images with these configurations are generated by the number of buildings.

54. One neighborhood
of Pittsburgh, PA

After generating the images, a deep learning model learns building information according to surrounding buildings and site conditions of the target parcels. We frame this experiment as regression problem because the label consisting the information of the target building footprint is a vector of points including x and y coordinates. We adopt VGG 19 model and data augmentation for training the DID. Based on the successful learning, the trained model can generate the footprints on the given and unfigured parcel. A specific measurement, 'Similarity', is developed to compare the generated footprints with the real footprints.

3.3.2 Setup for DID of Shadyside

We use Geographic Information System (GIS) data provided by Allegheny county⁵⁵ to get the information of building footprints and geometries of parcels: '2017 Allegheny County - Building Footprints' and '2017 Allegheny County - Parcels'. After importing

55. One neighborhood
of Pittsburgh, PA <http://www.pasda.psu.edu/uci/SearchResults.aspx?originator=Allegheny+County>

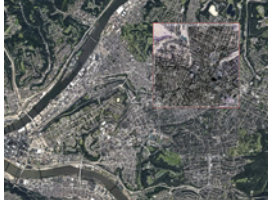


Fig. 36. Sample Site Range

.shp files in ArcGIS, we set an area covering about 1.5km around Pittsburgh at an intersection where several neighborhoods meet near the Shadyside, East Liberty, Friendship, Bloomfield, Garfield, Highland Park History District, and Larimer (Fig. 36). This range of the site includes 7,598 buildings and 8,459 parcels. After converting the building footprints and parcel geometries into .dwg files, Rhinoceros and Grasshopper are used to import those files and generates DID of occupancy model. For the diagrammatic images the process required setting a diagram drawing style by adjusting line weights or colors for the geometry. This style is customized by user intention for purposes such as the research described here. Here, target parcels are filled in solid red, and neighboring buildings filled in solid black. Neighbor parcels are represented 2px wide black lines.

The geometrical information of target building's footprint is a label of each diagrammatic image. The data type of the footprint is polyline which can be

Fig. 37. The Composition of the DID and Geometry Approximation for Constant Bin Size



represented by a set of points which the number varies depending on the shape. of the footprints. This is because a standardized vector format with a constant number of bin size is required. **Fig. 37** shows the example for the need for constant bin size through geometry approximation. One polyline (top) has four points, and this can be converted into a vector with 8 bins. Another polyline (bottom) has the six points, and this can be converted into a vector with 12 bins. Of course, the first vector can have 12 bins with pseudo-bins of 0. However, if the footprint with the most points has 100 bins, the data is distorted a lot (e.g. 92 bins are 0) Therefore, geometrical approximation to quadrangle helps a vector have constant bin size which can contribute the successful learning results.

For the geometrical approximation into a quadrangle, we test seven different types of algorithms: four longest distance, largest area, largest overlapping area, most similar with rectangle-shape, most similar variance, smallest variance, and Delaunay triangulation (Fig. 38). We chose the Delaunay triangulation because it creates a quadrangle successfully even in a concave geometry (Fig. 39).

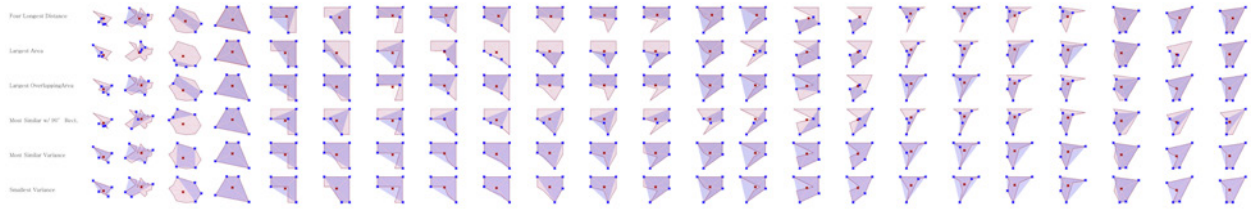


Fig. 38. Test Results of Five Different Geometry Approximation Algorithms

Regarding to the generation of DID, the ‘Conditions’ component checks several conditions. Firstly, the component checks whether the target lot is not empty or not. If it is empty, the component will return ‘False’. In the same way, the case that there are no neighbor buildings near target building, the component returns ‘False’. Other conditions check the collisions between geometries: target parcel and neighbor buildings, target parcel and target buildings, target buildings and the window geometry, target parcel and the window geometry, neighbor buildings and the window geometry, and neighbor parcels and the window geometry. If there are any collision, the component will return ‘False’. The target building size is too small (the area of target building / the area of the target parcel ≤ 0.15) like a temporary storage in yard, it will not be considered as a building. If the ‘Conditions’ component returns ‘False’, this data will not be converted as a diagrammatic image.

3.3.3 Models and Learning

The total number of the diagrammatic image (see Appendix 3.) is 2,080. This number of images are not enough to train the VGG 19 model for regression problem. We assume that the relations between context and footprint in Pittsburgh are weakly correlated with orientation. Therefore, in order to improve learning, we augment the dataset by rotating of sample images. The original building images and labels are rotated 3 degrees apart 8 times into counter-clockwise. In total, 18,720 images derived from 2,080 original images by this image augmentation method.

We compile the model with SGD (Stochastic Gradient Descent) optimizer, MSE (Mean Squared Error) loss function, and 0.00005 learning rate. Batch size is 64 for 800 epochs for learning. We test this model with different dataset by diagram styles and the number of images. With this trained model, un-figured images are given to predict the building footprints considering the given surrounding buildings and parcels.

3.3.4 Results & Discussion

In final learning, the model shows 94.47% of training accuracy and 93.14% validation accuracy (Fig. 41). We remap the predicted footprints on the target parcel and compared the similarity with the original target building by considering the overlapping area and difference of vertices positions. The best predicted footprint has 98% similarity with the original target buildings. However, when there are more than one target building

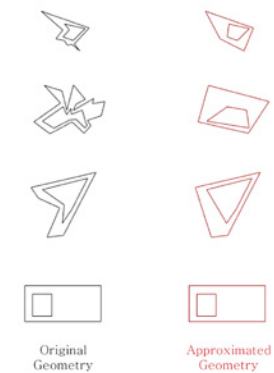


Fig. 39. Examples of geometrical approximation by Delaunay triangulation

Fig. 40. Similarity of Predicted Footprints, 98% (Left), 88% (Middle), 48% (Right)

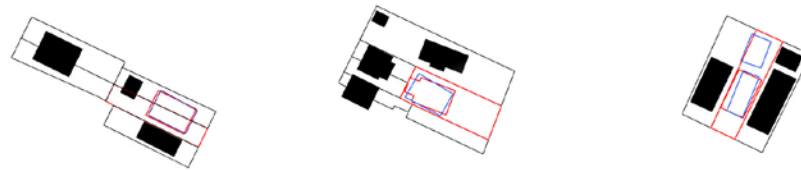
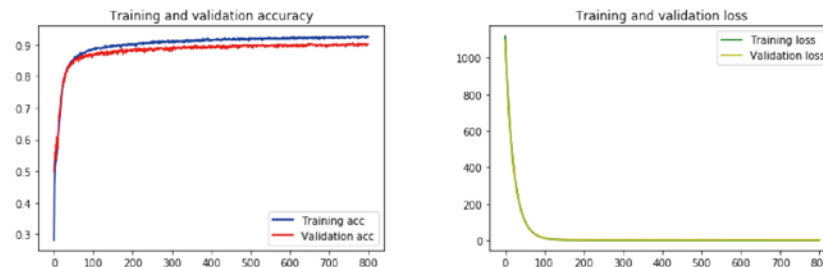


Fig. 41. Training and Validation Accuracy and Loss



in a parcel, it shows only one building footprint with 42% similarity. In additions, if the original footprint shape is complex, it shows a generated footprint with about 60% similarity due to the geometrical approximation process.

The results of this research show that the geometry can be predicted according to a given pattern. This research can provide designers new perspectives on architectural decision making within urban contexts and show the potential of the extended use of technology in architectural design. Moreover, this research demonstrates that deep neural networks can be trained through diagrams rather than satellite images. It means that a model can find specific patterns and features through abstracted images as well as usual images. This study has limitations in that only one learning model is tested and DID includes only the shapes of a small range of urban region. Despite these limitations, this research addresses the potential for designers to employ learning-based systems for architectural design tasks taking into consideration on urban and architectural contexts.

4 Urban Morphological Types: Analysis Architectural Contexts

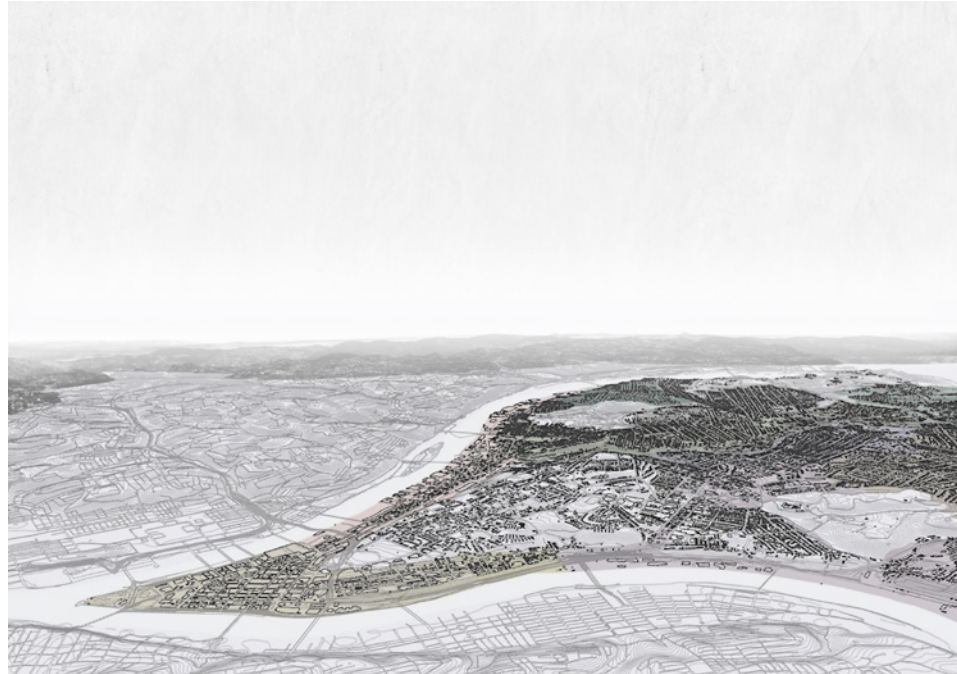
In the previous chapter, we have developed a novel data type, DID, which contains the architectural relationship of the city, and experimented with the recognition and use of this dataset. With the AI models to handle DID, the computer has shown the possibility of a design system that responds to situations by recognizing and patterning complex relationships. Particularly, by not only learning the footprint based on the information of surrounding architectural contexts conditions but also generating a new footprint on an unfigured site, as learning-based system, this system showed the potential that it will be able to overcome several bottlenecks of the rule-based system. Learning-based systems can be a new paradigm in architectural design beyond the existing rule-based system such as parametric architecture.

However, from the viewpoint of architectural design as the complex mental activity, learning system for a small-scale subject is not enough to intervene deeper into the it. This is because the process of architectural design includes materialization of spaces through logical transition from city scale to room or furniture scale. Even a simple design decision in architectural design requires reasons and clues usually derived from the outside space of the given site, not inside space, composed of urban forms such terrain, buildings, streets, etc. Therefore, a large scale understanding of architectural relationships in a city is essential for architectural design. For example, in the previous experiment (see 3.3), it was the miniaturized process of architectural design in that the building outline was created only based on the relationships among buildings nearby. However, in actual architectural design in practice, larger scale information is needed to set the outline of a building. This information is about the urban scale along with architectural scale, such as where it is located in urban space, what type of outside space it has, and so on.

A city is a space consisting of sets of relationships. Urban architecture should eventually be planned with these relationships in mind. As a cooperative partner of architecture, in order to engage deeper in architectural design, computers must be able to find these relationships and make them in the form of knowledge that can be employed in design process. Architects have the know-how to read these relationships with the notion of urban fabrics which enable the morphological understanding of urban space based on the collective relation of individual elements. The computer can recognize these relationships and interpret the urban space through a new definition from the DID with the context information in which the relationships are expressed and the target information representing the object together.

In this section, we try to develop the morphologically new interpretation of a city through DID. In this experiment, we have three expected contributions and results.

Fig. 42. Perspective,
Example of urban
fabrics in Pittsburgh



The first is a new data-driven maps of urban morphological types illustrating their architectural contexts. Assuming that the urban morphological type is a set of urban elements sharing morphological similarity, we generate the DID for this definition. Since the urban morphological types can be generated by clustering the urban space according to the contextual similarities of the individual elements through DID, they contain more information of architectural and urban space typology.

Secondly, this experiment provides a method to phenomenologically recognize the existing urban boundaries by discovering the relationship between the various existing boundaries of cities and the boundaries of urban morphological types. We examine the phenomena that occur within the urban morphological types and at the boundary of the fabrics and compare the difference with the conventional boundary created by the administration and urban infrastructures. Based on these differences, we suggest a different approach to the ambiguous urban boundary problem and new boundaries implying actual phenomena. Different interpretations of urban boundaries can lead different perceptions of cities from conventional one; the city is not made up of several large administrative assemblies but can be perceived as consisting of a large number and set of similar spaces.

Another expected contribution of this experiment is the new urban morphological analysis methodology that can be integrated with other non-morphological information, which can trigger the new understanding of city space. The morphological analysis of the city based on the collective information of the individual building and its surrounding traits is a bottom-up approach of urban fabrics which are different from conventional top-down approaches of other analysis for urban space. This is a foundational method that

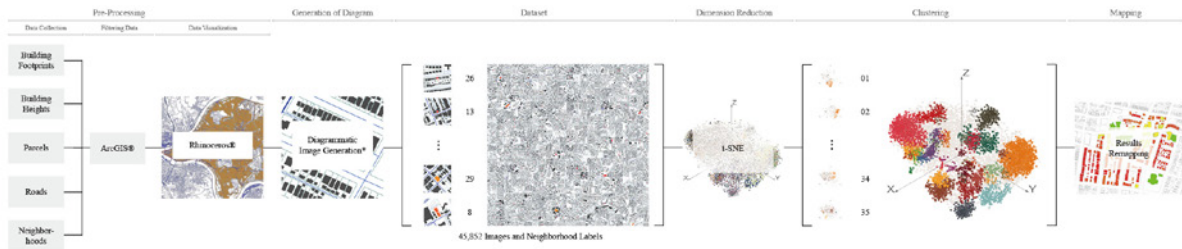


Fig. 43. Overall Process: a data-driven urban fabric map using DID-PGH

can connect urban elements to social, cultural, and economic phenomena, such as price, population, school district, etc., and analyze these phenomena in the aspect of urban space.

Pittsburgh, PA is chosen as an example of this city for case study (Fig. 43). A new type of city data is generated, called DID-PGH (4.1), and urban morphological types are defined in the data space by clustering similar architectural context conditions through a machine learning algorithm (4.2). Based on the clusters, we generate a data-driven map of urban morphological types by remapping the clustered data in the data space on the real-world map. Using this map, we analyze the characteristics of the architectural contexts that each urban morphological type shares. Then, interpreting boundaries of these types as the boundaries of urban fabrics, we sample the several areas where the boundary of urban fabrics are formed based on the map of urban morphological types, and examine the characteristics of what the boundary is, and how it differs from the perception of existing administrative boundaries through the site survey (4.3). In addition, we investigate the conflicts among the different urban fabrics based on the analysis and results of the map and the historical events of urban development to interpret the pattern of the map. This interpretation is a back tracking the result and legitimacy of computer morphological interpretation (4.4).

4.1 Dataset

The **Fig. 43** shows the overall process of generating a data-driven urban fabric map using the DID-PGH. The process of creating DID-PGH has three main steps: pre-processing, generation of diagram, and analysis of dataset's characteristics. In the pre-processing step, there are three detailed steps: data collection, filtering data, data visualization. Data collection is the process of obtaining information about the physical elements that affect urban fabrics. Filtering data is the step of extracting only necessary data from the collected data and matching and synchronizing them to each other. Data visualization is the process of assigning a drawing style to the refined data to draw the diagram. Generation of diagram is the step of generating the diagram filtering the assigned drawing style within the intended scope. Once all these steps are done, a diagram image dataset can be generated. The dataset generated by those steps above has different characteristics from existing GIS data. This section illustrates how DID as the new form of urban data has different characteristics compared to GIS.

4.1.1 Data Pre-Processing

The first step of pre-processing of the DID is to collect the GIS data including shapes of urban elements which will be contained in images with their tags. I mainly use the data provided by the government, Allegheny county (Table 4). Shape file (.shp) is one of the popular file formats in GIS. It contains not only the shape information of urban elements, but also the other numeric and string information by tagging them to the shape information. However, since shape file is not compatible with usual CAD (Computer-Aided Design) programs, such as AutoCAD and Rhinoceros, it needs to be converted as .dwg format through GIS program, such as ArcMap®.

Table. 4. Sources of Data

	Title	Source	Provider	Date	Format
Building Footprints	Allegheny County - Building Footprints	http://www.pasda.psu.edu/uci/SearchResults.aspx?originator=Allegheny+County	Allegheny County	2019	.shp
Parcels	Allegheny County - Parcels	http://www.pasda.psu.edu/uci/SearchResults.aspx?originator=Allegheny+County	Allegheny County	2019	.shp
Neighborhoods	Pittsburgh 2014 CDBG Block Groups	https://data.wprdc.org/dataset/city-of-pittsburgh-gis-data	City of Pittsburgh	2014	.shp
Streets	Pittsburgh	https://cadmapper.com/	Open Street Map	2018	.dwg
Digital Elevation Model	Allegheny County LiDAR 2017 - DEM 6ft	http://www.pasda.psu.edu/uci/SearchResults.aspx?originator=Allegheny+County	Allegheny County	2017	.lyr
LAS Data	Allegheny County LiDAR 2017 - las files	http://www.pasda.psu.edu/uci/SearchResults.aspx?originator=Allegheny+County	Allegheny County	2017	.las

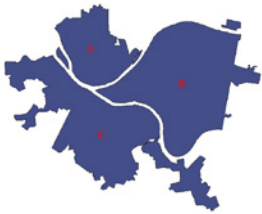


Fig. 44. Perspective, Example of urban fabrics in Pittsburgh

As the range of several data is in Allegheny County, not Pittsburgh City. It is necessary to set proper range of the data by reducing the data range from these regions to Pittsburgh in urban scale. Pittsburgh is made up of three large regions along the three rivers: Ohio river, Allegheny river, and Monongahela river. This research focus on the main region of Pittsburgh city, region B, where contains Downtown, excluding all geometries outside region B in whole data. In this study, Pittsburgh means only the region B of the city.

One of the important information consisting the DID-PGH is building height information. The value of building heights can be obtained through laser survey data, called 'Lidar'. This laser survey data is stored in LAS dataset. A LAS file is an industry-

Fig. 45. Decreased Range of the Data, Building Footprint, Parcel, Neighborhoods, Streets

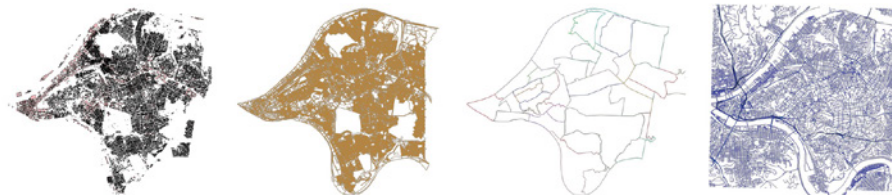
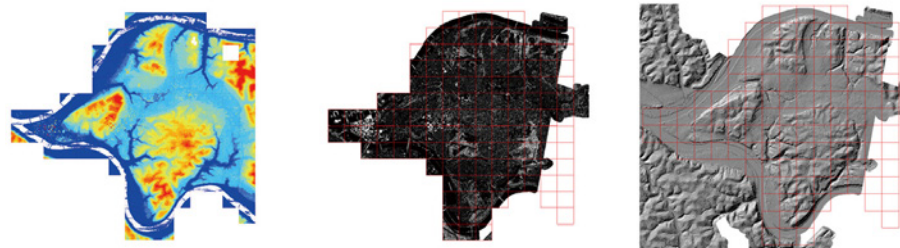


Fig. 46. LAS point clouds of Pittsburgh, DSM of Pittsburgh, DEM of Pittsburgh



standard binary format for storing airborne lidar data, and stores reference information for the location as well as to additional surface features. LAS data of Allegheny County can be downloaded from the website of Pennsylvania Geospatial Data Clearinghouse. The first step is importing the LAS data into ArcMap. The partitioned LAS files (.las) by each tile are imported as point clouds using the geoprocessing tool in ArcMap. An index tile in the northeast part of the Pittsburgh is missing because there are some military facilities. The next step is creating a Digital Surface Model (DSM) image using LAS data. See ArcMap's web page for more information about how to convert LAS into DSM⁵⁶.

56. <http://desktop.arcgis.com/en/arcmap/10.3/manage-data/las-dataset/lidar-solutions-creating-raster-dems-and-dsms-from-large-lidar-point-collections.html>

Once building footprints, DSM, and DEM are imported in ArcMap, the height of the building can be calculated by topological difference between DSM surface and DEM surface within give building boundaries (Fig. 47). Numerous sampling points are randomly generated inside of each building footprint and projected onto the DSM surface and DEM surface. The differences of projected points' Z-coordinate could be the building height. But as the point locations are random within the building footprints, average heights will be used to calculate the building height.

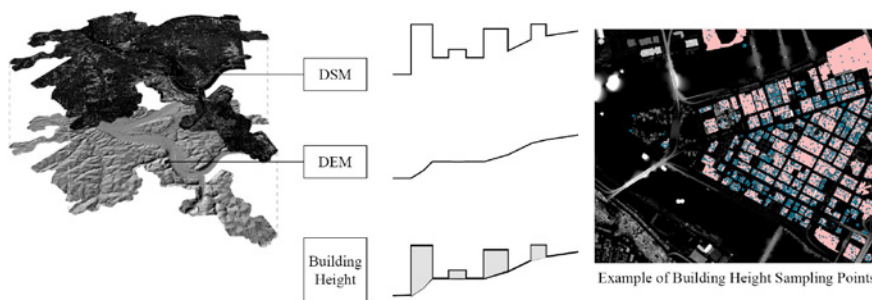


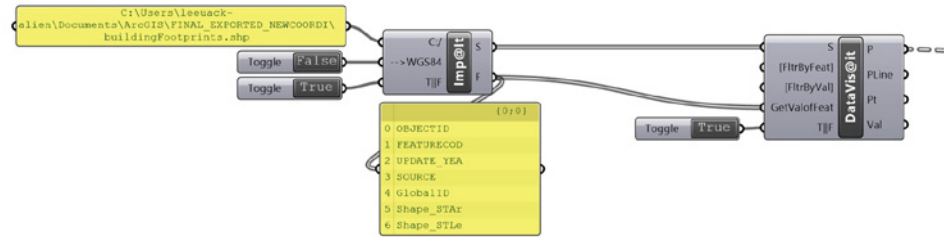
Fig. 47. Process of Building Height Calculation through DSM and DEM

Shape file (.shp) in ArcMap can be imported into Rhinoceros using Grasshopper. A specific plug-in is needed to import shape files based on their coordinates system (Fig. 48). I use '@IT'⁵⁷ plugin to bring the shape file of building footprints into Rhinoceros and visualize them. The first component is 'Im@IT' which has three input parameters: the directory of the shape file ('C:/'), coordinate settings ('WGS84'), and controller of importing ('T||F'). After connecting the directory path of the shape file with the first input parameter, set 'False' value to 'WGS84', because we already converted the coordinates systems of all shape files into 'GS_WGS_1984'. Once input parameters are set, the component shows the 'shapes' and their tags called 'features' in two output parameters: 'S' and 'F'. Each building footprint has a feature which is composed of seven information: OBJECTID, FEATURECOD, UPDATED_YEA, SOURCE, GlobalID, Shape_STAr, Shape_STLe.

57. <https://www.food4rhino.com/app/it>

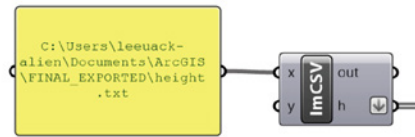
In order to visualize the shapes as Rhinoceros geometry, shapes at a output parameter 'S' of 'Im@IT' component should be connected to the input parameter 'S' of 'DataVis@

Fig. 48. Grasshopper Algorithms to Import the Shape File of Building Footprints



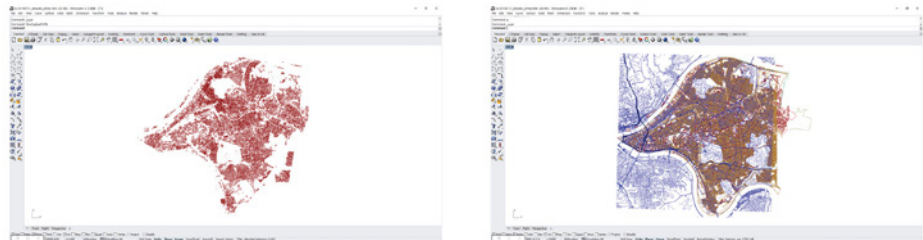
IT', and features at a output parameter 'F' of 'Im@IT' component should be connected to the input parameter 'GetValofFeat' of 'DataVis@IT' (Fig. 48).

Fig. 49. A Custom Grasshopper Component to Read CSV



Next step is to match the building heights to each building footprint. As the building heights data was exported as .csv file from ArcMap, I use my own Python® scripts to read and import .csv file into Rhinoceros tree data structure (Fig. 49). While reading the .csv file, I rounded-up the float numbers of the heights, because the map uses the meter unit and floats of the height values are minor information.

Fig. 50. Imported Shape Files in Rhinoceros Using Grasshopper (Left) and All Imported Geometries in Rhinoceros (Left)



The street, parcel, and neighborhood geometry files are also loaded in the same way as above or they can be converted into .dwg file and directly imported into Rhino without the Grasshopper (Fig. 50).

The next step of preparation for drawing a diagrammatic image is to set the image range of the window. Specifying a range to be called a surrounding context with a building is to set a range of other urban elements related to this building. This range can include a broader architectural contextual relationship, or just an immediate relationship, such as footprints of neighbor buildings. As this research is a study about urban fabrics, the range should be set smaller than city or community scale and larger than adjacent buildings. To get the proper range of the window for our research goal, I calculate the average area of all building footprints (202m²) and square root it (14m). I employ this number as the average radius of the target building because of the

various shapes of buildings. Especially, even if the same area, because there are many rectangular buildings, the average range of one building is 28m. Assuming that about three neighbor buildings are included on one side of the target building, half of the window range is set as 98m and one side of the window is 196m (Fig. 51).

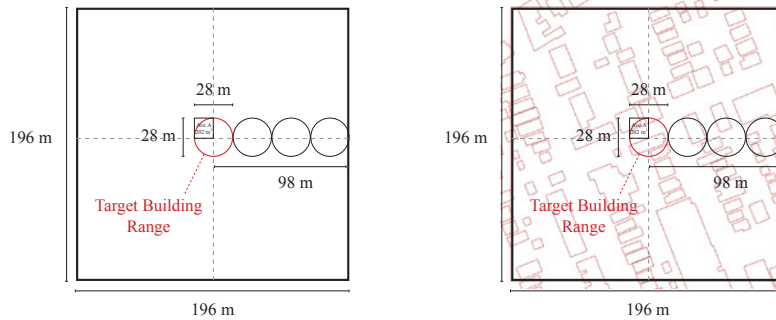


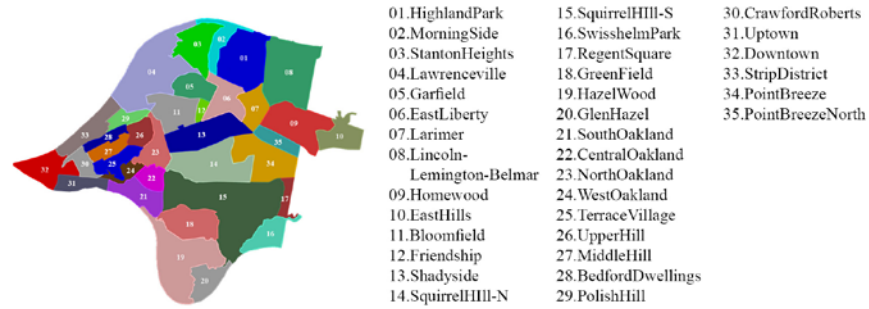
Fig. 51. Process of Setting the Image Range and Sample of Image Range

Drawing diagrammatic images requires the process to set the diagram drawing style. This style can be customized by the user's intention or the purpose of the researches. In this research, after setting the window for the image range, we filled the building footprints which is inside of the window with solid color. Particularly, the target building has a color from the color gradients by its height. The color gradient can be segmented as five different colors to categorize the classes of buildings height: 1m - 12m (max. 3 stories), 12m-25m (max. 7 stories, low-rise building), 25m-75m (max. 20 stories, mid-rise building), 75m-150m (max. 35 stories, high-rise building), 150m - (super high-rise building). We only colored the target building by its height, but when other analyses are needed, all buildings within the window can be colored by their heights. Besides, I set the graphic style for all the line elements such as parcels and streets. I use a 'Custom Preview Line weights' component from the plug-in named 'Human' to express the line color and its thickness: highway as 10 px wide dark blue lines, major roads as 7px wide blue lines, minor roads as 3px wide sky-blue lines, and pedestrian paths as 1px wide light blue lines. Once assigning the diagram style is finished, geometries appear in the window according to the given style.



Fig. 52. An Example of Diagrammatic Image Generation with Customized Drawing Style

Fig. 53. Pittsburgh
Neighborhood
Map and Index



4.1.2 DID of Pittsburgh (DID-PGH)

Regarding to the generation of DID-PGH, the site includes 48,976 buildings, and 64,019 parcels. Algorithm-based generative modeling software, Grasshopper, and the customized plug-ins of Grasshopper mentioned the previous chapter were used. The differences are label settings and the settings of the 'Conditions' component.

The target building's neighborhood stores as a label. There are total 35 neighbors in the data (Fig. 53). Each neighborhood is assigned a unique integer from 1 to 35. During the generating a DID, the 'Current Zone' component automatically detects which neighborhood the target building is belongs to by checking the containment of the center point of the target building footprint with all neighborhoods' boundaries. These values are saved into 'labels.csv' file. As the number of buildings increases, these values are accumulated in the .csv file. Label value can be 0, if the target building does not belong to any given neighborhood. This label system serves as a filter to sort data by neighborhood, and allows users to investigate the characteristics of each neighborhood in the data space.

In this research, we set two conditions. The first condition is to check whether the target parcel is not empty. If there is no building on the target parcel, it is difficult to translate the relationship between the target building and its context conditions. Another condition is to check whether there is an intersection between the target parcel geometry and the window geometry for the image boundary. If there is any intersection, the component will return 'False' because it is hard to detect the boundary of the target parcel.

Under these label settings and conditions, the DI generator looks 48,913 buildings and generates 45,852 diagrammatic images due to 3,061 images which are not satisfying the given conditions, the label file has 45,852 integers which is matching to the diagrammatic images. It takes 49.72 hours to check availability of image generation 48,913 buildings and create 45,852 images with 'Intel(R) Core (TM) i7-8700k @ 3.70GHz' and 64GB memory.

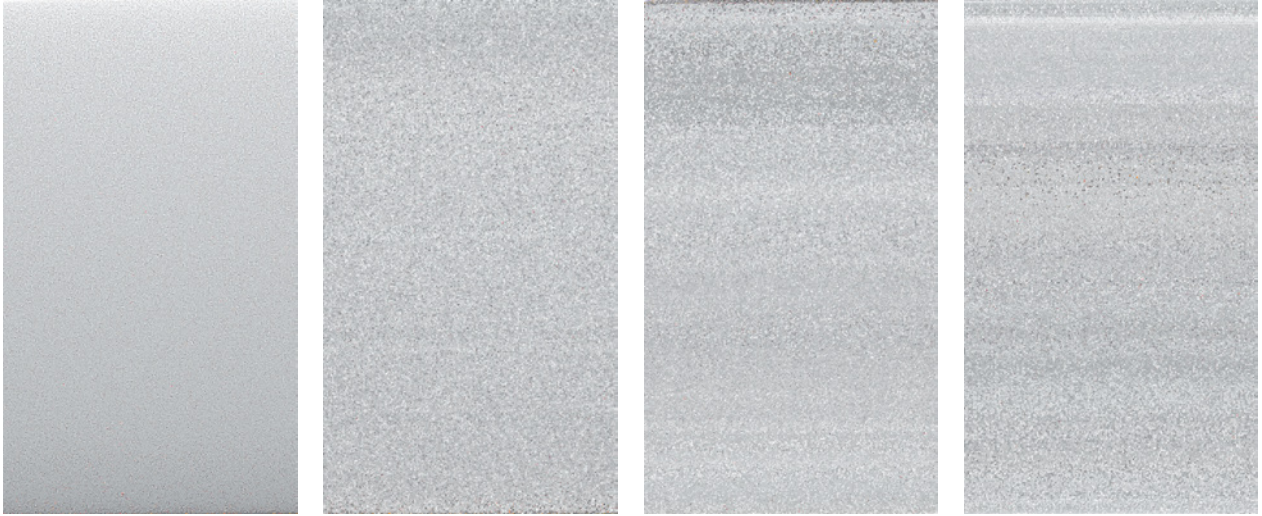


Fig. 54. Four Different Remapping and Arrangements of DID -PGH, By Neighborhood Density, Height, West to East, South to North

4.1.3 Characteristics of DID

The diverse ways to aggregate and arrange the all images representing all buildings and their context conditions in Pittsburgh show patterns of granularity that architects have not seen on regular map (see Appendix 4 for detailed images). The goal of this aggregation is not a simple arrangement of diagrammatic images, but to discover the new understandings of the city changing the perspective on how we interpret the data using AI techniques unlike the way to understand the city so far. The analysis of the city by transforming a city into data containing visual information will provide a chance to look closely at the urban space and collectivity through individuality like these aggregated diagrammatic images.

The aggregated diagrammatic images well show difference between GIS and DID-PGH in terms of approaches of the analysis method: top-down and bottom-up.

The use of GIS is popular to understand the city as a data-based object. The GIS system collects a variety of information and displays it on the map to help users understand new facts form the mapping. This way of analyzing the city is a top-down method. Particularly, when analyzing the urban space, fabric or texture with GIS, the general method is to divide the entire city into a set of patterns based on several specific urban structures. This method is effective for understanding the overall shapes of the entire city with specific themes due to huge scale of the analysis. However, in order to apply this analysis method and information to building or district design, it is necessary to change the scale. This change of the scale can result in distorted application due to arbitrary interpretation of the urban scale information.

DID is the material for a bottom-up approach to analyze the urban structures by focusing a building and its context conditions. This method will help designers minimize distortion when applying urban scale information to smaller-scale designs.

Rather than collectively setting urban patterns, this method finds similarities according to the vectorized characteristics of all buildings in the city and their surroundings. Therefore, unlike a global view, it contains all the situations in which a building is laid out. With this method, when designing space within a specific site, designers can realize what kinds of spaces and characteristics it has throughout the city.

DID differs from GIS mapping in that a diagrammatic image in DID is configured around one building. The dataset contains information describing the relationship between a building and its surrounding context conditions as many as the number of buildings in the city. With DID, it is possible to obtain the information of the urban structures by collecting, classifying, and clustering the information of the individual buildings in the city.

Table 5. Obtainable Information through DID by Scale

Topic	Urban Component	Single Image (Building Scale)	Multiple Images (Urban Scale)
Urban Space	Street Space	Conditions of street space of a building and its surroundings	Tendency and pattern of the relation between buildings and street space
	In-Between Space	Conditions of in-between space of a building and its surroundings	Tendency and pattern of the relation between buildings and in-between space
	Density	Conditions of density of a building and its surroundings	Tendency and pattern of the relation between buildings and density space
	Idle Space	Conditions of idle space of a building and its surroundings	Tendency and pattern of the relation between buildings and idle space space
Occupancy	Occupancy of a Parcel of Land	A building's occupancy of a parcel of land	Tendency and pattern of the buildings' occupancy of a parcel of land
		The occupancy relation of a parcel of land between a building and its surroundings	Tendency and pattern of the buildings' land occupancy relation between a building and its surroundings
	Occupancy of a Parcel of Street	A building's occupancy of a parcel of street	Tendency and pattern of the buildings' occupancy of a parcel of street
		The occupancy relation of a parcel of street between a building and its surroundings	Tendency and pattern of the buildings' street occupancy relation between a building and its surroundings
Types of Building	Shape of Building Footprint	Shape of a building's footprint	Tendency and pattern of the shape of building's footprints
	Area of Building Footprint	Area of a building's footprint	Tendency and pattern of the area of building's footprints
	Volume of a Building	Capacity of a building	Tendency and pattern of the capacity of a building

Moreover, DID can be employed not only to generate the results of descriptive and exploratory analysis, but also to predictive analysis by using deep learning models.

Table 5 shows the obtainable information from DID. The information that can be obtained from a single image and multiple images is different. A single image means

relationships on building scale, and multiple images mean relationships on urban scale. That is, one urban component is interpreted as an urban scale through a large number of images. **Table 6** shows possible urban components through the combinations of urban forms. The characteristics of these relationships are latent in the DID, and finding these characteristics is learning the AI model. Based on this learning the current situation, for the AI mode, it is also possible to predict the changed situation. For example, with DID, when two parcels are merged into one parcel, the effect on surrounding roads can be deduced. This is predictive analysis based on the current descriptive analysis.

	Physical / not Social · Cultural Aspect							Social · Cultural Aspect	
	Target Bldg	Target Parcel	Target Bldg Height	Neighbor Bldgs	Neighbor Parcels	Streets	Neighborhoods	Program	Date
Physical / not Social · Cultural Aspect	Target Bldg	The occupancy of parcel and the main orientation of a Bldg	The capacity of a Bldg	Similarity with context Bldgs, in-between space	The shape of the Bldg, according to the division of the lot	The main street for a building	The shape of Bldg, footprint by neighborhoods	The shape of Bldg, footprint by programs	The shape of Bldg, footprint by dates
	Target Parcel	The occupancy of parcel and the main orientation of a Bldg	The capacity of a parcel	The relationship between a Bldg, and patterns of context Bldgs.	Similarity and collectivity with context parcels	Segmentation and occupancy of the streets	The shape of parcels by neighborhoods	The shape of parcels by programs	The shape of parcels by dates
	Target Bldg Height	The capacity of a Bldg	The capacity of a parcel	The relationship between the Bldg, height and patterns of context Bldgs	The relationship between the Bldg, height and patterns of context parcels	Geometrical and spatial relationship between streets and height	The shape of Bldg, height by neighborhoods	The shape of Bldg, height by programs	The shape of Bldg, height by dates
	Neighbor Bldgs	Similarity with context Bldgs, in-between space	The relationship between a Bldg, and patterns of context Bldgs.	The relationship between the Bldg, height and patterns of context Bldgs	The occupancy of context parcels and the main orientation of Bldgs	The relationship between context Bldgs, and streets	The patterns of context Bldg, by neighborhoods	The patterns of context Bldg, by programs	The patterns of context Bldg, by dates
	Neighbor Parcels	The shape of the Bldg, according to the division of the lot	Similarity and collectivity with context parcels	The relationship between the Bldg, height and patterns of context parcels	The occupancy of context parcels and the main orientation of Bldgs	The relationship between context parcels and streets	The division method of lots by neighborhoods	The division method of lots by programs	The division method of lots by dates
	Streets	The main street for a building	Segmentation and occupancy of the streets	Geometrical and spatial relationship between streets and height	The relationship between context Bldgs, and streets	The relationship between context parcels and streets	The patterns of streets by neighborhoods	The patterns of streets by programs	The patterns of streets by dates
	Neighborhoods	The shape of Bldg, footprint by neighborhoods	The shape of parcels by neighborhoods	The shape of Bldg, height by neighborhoods	The patterns of context Bldgs, by neighborhoods	The division method of lots by neighborhoods	The patterns of streets by neighborhoods	The program distribution by neighborhoods	The completion date distribution by neighborhoods
Social · Cultural Aspect	Program	The shape of Bldg, footprint by programs	The shape of parcels by programs	The shape of Bldg, height by programs	The patterns of context Bldgs, by programs	The division method of lots by programs	The patterns of streets by programs	The program distribution by neighborhoods	The completion date distribution by programs
	Date	The shape of Bldg, footprint by dates	The shape of parcels by dates	The shape of Bldg, height by dates	The patterns of context Bldgs, by dates	The division method of lots by dates	The patterns of streets by dates	The completion date distribution by dates	The program distribution by dates

Table. 6. A Matrix of the Obtainable Information through DID

DID only containing physical aspects of city is suited for analysis of urban fabric and morphology. However, if this dataset would be combined with social and cultural aspects such as building programs, price, permission date, etc., it can derive other valuable results beyond the morphological analysis of the city. Technically, instead of increasing vector dimensions of images, label vectors can include this social and cultural information. Ultimately, this is a combination of various indicators of urban



Fig. 55. The Samples of Dimensionality Reduction Results by Different Algorithms, PCA, VAE, t-SNE

phenomena and urban spatial information, which can interpret them with individual spatial units.

4.2 Clustering Based on Similar Context Conditions

In statistics, machine learning, and information theory, dimensionality reduction or dimension reduction is the process of reducing the number of random variables under consideration by obtaining a set of principal variables⁵⁸. This section shows statistical methods that can detect the similarity of architectural context conditions of a building in a diagrammatic image through the visual information. Dimensionality reduction is employed in order to extract the main features of the architectural context conditions in an image and compute how similar they are to others. In this project, three feature projection techniques that reduce the dimension of DID-PGH are experimented to find the most appropriate methods for keeping principal variables: PCA (Principal Component Analysis), VAE (Variational Auto Encoder), t-SNE (t-distributed Stochastic Neighbor Embedding).

4.2.1 Dimensionality Reduction

Prior to the experiment of different feature projection algorithms, to efficiently run them, decreasing the image size is required because the original image size (512×512 pixels) is huge. I decrease the size of each image into 28×28 pixels. Generally, the higher the image resolution, the better the clustering results will be, because it contains more detailed information. However, in terms of clustering the DID-PGH, not always the higher resolution image dataset can result in the better clustering. This is because some information of pixels in the high-resolution images can be treated as a noise and disturbs the feature extraction.

PCA is to reduce the dimensionality of a data set in which there are a large number of interrelated variables, while retaining as much as possible of the variation present in the dataset. This reduction is achieved by transforming to a new set of variables, the principal components, which are uncorrelated, and which are ordered so that the first

58. Roweis and Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding."

few retain most of the variation present in all of the original variables⁵⁹. Although I test PCA for extracting the features of architectural context conditions in diagrammatic images and mapping the extracted features to 3D space, it is not suitable for finding and projecting major features of the dataset with high complexity such as DID-PGH. This is because PCA uses linear transformation, which is not proper to the data with non-linear patterns. As a result of the PCA calculation with DID-PGH dataset, it is difficult to find clusters that exhibit a certain tendency, despite of few clusters with vague and linear borders.

VAE models inherit autoencoder architecture, but make strong assumptions concerning the distribution of latent variables. They use variational approach for latent representation learning, which results in an additional loss component and specific training algorithm called Stochastic Gradient Variational Bayes (SGVB)⁶⁰. Regarding to the results of the dimensionality reduction with VAE model, it does not grasp the principle features of the DID-PGH dataset well. The **Fig. 55** shows the results of the VAE in 3D space. Each point is the location of an image's principal components in projected space, and ultimately means the features of their architectural context conditions. If two points are close enough to be categorized into a cluster, they are sharing similar architectural context conditions. In the **Fig. 55**, the results of VAE illustrates a better distribution of features than PCA. The sample of the latent space from VAE with DID-PGH (**Fig. 56**) shows that the VAE model can distinguish the context conditions in the images from the target building on the center of the images. However, there is no distinct pattern of clustering the data points that are still concentrated in the middle of the data space.

t-SNE is a tool to visualize high-dimensional data. It converts similarities between data points to joint probabilities and tries to minimize the Kullback-Leibler divergence between the joint probabilities of the low-dimensional embedding and the high-dimensional data. t-SNE has a cost function that is not convex, i.e. with different initializations we can get different results⁶¹. In this experiments, t-SNE algorithm is established and visualized in Tensorboard by using embedding projector. The hyper parameter setting is as follows: perplexity 35, learning rate 100, iterations 1500. Like the results of VAE, a data point in the reduced dimension space means the building, the location of the point means its characteristics of the building footprint and architectural context conditions, and the distance among the points means the degree of the similarity. The shorter the distance, the more similar context conditions.

The visualization of t-SNE results (**Fig. 57**) shows several clusters despite of the vague their borders. Compared to the t-SNE results with MNIST dataset⁶², it does not show the neat boundaries of the cluster. However, the difference of neatness of boundaries is a phenomenon caused by the characteristics of the dataset rather than the problem of the algorithm. Although, in DID-PGH, there are data that have morphologically distinct

59. Jolliffe, *Principal Component Analysis*.

60. Kingma and Welling, "Auto-Encoding Variational Bayes."

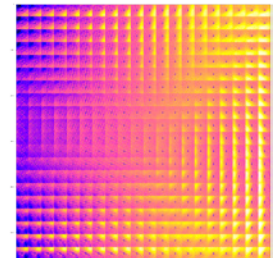


Fig. 56. The latent space of VAE from DID-PGH

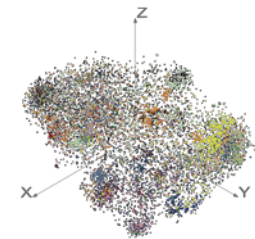


Fig. 57. t-SNE visualization of DID-PGH based on the Similarity of Architectural Context Conditions by Neighborhoods

61. <https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html>

62. The MNIST database of handwritten digits, available from this page, has a training set of 60,000 examples, and a test set of 10,000 examples. It is a subset of a larger set available from NIST. The digits have been size-normalized and centered in a fixed-size image, <http://yann.lecun.com/exdb/mnist/>

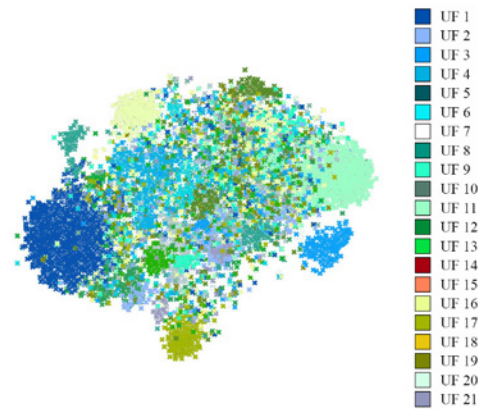
characteristics of contexts, there are also many data that are difficult to distinguish the characteristics of contexts. This data is placed between the clusters in the data space. However, t-SNE is successful to extract the principle components of images in DID-PGH and calculate the similarity of the architectural context conditions among them in that data points representing buildings in the same neighborhood are clustered in the same cluster.

4.2.2 Clustering by Similar Context Conditions

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm is used to find core samples of high density from the distributed points⁶³. A map describes the similarity of the architectural context conditions in Pittsburgh is created by mapping the results of the clustering.

63. <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html>

Fig. 58. t-SNE visualization clustered by DBSCAN



The visualization of t-SNE results (Fig. 57) only shows the spread and distribution of the data points of the images comprising the architectural context conditions. After applying DBSCAN, the data points distribution model illustrates the similarity of the contexts by the distance of the data points. The hyperparameter of DBSCAN is as follows: Epsilon 1000 (10×100), minimum number of samples 15, leaf size 30. The data on the normalized data space from 0 to 1 is imported to Rhino using Grasshopper. After increasing the scale of the data space 100 times to get better visualization for the numerous datapoints, Grasshopper performed the Scikit-learn's DBSCAN. Therefore, the value of epsilon is big and scaled too. According to the results of the DBSCAN (Fig. 58), in data space, Pittsburgh has 21 different distinct clusters that share the morphologically similar architectural context conditions.

4.3 Data-driven Map of Urban Morphological Types (DMUT)

The previous chapter shows the clustering results of buildings that shares the similar context conditions in data space. From this clustering results, two types of urban fabric map

can be derived: data-driven map of urban morphological types (DMUMT) by buildings and borders. I remap the clustering results on the real-world map by coloring each building with the clustering labels from the data space. This is the process of space conversion; from the 3D data space to the 2D world space. The map generated by remapping visualize which buildings share similar urban spatial structures in a city. By the algorithmic definition of the cluster boundary, another type of DMUMT illustrating the border and the area of the urban fabrics can be generated from the clustering information for buildings. Setting the boundaries of the urban fabrics helps users know the characteristics of inside the fabrics and the collision points among the fabrics. By observing the inside the fabrics, users can understand the characteristics of the urban structure. By observing the boundary where the collision occurs, users can infer the change of the urban fabrics.

This section investigates the spatial characteristic of 21 different and distinct urban morphological types of Pittsburgh based on these two types of map. This methodology can be applied to not only Pittsburgh but also other cities. Using the same algorithms and data generation settings, users can investigate the urban structure of other cities. Moreover, the maps from this method can be used as the measurement or indicator for identifying the similarity among cities through morphological comparison.

4.3.1 Map of Building Clusters and Urban Types' Boundaries

By coloring the footprints with different colors, a map containing the building footprints by the clusters from the reduced dimension space using t-SNE and DBSCAN can be generated. This remapping is an iterative process requiring various types of data: shape files of building footprints and their indices that was the base data of DID-PGH, points data of t-SNE results, clustering label of DBSCAN, and color set. After obtaining the clustering label of each building, the building color is selected by converting the label to a corresponding color in the color palette⁶⁴. Rhinoceros and Grasshopper is used as the integrated modeling platform, importing .shp file with its label files. Creating a custom component to make hatches following the building footprints polyline with the color labels, I generate the map illustrating the distribution of the clusters in the real-

64. To generate the 21 distinct colors, I used “I want hue (<http://tools.medialab.sciences-po.fr/iwanthue/>)” which is a tool to generate color palette based on k-means clustering.

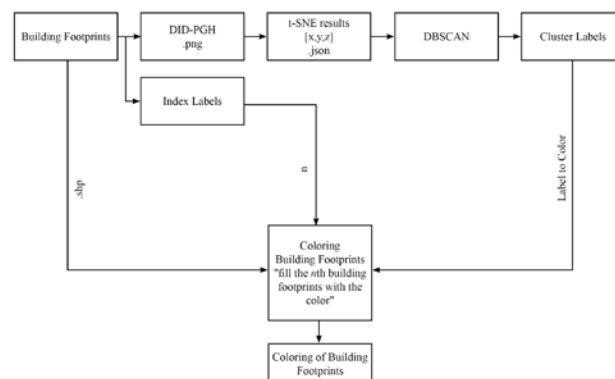


Fig. 59. Process of Remapping and Coloring the Building Footprints

Fig. 60. Data-driven
Urban Fabrics Map,
Building Footprints

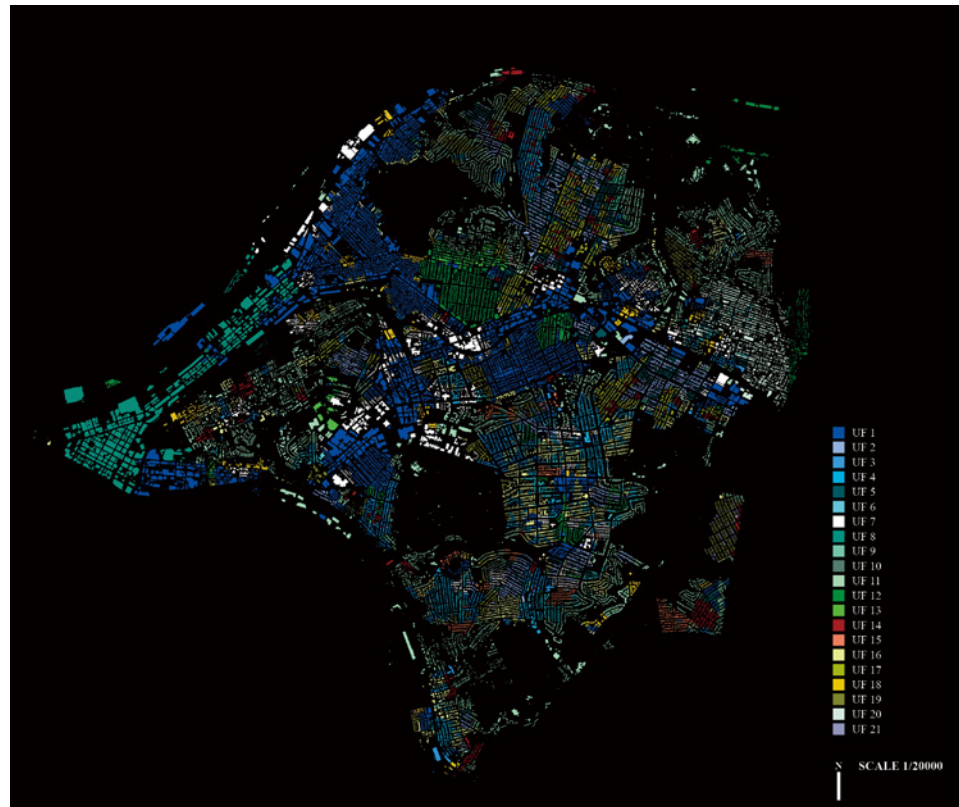
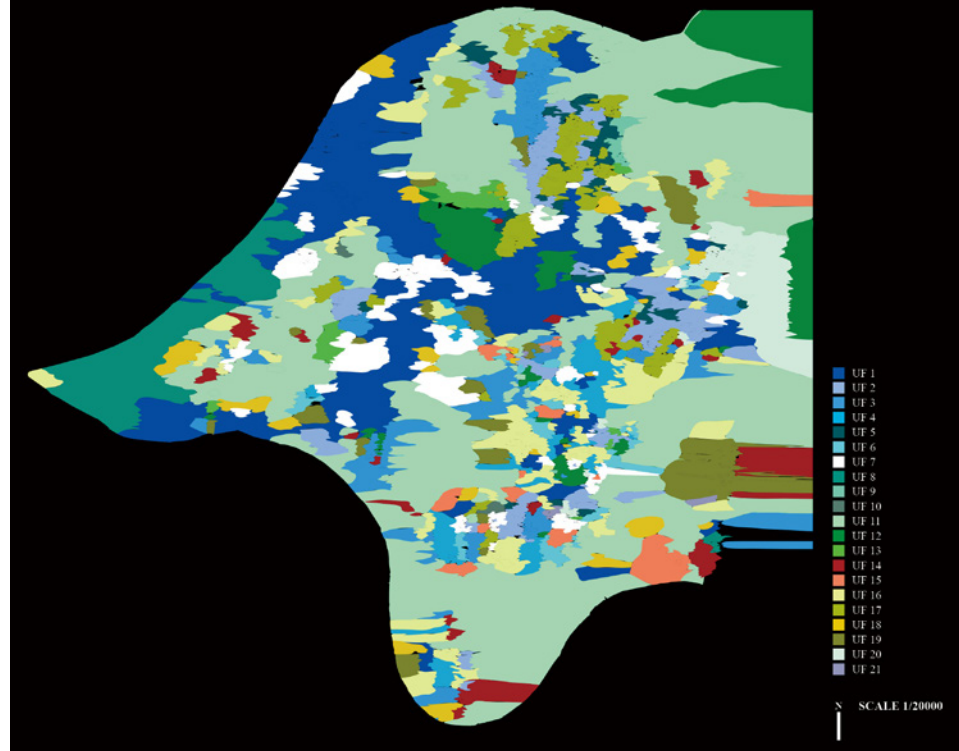


Fig. 61. Data-driven Urban
Fabric Map, Boundaries



world 2D space for buildings (Fig. 60).

After marking all buildings in the map with different colors by clusters in data space, boundaries are generated through KNN (k-nearest Neighbors) algorithms. First, we set the data point to be used as the data KNN by extracting the center point of each building footprint. Besides, we create the label file from the clustering labels for KNN. The hyperparameter of KNN is as follows: algorithm 'Auto', leaf size 30, metric 'minkowski', metric parameters None, number of jobs '1', number of neighbors 3, power parameter for the Minkowski metric 2, weights 'uniform'. Original boundaries of the KNN were fit to the rectangular windows. We trimmed out the original boundaries with the outline of Pittsburgh to create the DMUFMT for boundaries (Fig. 61).

DMUMT for boundaries provide the more distinct shapes of urban fabrics than DMUMT for footprints. DMUMT for footprints shows the information about which clusters the building is in and which buildings shares similar architectural context conditions in urban spaces. Otherwise DMUMT for boundaries shows the patterns of urban fabrics throughout the whole city, efficiently visualizing the tendency and morphological features of each urban fabric in a city.

4.3.2 Morphological Analysis by Clusters

With DMUMT, typical context conditions of new urban fabrics are sampled to examine the properties of urban clusters composed of 21 clusters. 3D models of typical context conditions of 21 different urban morphological types can show spatial features and information. In order to investigate the distinctive features of each urban morphological type, I used spatial quality descriptions and several statistical values: occupancy rate (%), occupancy rank, average height (m), average area (m2), average density, average building occupancy, and average stories.



Fig. 62. Statistics on morphological characteristics by cluster

'Occupancy Rate' is the relative measurement that how much a cluster cover the city area. 'Occupancy Rank' is the rank of the occupancy rate. 'Average Height' is the mean

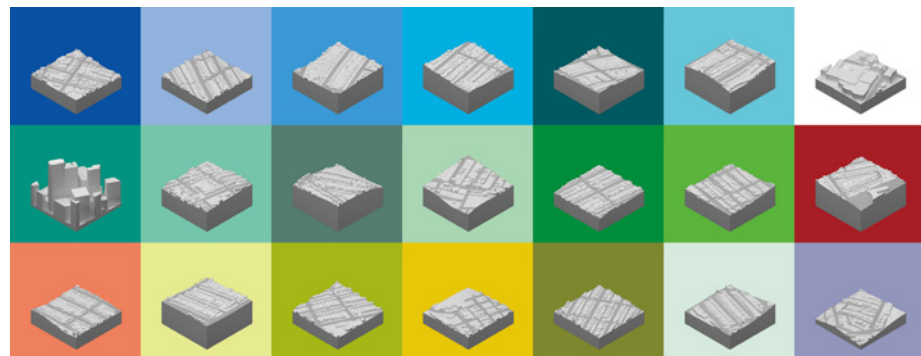
of the building's heights in a cluster. 'Average Area' is the mean of the building's areas in a cluster. 'Average Density' is the mean of the building's density in a cluster. Density is the floor area of the building multiplied by the total height divided by the floor height (3m). 'Average Building Occupancy' is the average of the building occupancy which is the ratio between the building floor and parcel area, 'Average Stories' is the average height divided by the floor height (3m).

From now on, in this section, for the sake of simplicity, the urban morphological type is regarded as a set of urban fabrics. Statistically, 'Urban Type 14' or 'Urban Fabric 14', which presents places where a certain number of empty parcels exists in neighbor and no distinctive shape of blocks, is the most popular type of urban fabrics. The next popular type of urban fabric is 'Urban Type 1'. Urban Type 1 is a street-centered and orthogonal intersecting space with different sizes of urban blocks. This is mainly distributed on the Lawrenceville, Bloomfield, Shadyside, and Oakland. Vertically distinct urban fabric is 'Urban Type 8'. This urban fabric consists of typical blocks with grid systems. This is mainly distributed on the Downtown and Strip District. Urban fabrics with a relatively large building area are 'Urban Type 13' and 'Urban Type 15'. They are mainly distributed on the Garfield, Northern Squirrel Hill, Green Field, and Swisshelm Park. Urban Type 6 has the highest level of land occupation. It can be interpreted that this fabric has the least outside space. When 'Average Area', 'Average Density', and 'Average Building Occupancy' are interpreted together, it can be seen that 'Urban Type 13' and 'Urban Type 21' have the highest density of buildings with the high level of FAR (Floor-Area Ratio).

In addition to quantitative surveys, 3D models are created by selecting sites that represent each urban fabric to conduct qualitative surveys. I sample the center of the largest fabric within a cluster. Two maps, DMUMT for building footprints and boundaries, can be divided into clusters, and 3D models illustrating typical conditions of the clusters can be used to describe the characteristics of each urban fabrics. (See Appendix 5. for details.)

Urban Type 1. This urban fabric is a street-centered and orthogonal intersecting space

Fig. 63. Typical Conditions of 21 Different Urban Fabrics from Clusters



with different sizes of urban blocks. This is mainly distributed on the Lawrenceville, Bloomfield, Shadyside, and Oakland. This urban type contains different types of streets by their width with hierarchy. There are various sizes of buildings along the main streets, and on the hinterland, there is a series of typical buildings along the minor streets with little side gap.

Urban Type 2. This urban fabric consists of blocks parallel to the streets. This is mainly distributed on the Highland Park Historic District, Point Breeze, and Upper Hill. The blocks in this urban type have rectangle shapes along the main streets and two rows of parcels. A road with a lower hierarchy than the road surrounding the block crosses a block and splits two rows of parcels.

Urban Type 3. This urban type consists of blocks parallel to the streets. This is mainly distributed on the Squirrel Hill, Oakland Historic District, and Green Field. The blocks in this urban type have elongated rectangle shapes along the main streets and two rows of parcels. Instead of a street that traverses a block, it has a mirrored building arrangement referenced to the block's center axis. The back yards of the buildings in a block of this urban type are facing each other.

Urban Type 4. This urban type consists of blocks parallel to the streets. This is mainly distributed on the Squirrel Hill and Green Field. The blocks in this urban type are U-shaped. Like blocks in other urban fabrics, they consist of two rows of parcels, but they tend to form dead ends inside of themselves. In addition, due to these dead ends, the depth of the block is wider than other blocks. Therefore, the side of the block also has a series of buildings.

Urban Type 5. This urban type consists of blocks parallel to the streets. This is mainly distributed on the Highland Park and Point Breeze. The blocks in this urban type are a parallelogram shape due to the shape of the surrounding streets and terrains. The blocks in this urban type has two rows of parcels and some of them has a series of buildings along the side of them. The back yards of the buildings in a block of this urban type are facing each other.

Urban Type 6. This urban type consists of blocks parallel to the streets. This is mainly distributed on the Squirrel Hill. The amount of this urban type throughout the city is very small, but its characteristics are clear. This urban type has arrays that form a simple urban space lined up along a single width of a street. Because of its long length, minor streets are created across the inside of the blocks. In addition, while other types are following the streets with the diagonal directions, this urban type is formed on the streets with the orthogonal directions.

Urban Type 7. Since the main elements of this urban types are not housing buildings, it has a relatively higher density than other urban types. There is no generalized shape that can be specified for this urban type, each block in it is represented by the shape of

buildings such as a hospital, school, government office, shopping mall, etc. Moreover, this urban type is located where the boundaries are ambiguous mainly due to the conflicts and collapse of urban types. This urban type can be called the ‘fragmented’ urban type which can be developed to the ‘urban island’.

Urban Type 8. This urban type consists of typical blocks with grid systems. This is mainly distributed on the Downtown and Strip District. Like Urban Type 7, there is no generalized shape that can be specified for this urban type, each block in it is represented by the shape of buildings. However, unlike Urban Type 7, high density buildings of this type are arrayed within a grid system.

Urban Type 9. This urban type consists of blocks with four edges surrounding by streets. This is mainly distributed on the Highland Park. The amount of this urban type throughout the city is very small, but its characteristics are clear. While the blocks in other urban type have generally two sides with a series of buildings, the blocks in this urban type have four sides with a series of buildings.

Urban Type 10. This urban type consists of blocks parallel to the streets. This is mainly distributed on the Green Field and Polish Hill. The amount of this urban type throughout the city is very small. The blocks in this urban type are linear polygon shapes along the streets. Since the streets are responding to the terrains of hills, they are not linear but irregular shapes. Therefore, the blocks in this type are also bent, crossed, broken, and irregularly shaped along the way.

Urban Type 11. This urban type represents places where a certain number of empty parcels exists and no distinctive shape of blocks. A typical example of this urban type is Middle Hill. There are a lot of empty parcels in that neighborhood, and a building is adjacent to some of them.

Urban Type 12. This urban type consists of blocks parallel to the streets. This is mainly distributed on the Bloomfield and Shadyside. The blocks in this urban type have rectangle shapes along the main streets and two rows of parcels. They look similar with Urban Fabric 2’s blocks, but parcels in these blocks have more longer depth and a building’s occupancy of a parcel is lower than Urban Type 2.

Urban Type 13. This urban type consists of blocks parallel to the streets. This is mainly distributed on the Garfield. This urban fabric consists of Long-shaped buildings that occupy several parcels and buildings which have large footprints within a parcel. In particular, the commercial area in the Garfield is the typical example of this urban type’s architectural contexts conditions.

Urban Type 14. This urban type consists of blocks that forms inside space of themselves. The inside space of a block can have another block or can be empty. Compared to other types, building size is similar with them, but block size is little bit

huge because the blocks surround other space of themselves.

Urban Type 15. This urban type consists of blocks parallel to the streets. This is mainly distributed on the Green Field and Swisshelm Park. Like Urban Fabric 4, the blocks in this urban type are U-shaped. The side of the block also has a series of buildings. But there are no streets crossing the blocks.

Urban Type 16. This urban type consists of blocks parallel to the streets. This is mainly distributed on the Squirrel Hill, Green Filed, and Hazel Wood. Like Urban Type 3, the blocks in this urban type have elongated rectangle shapes along the main streets and two rows of parcels. But it has shorter depth of blocks and more compact shape because houses of two rows forming back decks are closer to each other than Urban Type 3.

Urban Type 17. This urban type consists of blocks parallel to the streets. This is mainly distributed on the Highland Park, East Liberty and Friendship. Like Urban Type 2 and Urban Type 16, the blocks in this urban type have rectangle shapes along the main streets and two rows of parcels. However, unlike Urban Type 2, there are no streets crossing the blocks and splitting them into two rows of parcels. While it seems similar to Urban Type 16 regarding to the depth of blocks, its width is shorter than Urban Type 16.

Urban Type 18. This urban type has basically same with the Urban Type 7. But this urban type has shorter and more fragmented blocks than Urban Type 7 and more empty spaces due to the smaller parcel and building size. In addition, it tends to exist around big buildings and Urban Type 7. This is another type of the ‘fragmented’ urban type.

Urban Type 19. This urban type has basically same with the Urban Type 17. Only the difference between this urban type and Urban Type 17 is the streets crossing the blocks. This is mainly distributed on the Larimer, Shadyside, and Regent Square.

Urban Type 20. This urban type consists of blocks parallel to the streets. This is the main urban type of the Homewood. The form of the block is similar to Urban Type 16, but the variety of building shapes is lower than Urban Type 16. One building’s footprint looks uniform to adjacent buildings.

Urban Type 21. This urban type has the lowest occupancy rate in Pittsburgh. It mainly exists on Regent Square. Especially, this urban type coexists with the Urban Type 19. It has diagonal and irregular forms of blocks. These blocks fill the portion of the Urban Type 19 when it is bent or split.

4.4 Interpretation of DMUMT

The interpretation of the data-driven map is up to the architects. In this section, as architects, I investigate whether the urban types’ boundaries created by the computer are appropriate, what spaces exist inside the urban fabric, and how different urban fabrics are

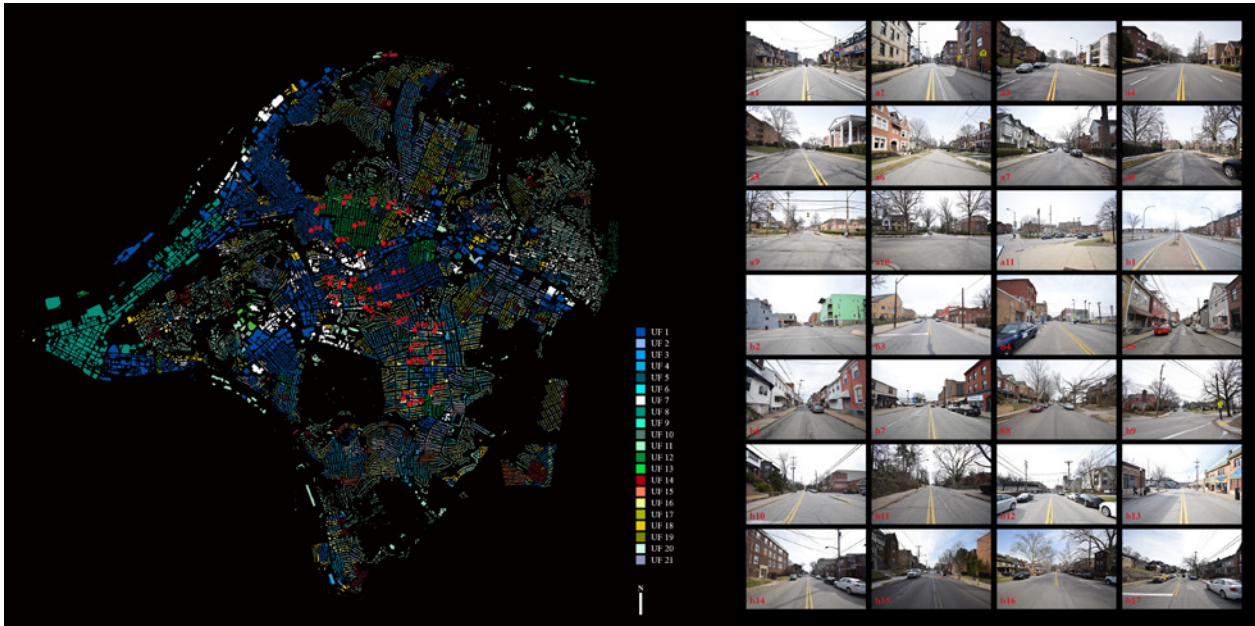


Fig. 64. Site Survey, Places Where City Fabrics Conflict

conflicting. For this investigation, I execute the site survey and record the current situations of the site that urban fabrics are conflicting. The selected two of these sites are more precisely analyzed to identify that how computer detect the place of conflicts of urban type.

I also analyze the characteristics of urban type boundaries' pattern in DMUMT and infer the rationale and development of the urban type boundaries in the history of urban development. By analyzing the remaining historical maps and literature data in the city archives, museums, and university library, I will discover what events in Pittsburgh development affected the formation of the current urban structure and urban morphological types.

4.4.1 Conflicts among Urban Fabrics

Investigating the conflicts of urban morphological types, or urban fabrics, begins with selecting the sites of investigation. In order to select the sites, I focus on two parts of Pittsburgh: Shadyside area, where large fabrics collide, and Squirrel Hill area, where small fabrics collide. In these two parts, the main conflict points are set around the streets. I record not only the situations of the conflicts among urban fabrics by tracing photographs, but also the inside situations of the fabrics, to see how the typical conditions form the actual urban space. In this section, I will look at three of these sites in detail.

First, there is a conflict in the Shadyside, Pembroke Place and St. James Street (marked as 'a07' on Fig. 64). This point of conflicts shows the boundaries of Shadyside's expensive price area and average price area. They have different architecture types: mansion style house (Urban Fabric 19) and compact collective house (Urban Fabric 1).

It can be seen that even the same neighborhood has a different architectural and urban environment. Of course, architects can recognize such difference, but it is hard for them to precisely define where the difference starts and ends. This is because the boundary is ambiguous in 2D map.



Fig. 65. Pembroke Place and St. James Street, Facing South (Left) and North (Right), Pittsburgh

This conflict is clearly described in the **Fig. 65** facing the north and south at Pembroke Place and St. James Street. The east side of the street is the vernacular⁶⁵ of common collective housing in Pittsburgh, Shady Side. This type of housing forms the urban space in a way that similar size buildings are tightly lined up. The west side of the street has the buildings largely occupying the ground locating at a distance from each other, called mansion-styled houses as the type of independent residence with individuality.

65. Ulrich Flemming et al, *A Pattern Book For Shadyside*, p.4-8

This difference between fabrics is subtle, constructive, undisclosed, and experience



Fig. 66. Façade of Vernacular of Shadyside (top, bottom-left), Mansion in Shadyside

-based knowledges that are difficult to perceive unless one spends a long time or has sharp discernment in the city. The computer-generated urban fabric maps based on the urban contexts can show that computers can detect this subtle difference.

Fig. 67. Fragmented Urban Fabric (Left) and Holy Spirit Byzantine Catholic, Fifth Avenue and Clyde Street



Fifth Avenue and Clyde Street (marked as ‘a11’ on Fig. 64) has the fragmented urban fabrics by various huge monumental buildings. There are various building programs such as houses, churches, broadcasting stations, university facilities, government offices, apartments. Besides this area has various scale of buildings and no distinct streets inside of the block. Particularly, the impact of the two churches (Holy Spirit Byzantine Catholic and Rodef Shalom Congregation) on the surrounding urban fabric is enormous. They cut down the housing fabrics that runs from north to south and keeps it from going any further. As a result, at this intersection, large monumental buildings can exist regardless of the surrounding urban fabrics. In the **Fig. 67**, fragmentation of urban fabrics is more noticeable. While other photographs illustrate that the urban fabrics form a city space centered on a certain street, this urban fabric has no distinct streets, and the urban forms of fabrics are scattered as objects.

The fragmentation of urban fabrics is not always negative. With the information exactly where the fragmentation of urban fabrics is located, one can increase the efficiency of the urban space by enlivening the surrounding fabrics or by implanting the third urban fabrics through the public development or city planning guidelines. It is encouraging that computer can distinguish the urban space with those potentials from other urban fabrics without any input knowledge.

Fig. 68. Hobart Street and Wendover Street (Left) and Wightman Street and Marlborough Road (Right)



Hobart Street and Wendover Street and Wightman Street and Marlborough Road (marked as 'b14' and 'b16' on Fig. 64), these areas have numerous small-sized conflicts among small fabrics, unlike the above two areas. As the scale of conflicts in 'b14' is smaller than 'b16's one, the urban spatial change in 'b14' is interesting. The photograph taken at 'b14' includes different types of buildings that consist different fabrics: 1950s apartment houses (Urban Fabric 11), 2-story private houses (Urban Fabric 16), 3-storey residential apartments (Urban Fabric 1), etc. 'b16', on the other hand, illustrates the typical conflict of two different fabrics of Squirrel Hill. The street between the apartment houses and the condominium apartment is the border of the conflict. Compared with 'b16', the degree of change in 'b14' is low but there are various kinds of change.

The examples of the three regions above show that the computers have successfully distinguish the urban fabrics. Computers' classification of urban fabric is not only highly accurate, but is also based on characteristics derived from the form of urban space. Through the site survey, I surprisingly confirm that computers' classification of urban fabrics and their boundaries is the way to precisely distinguish 3D space in a city through 2D urban information. During the survey in the field, I have also found that this technology can be applied to various aspects of urban architecture, such as urban planning and development. This technology has been proved to be a method that can be implemented by efficiently strengthening the urban space analysis which requires long time experience in a city and intensely trained discernment in urban space.

4.4.2 Historical Reasoning

Beyond the analysis of each urban fabric and the conflicts among the fabrics, observing DMUMT from a larger perspective reveals certain patterns of the city structure. The neighborhoods near the Monongahela River, such as Squirrel Hill, Green Field, Hazel Wood, and Point Breeze are consisted of complex and small urban fabrics. On the other hand, the neighborhoods near the Allegheny River, such as Downtown, Strip District, Lawrenceville, Shadyside, and East Liberty, are consisted of the huge fabrics. In other words, the northwestern part of the city is occupied by large fabrics, and the southeast is collaged by a large number of small fabrics.

These different forms of urban fabrics do not simply remain in shape or morphological aspects of a city, but have a history of development and collective memory of the city⁶⁶. Reverse tracing the basis of patterns created by urban fabrics in Pittsburgh not only elicits the computer generate data that can infer non-physical information from physical information, but also provides a proof that the inference can be of a qualitatively reliable level.

66. Rossi, *The Architecture of the City*, p.18

The starting point for analyzing the characteristics of DMUMT patterns is the railroad system in Pittsburgh. **Fig. 69** shows two urban fabric maps by different times: the map

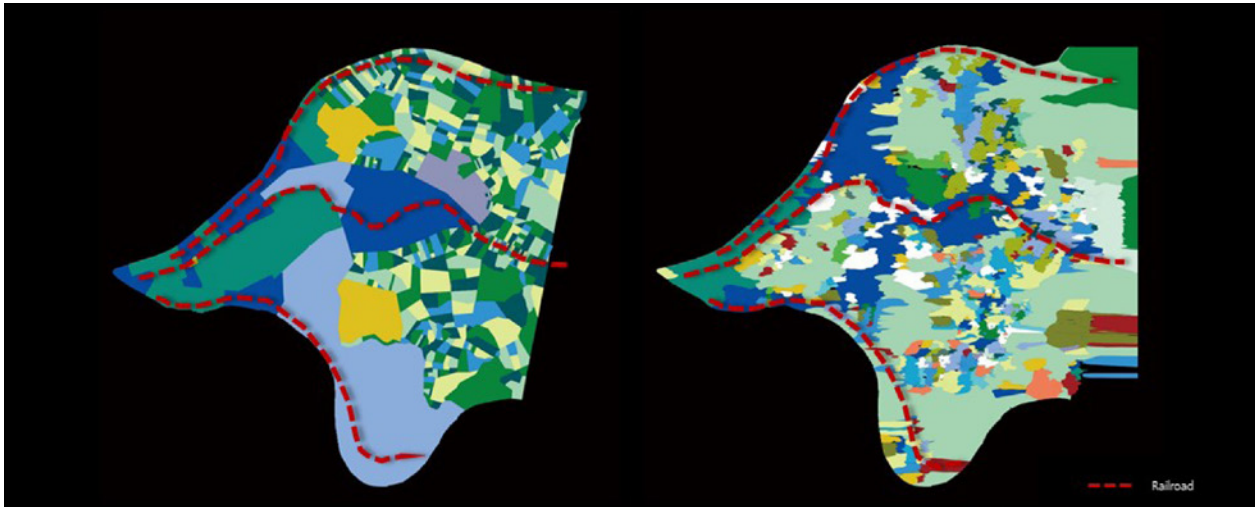


Fig. 69. Urban Fabrics Map with Railroad 1872 (Left) and 2017 (Right)

67. '1872 – Atlas of the cities of Pittsburgh, Allegheny, and the adjoining boroughs', University of Pittsburgh, <https://historicpittsburgh.org/maps-hopkins/1872-atlas-pittsburgh-allegheny>

68. Evans, "Notes on Pittsburgh Transportation Prior to 1890.", p.163-164

69. Evans, "Notes on Pittsburgh Transportation Prior to 1890.", p.242

70. Schusler, "The Railroad Comes To Pittsburgh.", p.259-260

71. Ulrich Flemming et al, *A Pattern Book For Shadyside*, p.3-3

72. Daly, "Zoning.", p.104

73. Schusler, p.260

on the left side is the urban fabrics defined by the occupation owner in 1872 Pittsburgh⁶⁷ and the map on the right side is the urban fabrics in 2017. The red dotted line represents (planned) railway lanes at the time of 1872. It clearly shows that there are large urban fabrics in the vicinity of the railroads.

After the French and Indian War, the settlement of the title to western Pennsylvania, and the Revolution, Pittsburgh named 'Gateway to the West'. At this time, Pittsburgh had been developed as lodging town with hotels, taverns, stables along the Monongahela river. The main facilities were in the Downtown near the canal⁶⁸(Fig. 70), the rest of the area was suburb with farming and wilderness⁶⁹.

The deterioration of canal-based transportation in Pittsburgh and the emergence of railroads connecting Pittsburgh to Philadelphia make it possible to commute from the suburban East End to the downtown area. As a result, suburbs are recognized as part of Pittsburgh and various types of intensive developments take place⁷⁰. Shadyside and East Liberty are the proper exemplars of these developments. As the population expanded due to railway connections in these areas where the farms and mansions were located⁷¹, the government began to control the amelioration and deterioration of the city through zoning⁷². East Liberty was the center of the East End, played a role of the lodging town, instead of the downtown area of Pittsburgh, and led the rapid development of the eastern suburbs⁷³. In addition, there is another lane of Pennsylvania Railroad company, which starts from Downtown, goes through Strip District and Lawrenceville, and then goes northeast. In these areas which were heavily industrialized areas of heavy and light industry, huge urban fabrics are formed due to the development planning by companies and the government.

After the rise and fall of Pittsburgh, the size of East Liberty's city structures has been relatively reduced, but the city where the railway had passed has maintained the framework of the city fabrics that was created 1890s.

There were farmland and forests only a few years ago around the Squirrel Hill⁷⁴ where the railway did not pass. Since the 1760s, proprietary government of Pennsylvania had started to sell the land. Unlike the early time of mansions and farms in Shadyside, the lands in Squirrel Hill were compartmentalized and sold through land surveying. Some owners of the land came to immigrate to live in Squall Hill, but some others had only owned the land. As the area of Squirrel Hill is large and the population is small, it has remained in the suburbs for a long time.

Major change occurred in 1893 with the coming of the electric trolley. It connected the bustling steel town of Homestead with Pittsburgh's East End, stimulating the building of hundreds of large homes on Shady Avenue and Denniston Street for middle management of the mills⁷⁵. The owners of the land sold their properties to the developers or they developed their lands by themselves. These private-led developments, unlike the government-led one, have persisted for many years, creating a variety of living and urban spaces. As a result, the city becomes a collage of various urban fabrics.

Whether this data-driven map cultivates intuition on historical knowledge is still debatable. Since Pittsburgh's urban development history might be an exceptional case for reverse tracing, applying to more cities can confirm that data-driven map cultivates the intuition on historical knowledge. However, historical reasoning can be a qualitative evaluation for the reliability of data-driven map, because the present urban structure is not independent of the past urban structure.



Fig. 70. 1850s Canals in Pittsburgh

74. Frew, "Squirrel Hill.", p.242-243

75. Squirrel Hill Historical Society, Squirrel Hill History, <https://squirrelhillhistory.org/squirrel-hill-history/>

5 Discussion and Applications

The experiment described in the previous chapter treat a city as a single data space, substitute data for urban elements, and creates a data type representing the architectural context conditions. Through the data type, computer generated urban fabric maps. Does the computer understand a city through the map of the urban fabrics? Or Does it simply search and calculate parameters consisting the characteristic of data? These questions provide an important point of discussion about defining the relationship between human and computer in urban architecture in the future.

This thesis treats AI as a tool to calculate the gradients and features of urban data containing complex relationships for defining and visualizing the urban fabrics in a city. Reviewing not only the new knowledge found in urban context analysis, but also the logical concept of this tool, in this section, I discuss the value of the map of urban fabrics generated by AI as a tool.

5.1 The Value of the Urban Analysis through Context-rich Dataset

For several decades, AI techniques have been employed in architecture and showed several achievements in terms of executing specific architectural problems, such as room layout. Most works that use AI techniques in architectural design have focused on the internal aspects of buildings. In other words, the architectural contexts which is implicitly and explicitly representing various relationships were not well considered and integrated with AI techniques in architectural design.

While there might be various reasons why AI techniques cannot be actively intervened and used with architectural contexts, one crucial reason is the limitations of computer performance. Despite of the fact that without considering the architectural contexts, practical uses of AI cannot be developed in architectural design, unfortunately, calculating and learning the contextual information requires a heavy load of computing resources, such as fast data transfer rates, large amounts of data operation memory, amplified calculating units, etc. Limits of computation power were not the problem only in the architectural design discipline, but have arisen in many areas of research involving AI.

However, now the situation has changed. Advances in hardware technology have overcome the limitations of computing performance and led the popularization of contemporary AI techniques. Due to these advances, the development of AI technology for architectural design has also become more popular than before, expanding the potential that it can intervene more with architectural design. Specifically, in urban architecture, if assuming a city is a space where highly complex data composed of multiple elements are buried, it is possible to discover new layers of urban spatial characteristics through AI techniques. Ultimately, developing AI models through the dataset including various types

of relationships that can be contextual information in the city scale, will help architects and urban designers overcome one of the reasons that AI has not been widely used in architectural design.

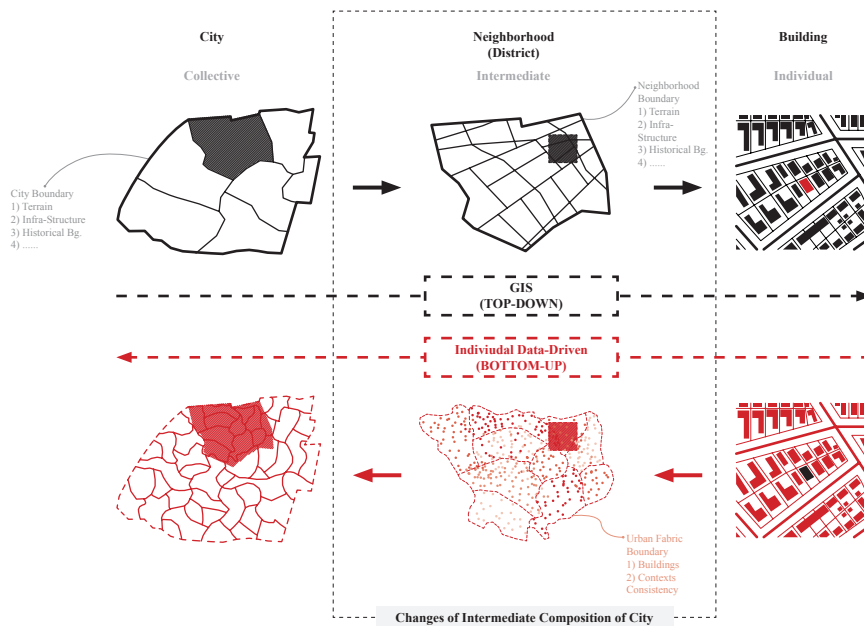
Urban morphological analysis through DMUMT shows the potential that with the proper representation methods, datasets, and models for the learning process, it is possible to develop a system that deals with this complexity of urban data and generate the meaningful results from the analysis of the data. This system will result in the unprecedented interpretations of the city using AI technology, which can be used as the foundation for the various architectural design decision.

DMUMT is a map of urban morphological types generated by the computer based on architectural context conditions. With the 45,852 buildings in DID-PGH, the computer classifies and clustered each building according to the relationship between the building and its surrounding buildings, sites, and streets. Defining the urban fabric as the cluster of the buildings sharing similar architectural context conditions, the urban morphological types are visualized by remapping the cluster information. The visualized urban morphological types are forming some patterns in the DMUMT, and these patterns have revealed several ambiguous knowledges that cannot clearly be articulated in architectural and urban studies.

First, DMUMT enabled users to quantitatively and qualitatively analyze the characteristics of urban fabric. Urban fabric is a unit that contains the qualitative characteristics of urban space. DMUMT shows the exemplary method to set qualitative definition of urban elements through quantitative information. In other words, DMUMT provides an approach to correlate the characteristics of urban space with quantitative analysis based on data, without solely relying on qualitative analysis. For example, fragmented urban fabrics and their boundaries, which has been difficult to be defined or detected due to their qualitative definition, are detected by the feature extraction techniques in machine learning from DID-PGH. Architects, urban designers, and researchers share the meaning of the fragmented urban fabric, but it is vaguely defined. It can be established through the particular features of the architectural contexts in urbanism studies. The great advantage of DMUMT is that it quantifies large amounts of urban data in morphology and uses it to generate qualitative definitions.

Second, DMUMT provides the information of conflicting urban fabrics. It shows the situation of two or more fabrics are conflicting and where it happens. One of the reasons for analyzing the urban fabrics is to discover a basis for considering what is happening and what design is needed when different urban fabrics meet. But finding the boundaries is closely related to the definition of urban fabrics. As the definition of urban fabric was ambiguous, its boundaries were also ambiguous. Definition of urban fabrics based on data that can calculate similarity of architectural context conditions leads to the definite boundaries of urban fabrics. Visualizing the definite boundaries of the urban fabric does not

Fig. 71. The Concept of Reverse Analysis for City



simply indicate the point of conflicts. It shows how the fabrics expand and disappear, what the main space occupying the city is, and what city space to design on given sites.

Lastly, DMUMT shows the characteristics of not only each urban fabric but also the urban spaces through the overall patterns consisted of the urban fabrics. The information about the characteristics of the city through urban fabrics provides the identity of the urban space of the city. Qualitative studies make the task for urban spatial identity inexecutable because of the difficulty on keeping the consistent standards for urban fabrics. However, quantitative studies make this task feasible because of the advantages of large amounts of data helping computers make more informed decisions. The numerical calculation of the architectural context conditions of all the buildings in the whole city shows how and which urban fabrics spread throughout the city, understanding the historical and morphological characteristics of the city, as we have seen in Section 4.4.2. The value of the urban analysis technique using DID and machine learning is not limited to simply mapping urban fabrics in new ways. DMUMT is only one exemplary experiment in which a computer uses architectural contextual information to provide a new perspective to the analysis of urban space. Through this experiment, I have established criteria for several concepts that I could not have seen in the past or depend on the intuitions of architects and city planners, and confirmed that these criteria work well in real world. What more remarkable is the approach of this method to read and interpret the city (Fig. 71). There is a limit for architects to deal with numerous information thoroughly. A city can be seen as a data space where this granular and numerous information is well-suited and intertwined.

In the existing GIS, the city itself was regarded as a collection of information, rather than a complex data space. This recognition does not interpret the data from a computational point of view but still allows the data to be processed from a human perspective. By segmenting into states, cities, neighborhoods, streets, and parcels, people have classified, categorized, and understood the complex data of cities. This was an urban interpretation based on approaches based on segmentation.

But if people have a tool that allows them to observe a myriad of infinite pieces of information one by one, they can look at the various information in the city and identify certain rules and patterns. They can find and differentiate the overall patterns in dataset by integrating individual data instead of segmenting it. The possibilities of AI technologies in architectural design will be amplified when thoroughly dealing with a data space called a city with complex and innumerable relationships. If people have a new method of integral logics to calculate the overall shapes of urban data rather than segmenting them, it will be another way to interpret the city analyzing the implicit and explicit phenomenon more flexibly and efficiently.

Moreover, this integral approach is not limited to physical data. This approach can be applied to non-physical data. If urban spatial information is linked to non-physical data, it will be possible to analyze social, cultural, and economic factors reflecting the subtle difference for various urban phenomena in building scale, interpreting them in terms of urban space. Urban granularity is derived from urban forms, but human activities interacting with urban forms also have granularity. In the meantime, people have related non-physical data with urban space using only segmental thinking. However, this way of thinking eventually led to a generalization error due to unit or resolution, and subtle differences were ignored. The non-physical data that represent the city phenomenon already exists, such as price, history of buildings, census, usage of mobile data, etc. If people can find patterns that take into account the subtle differences of these data, the analysis of the city will be in a different paradigm.

5.2 The Applicable Scenarios

Scenarios where this urban analysis method can be applied are distinguished according to the users.

First, it can be applied to urban planners, policy makers and researchers as ‘context-rich urban space and phenomena analysis method’. Rather than analyzing various urban phenomena as one measurement or index, it is an analytical methodology that investigate urban phenomena in relation to the surroundings. For example, when generating the DUFM, only morphological information was input into the vector representing the architectural context condition. However, it is possible to create a new dataset by concatenating the nonphysical information, such as the resident's age information

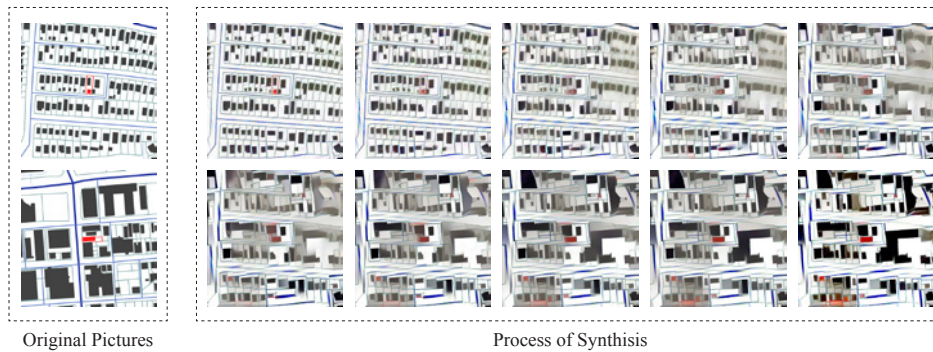


Fig. 72. The Concept of Urban Synthesis

extracted from the census data, into this vector. This newly produced dataset contains the relationships between architectural context similarity and resident's age, and can be remapped to urban space to visualize a pattern showing those relationships. It is an application that can contribute to identifying the relationship between urban spatial structure and urban phenomena.

Second, this technology can be a tool for designers to synthesize different urban morphological types based on the learning data of different architectural context conditions. Technically, after learning data that contains information of different urban fabrics, it searches for latent spaces and reorganizes the information. Or users can synthesize different styles of urban morphological types using style transfer techniques. **Fig. 72** is the concept of this synthesis, showing the process of synthesizing two different urban morphological types. With this synthesis, designers can simulate the city structures by setting the urban morphological types to be synthesized, the degree of compositing, and the place to apply. Beyond simple simulations, this synthesis will contribute to creating the intermediate or 'third' urban morphological type. Based on the characteristics of existing urban types, it is possible to create a new urban morphological type that does not belong to any existing urban morphological type and boundaries with gradients instead of rigid boundaries between urban morphological types.

The third application is for citizens' use, allowing them to understand the urban spatial characteristics of their current residence area and search for areas with similar characteristics. Most real estate search engines show properties by price and area. However, there is no search for the characteristics of the space that contain the life of the citizens, that is, the quality of the space. Citizens need to look at the properties that fit the price and judge the spatial characteristics of the property by themselves. Citizens who have no background for the city or no expertise on understanding urban space, the width of judgment is inevitably narrow. DMUMT can efficiently provide useful information for the desired urban space with appropriate conditions. In addition to finding the desired urban context in a city, I can expect to use it more widely for finding similar urban contexts among other cities.

6 Conclusion

In this thesis, in contrast to analyzing the morphological homogeneity of urban space in a top-down method using existing GIS data only focusing on individual buildings, this thesis illustrates a bottom-up method to investigate a tendency and patterns through the dataset comprising individual buildings and their surrounding situation in an urban scale. The novelty of this method opens the potential that morphological analysis can be integrated with the non-morphological aspects, such as economic and socio-cultural data, cultivating the new insight on the understandings of city space. This integration can not only extend the physical relationship of a building to a non-physical relationship but also catch the phenomenal patterns derived from these two entangled relationships. Therefore, this method has value as a research tool that hybridizes with other research methods and triggers new knowledge in urbanism.

Through this method, I was able to identify that Pittsburgh has 21 different distinct clusters in data space that share the similarity of the morphological traits. The neighborhoods near the Monongahela River, such as Squirrel Hill-South, Green Field, Hazel Wood, and Point Breeze have the multiple small and segmented clusters that indicate several different types of architectural context conditions. Otherwise, the northern neighborhoods, such as Lawrenceville, Bloomfield, Shadyside, and Oakland, have a large size cluster meaning one type of architectural context conditions. By investigating the historical maps and literature, I can infer that these two different types of clustering in Pittsburgh were derived from the two different types of development in the early history of Pittsburgh: government leading and private development.

Through this research, we have invented and developed ‘context-rich’ datasets containing the relationships between buildings and their architectural context conditions in a city. These datasets are employed to train the deep learning model and they shows the high recognition rate of the datasets and the potential to be used for design decision process. Beyond simply using the dataset for training of the AI model, I classified and clustered the buildings in the datasets by their space-contextual characteristics for the new definition of urban fabrics in a city. By visualizing the results of the classification and clustering, I created the data-driven urban fabric map. This map allows the grey area of urban analysis which have depended on the personal know-how and experiences, such as urban spatial characteristics of each fabrics, fragmented fabrics, transition and changes of fabrics, to be clearer.

Architectural design, especially the architectural activities in urban space cannot be defined through the independent events in virtual space, but can be defined by the relationships with lands, surrounding buildings, streets, urban spaces, etc. This relationship appears as the quality of space through the nature of public and private, outside and inside, light and

the darkness, openness and closeness, and so on. In other words, the architectural design is to define the characteristics of the space through the relationship with the context of the given site.

For computers, in order to intervene more deeply in this essential value of architectural design, they must go beyond passive design tools which are providing the visualization and drawing functions. The active design tools that can make design decisions in some level will complement the architect's cognitive limitations and can trigger new design paradigms. The function to understand the architectural contexts is indispensable for computers to become an active design tool. As the first step in the computer as an active design tool, in this research, I have examined the possibility of quantifying architectural context conditions and applying them to design through modern computer technology.

Different from room layout and form-finding in the field of architectural design with AI technology, this thesis addresses the task of urban analysis through architectural contexts. I have developed a new data set format called DID, experimented with its recognition and use to ensure the possibility of design tools that take context in architectural design, applied to a city which is a space with rich and complex relationship, and generated DMUMT through DID for investigating the characteristics of urban space. DMUMT compensated human's limitation providing the quantitative definition for the vague qualitative one and thoroughly investigating the numerous data in a city.

This approach shows the different aspect for the usage of AI in architectural design from precedents in that computers can make design decisions based on the dataset containing the architectural or urban context information. In previous researches, users set the major features affecting the tasks, computers calculate it without the consideration on the contexts. Otherwise, following the approaches demonstrated in this thesis, computer can execute the task by learning and feature extraction from the data containing the context information which seems similar with the data that architects use.

This research shows one reason why the rapid and remarkable development of AI technology has not been deeply involved in architectural design and has not led to innovative changes in the architectural design industry. That is the lack of consideration on the architectural contexts for AI. Until now, the attempt to integrate AI into architectural design quantified architectural problems and replaced them with an AI problem. Contextuality, the essential value of architectural design, did not have room for intervention. With the appropriate format of the dataset and AI model, the consideration on contextuality can provide more enriched opportunity that AI can be involved more deeply in architectural design.

For further studies, there are three different ways to expand to this research: various samples of cities, analysis methods integrated with other data in urban society, and more three-dimensional diagrams. The achievements from this research can be applied to

other cities too. If DMUMTs of diverse cities can be generated in this method, the spatial and morphological identity of each city can be investigated by comparing the types of boundaries of urban fabrics and distributed points model in data space. Therefore, it is planned to generate and analyze diagram data for various cities. Moreover, I am planning an integrated analysis of the DID with other data related to urban society (e.g. historical map, surveys, census, price, etc.) for investigating the relationship between urban phenomena and urban morphological structure. Lastly, as current dataset only describes the height of the buildings, there is a lack of three-dimensional analysis of the architectural context. I will experiment with a more detailed architectural context analysis through a diagram with additional three-dimensional information such as terrain.

Increasing the consideration of architectural context through the above-mentioned further advanced studies accelerates integration of architecture, city and AI, and ultimately establishes the foundation of the new design paradigm. I hope this study will be the cornerstone of this new design paradigm.

Appendix

Appendix 1. 43 architectural and urban design projects that used AI techniques from 1960s to 2017.

Appendix 2. 15,284 images of DID for Seoul

Appendix 3. 2,080 images of DID for Shadyside

Appendix 4. 45,852 images of DID-PGH, arranged by neighborhood's density

Appendix 5. 45,852 images of DID-PGH, arranged by target building's density

Appendix 6. 45,852 images of DID-PGH, arranged by target building's height

Appendix 7. 45,852 images of DID-PGH, arranged by the direction from West to East

Appendix 8. 45,852 images of DID-PGH, arranged by the direction from South to North

Appendix 9. Data-driven Urban Fabric Map: Footprints

Appendix 10. Data-driven Urban Fabric Map: Boundaries

Appendix 11. Typical conditions of the clusters can be used to describe the characteristics of each urban fabrics

Appendix 1

43 architectural and urban design projects that used AI techniques from 1960s to 2017.

	Author	Title	Task	Algorithm	Contextuality	Year
1	N.Cross	The Automated Architect	Room Layout	custom heuristic algorithm, CARD	Contextless	1967
2	N.Negroponte	The Architecture Machine	Urban Planning	Evolutionary Algorithm, URBAN5	Context	1967
3	N.Cross	The Automated Architect	Building Layout	Own Minimization Algorithm, STUNI	Context	1970
4	C.Eastman	Heuristic Algorithms For Automated Space Planning	Room Layout	Tree Search, custom Algorithm, S-Relation	Contextless	1971
5	G.Stiny and W.J.Mitchell	The Palladian Grammar	Room Layout	Shape Grammar	Contextless	1978
6	G.Schmitt	Archplan: An Architectural Planning Front End To Engineering Design Expert Systems	Building Form	Rule-based	Context	1988
7	U.Flemming	More On The Representation And Generation Of Loosely Packed Arrangements Of Rectangles	Room Layout	Rule-based	Contextless	1989
8	S.Gero and A.Kazakov	Learning And Re-Using Information In Space Layout Planning Problems Using Genetic Engineering	Room Layout	Genetic Algorithm	Contextless	1996
9	C.Damski and S.Geroan	Evolutionary Approach To Generating Constraint-Based Space Layout Topologies	Room Layout	Genetic Algorithm	Contextless	1997
10	J.C.Damski and J.S.Gero	An Evolutionary Approach To Generating Constraint-Based Space Layout Topologies	Room Layout	Evolutionary Algorithm	Contextless	1997
11	J.S.Gero and V.A. Kazakov	Learning And Re-Using Information In Space Layout Planning Problems Using Genetic Engineering	Room Layout	Genetic Algorithm	Contextless	1997
12	G.Weinzapfel et al	Image: An Interactive Computer System For Multi-Constrained Spatial Synthesis	Room Layout	Least Mean Square Error	Context	2000
13	T.Vardouli	"Architecture-By-Yourself": Early Studies In Computer-Aided Participatory Design	Room Layout	Heuristic, YONA	Contextless	2000
14	L.Caldas	An Evolution-Based Generative Design System:using Adaptation To Shape Architectural Form	Building Form	Genetic Algorithm	Context	2001
15	J.J.Michalek et al	Architectural Layout Design Optimization	Room Layout	Gradient-based search algorithms, Simulated Annealing, Genetic Algorithms	Contextless	2002
16	J.Michalek and R.Choudhary	Interactive Design Optimization Of Architectural Layouts	Room Layout	Pareto Set	Contextless	2002
17	C.KUO	Spatial Analysis Of Chinese Garden Designs With Machine Learning	Building Layout	Machine Learning	Context	2003
18	J.Moon	Shape Grammar For Mies Van Der Rohe's High-Rise Apartment	Room Layout	Shape Grammar	Contextless	2007
19	D. Donath	Massing Study Support: A New Tool For Early Stages Of Architectural Design	Building Form	Parametric System	Context	2008
20	P.Merrell et al	Computer-Generated Residential Building Layouts	Room Layout	Machine Learning	Contextless	2010
21	K.Knecht and R.Koenig	Generating Floor Plan Layouts With K-D Trees And Evolutionary Algorithms	Room Layout	K-D tree, Evolutionary algorithm,	Contextless	2010
22	H. Hua and T. Jia	Floating Bubbles	Building Layout	Physics, Heuristic	Contextless	2010

23	G.Pellitteri et al	A Generative Design System To Interactively Explore Different Urban Scenarios	Building Layout	Parametric System	Context	2010
24	M.Mirahmadi and A.Shami	A Novel Algorithm For Real-Time Procedural Generation Of Building Floor Plans	Room Layout	Tree Search, Rule-based, Squarified Treemap algorithm	Contextless	2012
25	M.Mathias	Automatic Architectural Style Recognition	Building Style	Deep Learning	Context	2012
26	S.Becker et al	Combined Grammar For The Modeling Of Building Interiors	Room Layout	Shape Grammar	Contextless	2013
27	M.E.Bayraktar and G.Cagdas	Fuzzy Layout Planner	Room Layout	Physics	Context	2013
28	A.Bakunowicz	Neural Self-Organising Maps And Genetic Algorithm: Evolving 3D Cellular Automata Architectural Model	Building Form	Genetic Algorithm	Contextless	2013
29	R.Koenig and K. Knecht	Comparing Two Evolutionary Algorithm Based Methods For Layout Generation: Dense Packing Versus Subdivision	Room Layout	Evolutionary algorithm, Search	Contextless	2014
30	R.A.Frenando	Space Planning And Preliminary Design Using Artificial Life	Room Layout	Evolutionary Algorithm	Contextless	2014
31	R.Puusepp	Agent-Based Models For Computing Circulation	Room Layout	Flocking	Contextless	2014
32	P.Veloso	Exploring The Bubble Diagram	Room Layout	Physics, Least Mean Square Error	Contextless	2014
33	R.Koenig and K.Knecht	Comparing Two Evolutionary Algorithm Based Methods For Layout Generation: Dense Packing Versus Subdivision	Room Layout	Genetic Algorithm	Contextless	2014
34	Boon et al	Optimizing Spatial Adjacencies Using Evolutionary Parametric Tools: Using Grasshopper And Galapagos To Analyze, Visualize, And Improve Complex Architectural Programming.	Building Form	Genetic Algorithm	Contextless	2015
35	T.Ireland	A Cell-Inspired Model Of Configuration	Room Layout	Agent-Based, pherom	Contextless	2015
36	F.Bazalo and T. J.Moleta	Responsive Algorithms	Building Layout	Physics	Context	2015
37	S.Das et al	Space Plan Generator. Rapid Generation & Evaluation Of Floor Plan Design Options To Inform Decision Making	Urban Planning	K-D tree, Search	Context	2016
38	J.Silvestre et al	Artificial Imagination Of Architecture With Deep Convolutional Neural Network	Building Style	Neural Networks	Context	2016
39	T.Strobbe et al	Automatic Architectural Style Detection Using One-Class Support Vector Machines And Graph Kernels	Building Style	SVM, Graph	Contextless	2016
40	J.Harding	Evolving Parametric Models Using Genetic Programming With Artificial Selection	Room Layout	Genetic Algorithm	Contextless	2016
41	A.Pasternak	Optimization Of The Building In Relation To The Insolation Conditions Of Premises In Adjacent Buildings	Room Layout	Genetic Algorithm	Context	2016
42	D.Nagy et al	Beyond Heuristics. A Novel Design Space Model For Generative Space Planning In Architecture	Urban Planning	Genetic Algorithm	Context	2017
43	T.Narahara	Collective Construction Modeling And Machine Learning Potential For Architectural Design	Building Form	Agent-Based, Genetic Algorithm, Neural Networks, Flocking	Contextless	2017

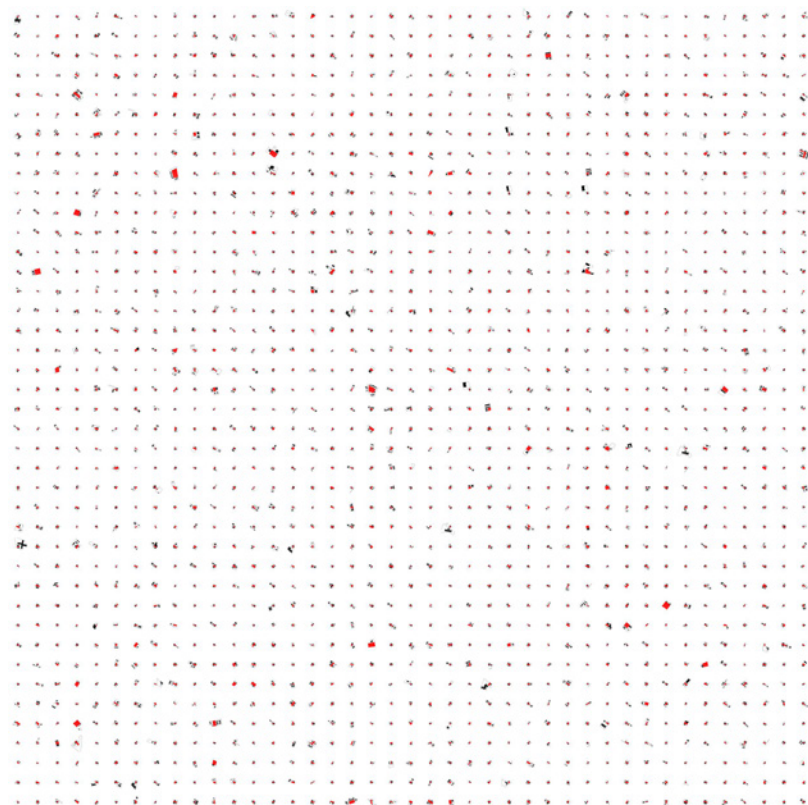
Appendix 2

15,284 images of
DID for Seoul



Appendix 3

2,080 images of DID
for Shadyside





Appendix 4

45,852 images of DID-PGH, arranged by neighborhood's density

Appendix 5

45,852 images of DID-
PGH, arranged by target
building's density





Appendix 6

45,852 images of DID-PGH, arranged by target building's height

Appendix 7

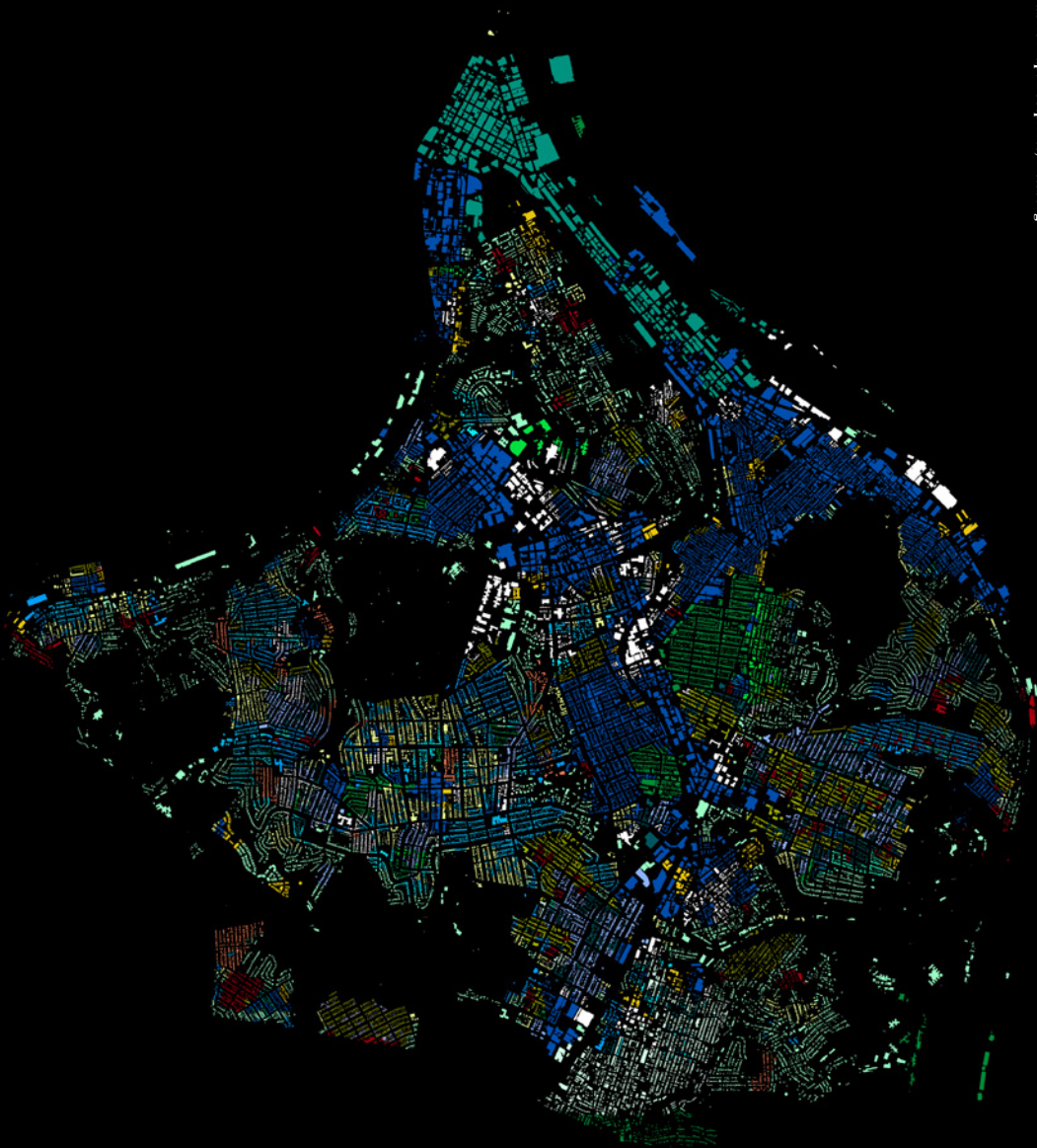
45,852 images of DID-
PGH, arranged by the
direction from West to East





Appendix 8

45,852 images of DID-PGH, arranged by the direction from South to North



Date-Driven Urban Fabric Map: Cluster Boundaries, Pittsburgh



SCALE 1:20,000

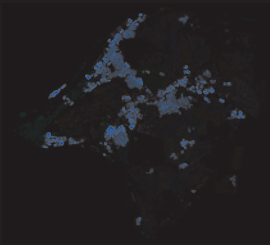
Appendix 11

Typical conditions of the clusters can be used to describe the characteristics of each urban fabrics

Cluster 1

Occupancy Rate (%):	13.82
Occupancy Rank:	2
AVR. Height (m):	6.69
AVR. Area (m²):	153.35
AVR. Density:	2.27
AVR. Building Occ.	0.34
AVR. Stories	2.23

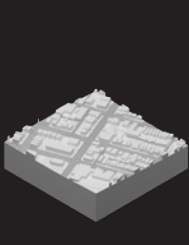
This urban fabric is a street-centered and orthogonal intersecting space with different sizes of urban blocks. This is mainly distributed on the Lawrenceville, Bloomfield, Shadyside, and Oakland. This urban fabric contains different types of streets by their width with hierarchy. There are various sizes of buildings along the main streets, and on the hinterland, there is a series of typical buildings along the minor streets with little side gap.



Building Footprints of Cluster 1



Urban Fabric Boundaries of Cluster 1

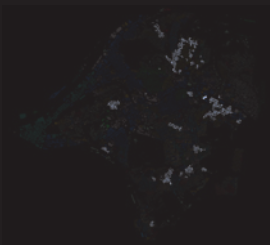


Sample 3D models of Cluster 1

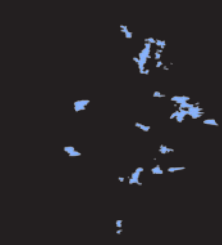
Cluster 2

Occupancy Rate (%):	3.56
Occupancy Rank:	9
AVR. Height (m):	7.06
AVR. Area (m²):	195.48
AVR. Density:	2.19
AVR. Building Occ.	0.31
AVR. Stories	2.35

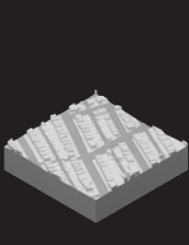
This urban fabric consists of blocks parallel to the streets. This is mainly distributed on the Highland Park Historic District, Point Breeze, and Upper Hill. The blocks in this urban fabric have rectangle shapes along the main streets and two rows of parcels. A road with a lower hierarchy than the road surrounding the block crosses a block and splits two rows of parcels.



Building Footprints of Cluster 2



Urban Fabric Boundaries of Cluster 2

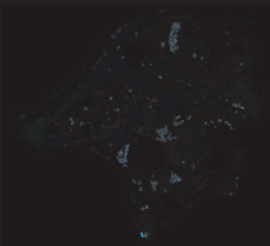


Sample 3D models of Cluster 2

Cluster 3

Occupancy Rate (%):	4.61
Occupancy Rank:	7
AVR. Height (m):	6.94
AVR. Area (m²):	175.24
AVR. Density:	2.71
AVR. Building Occ.	0.39
AVR. Stories	2.31

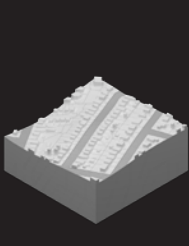
This urban fabric consists of blocks parallel to the streets. This is mainly distributed on the Squirrel Hill, Oakland Historic District, and Green Field. The blocks in this urban fabric have elongated rectangle shapes along the main streets and two rows of parcels. Instead of a street that traverses a block, it has a mirrored building arrangement referenced to the block center axis. The building back decks in a block of this urban fabric are facing each other.



Building Footprints of Cluster 3



Urban Fabric Boundaries of Cluster 3

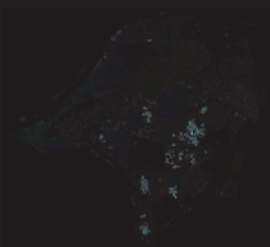


Sample 3D models of Cluster 3

Cluster 4

Occupancy Rate (%):	2.09
Occupancy Rank:	13
AVR. Height (m):	7.15
AVR. Area (m²):	167.56
AVR. Density:	2.29
AVR. Building Occ.	0.32
AVR. Stories	2.38

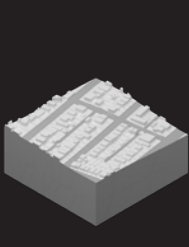
This urban fabric consists of blocks parallel to the streets. This is mainly distributed on the Squirrel Hill and Green Field. The blocks in this urban fabric are U-shaped. Like blocks in other urban fabrics, they consist of two rows of parcels, but they tend to form dead ends inside of themselves. In addition, due to these dead ends, the depth of the block is wider than other blocks. Therefore, the side of the block also has a series of buildings.



Building Footprints of Cluster 4



Urban Fabric Boundaries of Cluster 4

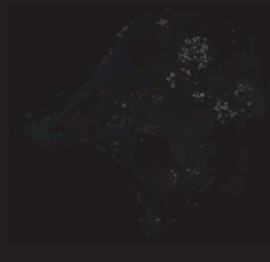


Sample 3D models of Cluster 4

Cluster 5

Occupancy Rate (%)	1.1
Occupancy Rank	16
AVR. Height (m)	6.7
AVR. Area (m ²)	156.57
AVR. Density	2.21
AVR. Building Occ.	0.33
AVR. Stories	2.23

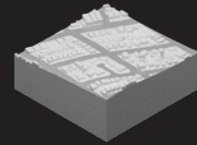
This urban fabric consists of blocks parallel to the streets. This is mainly distributed on the Highland Park and Point Breeze. The blocks in this urban fabric are a parallelogram shape due to the shape of the surrounding streets. The blocks in this urban fabric has two rows of parcels and some of them has a series of buildings along the side of them. The building back decks in a block of this urban fabric are facing each other.



Building Footprints of Cluster 5



Urban Fabric Boundaries of Cluster 5

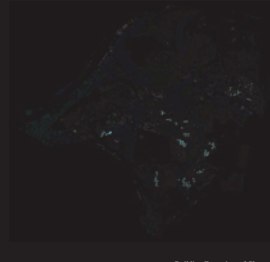


Sample 3D models of Cluster 5

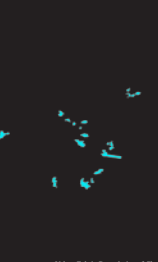
Cluster 6

Occupancy Rate (%)	0.96
Occupancy Rank	17
AVR. Height (m)	7.28
AVR. Area (m ²)	152.76
AVR. Density	10.56
AVR. Building Occ.	1.45
AVR. Stories	2.43

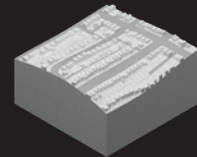
This urban fabric consists of blocks parallel to the streets. This is mainly distributed on the Squirrel Hill. The amount of this urban fabric throughout the city is very small, but its characteristics are clear. This urban fabric has arrays that form a simple urban space lined up along a single width of a street. Because of its long length, minor streets are created across the inside of the blocks. In addition, while other fabrics are following the streets with the diagonal directions, this urban fabric is formed on the streets with the orthogonal directions.



Building Footprints of Cluster 6



Urban Fabric Boundaries of Cluster 6

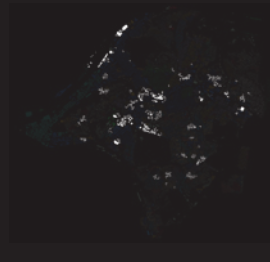


Sample 3D models of Cluster 6

Cluster 7

Occupancy Rate (%)	5.31
Occupancy Rank	5
AVR. Height (m)	6.75
AVR. Area (m ²)	145.02
AVR. Density	3.17
AVR. Building Occ.	0.47
AVR. Stories	2.25

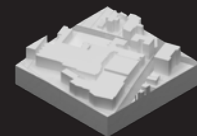
Since the main elements of this urban fabrics are not housing buildings, it has a relatively higher density than other urban fabrics. There is no generalized shape that can be specified for this urban fabric, each block in it is represented by the shape of buildings such as a hospital, school, government office, shopping mall, etc. Moreover, this urban fabric is located where the boundaries are ambiguous mainly due to the conflicts and collapse of urban fabrics. This urban fabric can be called the 'fragmented' urban fabric which can be developed to the 'urban island'.



Building Footprints of Cluster 7



Urban Fabric Boundaries of Cluster 7



Sample 3D models of Cluster 7

Cluster 8

Occupancy Rate (%)	4.1
Occupancy Rank	8
AVR. Height (m)	28.46
AVR. Area (m ²)	166.3
AVR. Density	27.89
AVR. Building Occ.	0.98
AVR. Stories	11.58

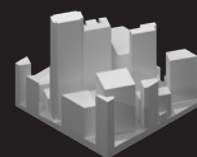
This urban fabric consists of typical blocks with grid systems. This is mainly distributed on the Downtown and Strip District. Like Cluster 7, there is no generalized shape that can be specified for this urban fabric, each block in it is represented by the shape of buildings. However, unlike Cluster 7, high density buildings of this fabric are arrayed within a grid system.



Building Footprints of Cluster 8



Urban Fabric Boundaries of Cluster 8

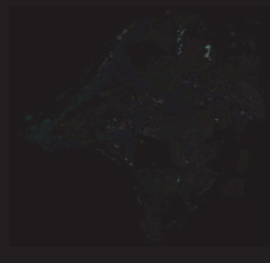


Sample 3D models of Cluster 8

Cluster 9

Occupancy Rate (%)	0.2
Occupancy Rank	20
AVR. Height (m)	6.72
AVR. Area (m ²)	121.03
AVR. Density	3.16
AVR. Building Occ.	0.47
AVR. Stories	2.24

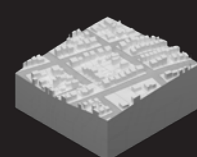
This urban fabric consists of blocks with four edges surrounding by streets. This is mainly distributed on the Highland Park. The amount of this urban fabric throughout the city is very small, but its characteristics are clear. While the blocks in other urban fabric have generally two sides with a series of buildings, the blocks in this urban fabric have four sides with a series of buildings.



Building Footprints of Cluster 9



Urban Fabric Boundaries of Cluster 9



Sample 3D models of Cluster 9

Cluster 10

Occupancy Rate (%):	0.31
Occupancy Rank:	19
AVR. Height (m):	6.49
AVR. Area (m ²):	130.12
AVR. Density:	1.49
AVR. Building Occ.	0.23
AVR. Stories	2.16

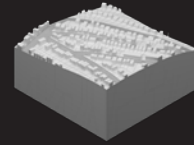
This urban fabric consists of blocks parallel to the streets. This is mainly distributed on the Green Field and Polish Hill. The amount of this urban fabric throughout the city is very small. The blocks in this urban fabric are linear polygon shapes along the streets. Since the streets are responding to the terrains of hills, they are not linear but irregular shapes. Therefore, the blocks in this fabric are also bent, crossed, broken, and irregularly shaped along the way.



Building Footprints of Cluster 10



Urban Fabric Boundaries of Cluster 10

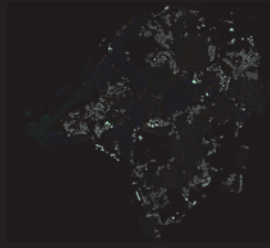


Sample 3D models of Cluster 10

Cluster 11

Occupancy Rate (%):	33.25
Occupancy Rank:	1
AVR. Height (m):	6.88
AVR. Area (m ²):	161.43
AVR. Density:	3.37
AVR. Building Occ.	0.49
AVR. Stories	2.29

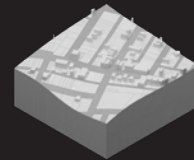
This urban fabric represents places where a certain number of empty parcels exists and no distinctive shape of blocks. A typical example of this urban fabric is Middle Hill. There are a lot of empty parcels in that neighborhood, and a building is adjacent to some of them.



Building Footprints of Cluster 11



Urban Fabric Boundaries of Cluster 11

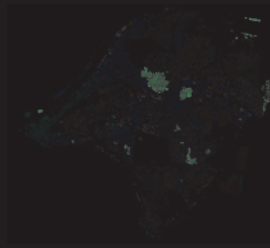


Sample 3D models of Cluster 11

Cluster 12

Occupancy Rate (%):	5.45
Occupancy Rank:	4
AVR. Height (m):	6.33
AVR. Area (m ²):	128.74
AVR. Density:	1.90
AVR. Building Occ.	0.3
AVR. Stories	2.11

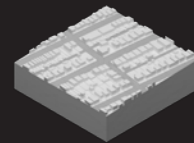
This urban fabric consists of blocks parallel to the streets. This is mainly distributed on the Bloomfield and ShadySide. The blocks in this urban fabric have rectangle shapes along the main streets and two rows of parcels. They look similar with Cluster 2's blocks, but parcels in these blocks have more longer depth and a building's occupancy of a parcel is lower than Cluster 2.



Building Footprints of Cluster 12



Urban Fabric Boundaries of Cluster 12

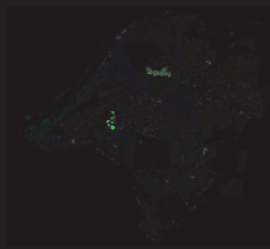


Sample 3D models of Cluster 12

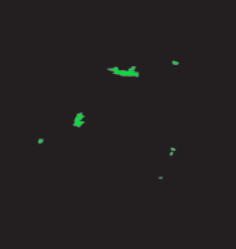
Cluster 13

Occupancy Rate (%):	0.72
Occupancy Rank:	18
AVR. Height (m):	7.13
AVR. Area (m ²):	281.07
AVR. Density:	6.84
AVR. Building Occ.	0.96
AVR. Stories	2.38

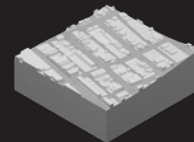
This urban fabric consists of blocks parallel to the streets. This is mainly distributed on the Garfield. This urban fabric consists of Long-shaped buildings that occupy several parcels and buildings which have large footprints within a parcel. In particular the commercial area in the Garfield is the typical example of this urban fabric's architectural contexts conditions.



Building Footprints of Cluster 13



Urban Fabric Boundaries of Cluster 13



Sample 3D models of Cluster 13

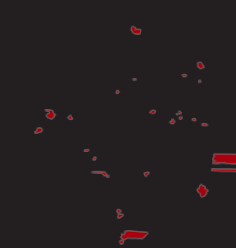
Cluster 14

Occupancy Rate (%):	2.18
Occupancy Rank:	12
AVR. Height (m):	7.19
AVR. Area (m ²):	159.99
AVR. Density:	2.37
AVR. Building Occ.	0.33
AVR. Stories	2.4

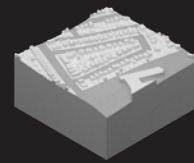
This urban fabric consists of blocks that forms inside space of themselves. The inside space of a block can have another block or can be empty. Compared to other fabrics, building size is similar with them, but block size is little bit huge because the blocks surround other space of themselves.



Building Footprints of Cluster 14



Urban Fabric Boundaries of Cluster 14



Sample 3D models of Cluster 14

Cluster 15

Occupancy Rate (%)	1.46
Occupancy Rank	15
AVR. Height (m)	7.69
AVR. Area (m ²)	234.63
AVR. Density	3.00
AVR. Building Occ.	0.39
AVR. Stories	2.56

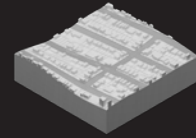
This urban fabric consists of blocks parallel to the streets. This is mainly distributed on the Green Field and Swedish Park. Like Cluster 4, the blocks in this urban fabric are U-shaped. The side of the block also has a series of buildings. But there are no streets crossing the blocks.



Building Footprints of Cluster 15



Urban Fabric Boundaries of Cluster 15



Sample 3D models of Cluster 15

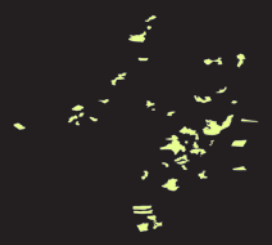
Cluster 16

Occupancy Rate (%)	5.19
Occupancy Rank	6
AVR. Height (m)	7.13
AVR. Area (m ²)	172.48
AVR. Density	6.92
AVR. Building Occ.	0.97
AVR. Stories	2.38

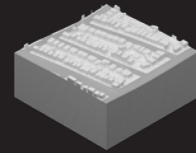
This urban fabric consists of blocks parallel to the streets. This is mainly distributed on the Squirrel Hill, Green Field, and Hazel Wood. Like Cluster 3, the blocks in this urban fabric have elongated rectangle shapes along the main streets and two rows of parcels. But it has shorter depth of blocks and more compact shape because houses of two rows forming back decks are closer to each other than cluster 3.



Building Footprints of Cluster 16



Urban Fabric Boundaries of Cluster 16

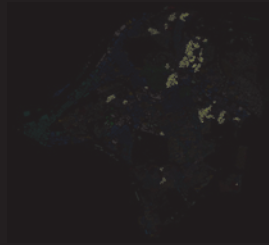


Sample 3D models of Cluster 16

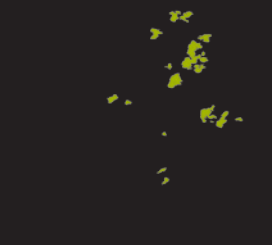
Cluster 17

Occupancy Rate (%)	8
Occupancy Rank	3
AVR. Height (m)	7.12
AVR. Area (m ²)	149.72
AVR. Density	2.06
AVR. Building Occ.	0.29
AVR. Stories	2.37

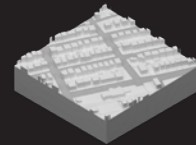
This urban fabric consists of blocks parallel to the streets. This is mainly distributed on the Highland Park, East Liberty and Friendship. Like Cluster 2 and Cluster 16, the blocks in this urban fabric have rectangle shapes along the main streets and two rows of parcels. However, unlike Cluster 2, there are no streets crossing the blocks and splitting them into two rows of parcels. While it seems similar to Cluster 16 regarding to the depth of blocks, its width is shorter than Cluster 16.



Building Footprints of Cluster 17



Urban Fabric Boundaries of Cluster 17

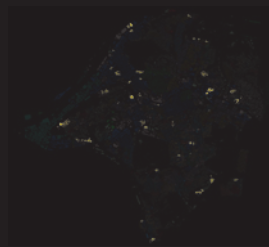


Sample 3D models of Cluster 17

Cluster 18

Occupancy Rate (%)	1.84
Occupancy Rank	14
AVR. Height (m)	6.62
AVR. Area (m ²)	122.86
AVR. Density	1.85
AVR. Building Occ.	0.28
AVR. Stories	2.21

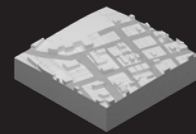
This urban fabric has basically same with the Cluster 7. But this urban fabric has shorter and more fragmented blocks than Cluster 7 and more empty spaces due to the smaller parcel and building size. In addition, it tends to exist around big buildings and Cluster 7. This is another type of the 'fragmented' urban fabric.



Building Footprints of Cluster 18



Urban Fabric Boundaries of Cluster 18

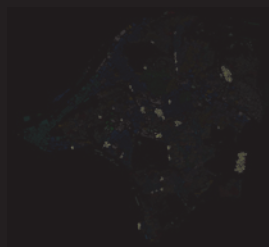


Sample 3D models of Cluster 18

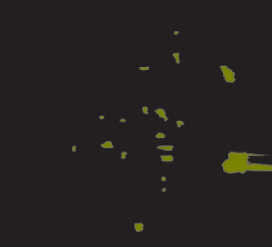
Cluster 19

Occupancy Rate (%)	3.07
Occupancy Rank	10
AVR. Height (m)	6.69
AVR. Area (m ²)	153.25
AVR. Density	2.34
AVR. Building Occ.	0.35
AVR. Stories	2.23

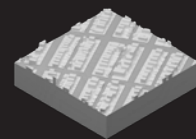
This urban fabric has basically same with the Cluster 17. Only the difference between this urban fabric and Cluster 17 is the streets crossing the blocks. This is mainly distributed on the Larimer, ShadySide, and Regent Square.



Building Footprints of Cluster 19



Urban Fabric Boundaries of Cluster 19

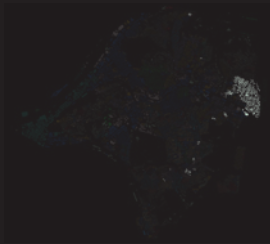


Sample 3D models of Cluster 19

Cluster 20

Occupancy Rate (%):	2.66
Occupancy Rank:	11
AVR. Height (m):	6.68
AVR. Area (m ²):	124.19
AVR. Density:	5.34
AVR. Building Occ.	0.8
AVR. Stories	2.23

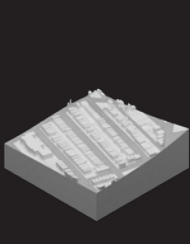
This urban fabric consists of blocks parallel to the streets. This is the main urban fabric of the Homewood. The form of the block is similar to Cluster 16, but the variety of building shapes is lower than Cluster 16. One building's foot print looks uniform to adjacent buildings.



Building Footprints of Cluster 20



Urban Fabric Boundaries of Cluster 20

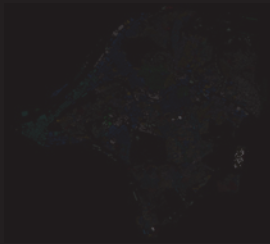


Sample 3D models of Cluster 20

Cluster 21

Occupancy Rate (%):	0.13
Occupancy Rank:	21
AVR. Height (m):	6.74
AVR. Area (m ²):	226.47
AVR. Density:	7.14
AVR. Building Occ.	1.06
AVR. Stories	2.25

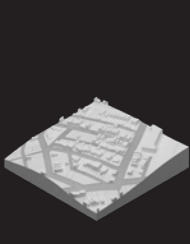
This urban fabric has the lowest occupancy rate in Pittsburgh. It mainly exists on Regent Square. Especially, this urban fabric coexists with the Cluster 19. It has diagonal and irregular forms of blocks. These blocks fill the portion of the cluster 19 when it is bent or split.



Building Footprints of Cluster 21



Urban Fabric Boundaries of Cluster 21



Sample 3D models of Cluster 21

References

1. Albert, Adrian, Jasleen Kaur, and Marta C. Gonzalez. 2017. "Using Convolutional Networks and Satellite Imagery to Identify Patterns in Urban Environments at a Large Scale." In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '17*, 1357–66. Halifax, NS, Canada: ACM Press. <https://doi.org/10.1145/3097983.3098070>.
2. Bakunowicz, A. 2013. "NEURAL SELF-ORGANISING MAPS AND GENETIC ALGORITHM: EVOLVING 3D CELLULAR AUTOMATA ARCHITECTURAL MODEL," 134.
3. Bard, Joshua, Madeline Gannon, Zachary Jacobson-Weaver, Michael Jeffers, Brian Smith, and Mauricio Contreras. 2014. "Seeing Is Doing: Synthetic Tools for Robotically Augmented Fabrication in High-Skill Domains." In *ACADIA 14: Design Agency [Proceedings of the 34th Annual Conference of the Association for Computer Aided Design in Architecture (ACADIA) ISBN 9781926724478]* Los Angeles 23-25 October, 2014), Pp. 409-416. CUMINCAD. http://papers.cumincad.org/cgi-bin/works/paper/acadia14_409.
4. Bayraktar, Mehmet Emin, and Gülen Çağdaş. 2013. "Fuzzy Layout Planner: A Simple Layout Planning Tool for Early Stages of Design." In *Computation and Performance: Proceedings of the 31st ECAADe Conference*, edited by Stouffs Rudi and Sariyildiz Sevil, 1:375–81. Delft: TUDelft.
5. Bazalo, Frano, and Tane J. Moleta. 2015. "Responsive Algorithms." In *Emerging Experience in Past, Present and Future of Digital Architecture: Proceedings of the 20th CAADRIA Conference*, edited by Y. Ikeda, C. M. Herr, D. Holzer, S. Kaijima, M. J. Kim, and M. A. Schnabel, 209–18. Hong Kong: CAADRIA.
6. Becker, Susanne, Michael Peter, Dieter Fritsch, Damian Philipp, Patrick Baier, and Christoph Dibak. 2013. *Combined Grammar for the Modeling of Building Interiors*. Vol. II-4/W1. <https://doi.org/10.5194/isprsannals-II-4-W1-1-2013>.
7. Beigel, Florian. 2010. *Architecture as City: Saemangeum Island City*. Vienna ; Springer.
8. Boon, Christopher, Corey Griffin, Nicholas Papaefthimious, ZGF Architects, Jonah Ross, and Kip Storey. 2015. "OPTIMIZING SPATIAL ADJACENCIES USING EVOLUTIONARY PARAMETRIC TOOLS: Using Grasshopper and Galapagos to Analyze, Visualize, and Improve Complex Architectural Programming." *SPECIAL ISSUE: FUTURE OF ARCHITECTURAL RESEARCH* 07 (02): 14.
9. Brandão, Filipe, and Alexandra Paio Ricardo Correia. 2017. "Rhythms of Renewal of the City." In *SIGraDi 2017 [Proceedings of the 21th Conference of the Iberoamerican Society of Digital Graphics - ISBN: 978-956-227-439-5]* Chile, Concepción 22 - 24 November 2017, Pp.534-540. CUMINCAD. http://papers.cumincad.org/cgi-bin/works/paper/sigradi2017_078.
10. Caldas, Luisa. 2001. "An Evolution-Based Generative Design System: Using Adaptation to Shape Architectural Form." *Doctor of Philosophy in Architecture: Building Technology*, Cambridge, Massachusetts: MASSACHUSETTS INSTITUTE OF TECHNOLOGY.
11. Cardoso Llach, Daniel. 2015. *Builders of the Vision: Software and the Imagination of Design*. 1st ed. Routledge. <https://doi.org/10.4324/9781315798240>.
12. Colaninno, Nicola, Josep Roca Cladera, and Karin Pfeffer. 2011. "An Automatic Classification of Urban Texture: Form and Compactness of Morphological Homogeneous Structures in Barcelona," 16.
13. Cowan, Robert. 2005. *The Dictionary of Urbanism*. Streetwise Press.
14. Cross, Nigel. 1977. *The Automated Architect. Research in Planning and Design* ; 4. London: Pion.
15. Daly, Janet R. 1988. "Zoning: Its Historical Context and Importance in the Development of Pittsburgh." *Western Pennsylvania History: 1918 - 2018* 71 (2): 99-126–126. <https://journals.psu.edu/wph/article/view/4103>.
16. Damski, José C., and John S. Gero. 1997. "An Evolutionary Approach to Generating Constraint-Based Space Layout Topologies." In *CAAD Futures 1997*, edited by Richard Junge, 855–64. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-011-5576-2_65.

17. Danil Nagy, Lorenzo Villaggi, Dale Zhao, and David Benjamin. 2017. "A Novel Design Space Model for Generative Space Planning in Architecture," 10.
18. Das, Subhajit, Colin Day, John Hauck, John Haymaker, and Diana Davis. 2016. "Space Plan Generator: Rapid Generation & Evaluation of Floor Plan Design Options to Inform Decision Making." In *ACADIA // 2016: POSTHUMAN FRONTIERS: Data, Designers, and Cognitive Machines* [Proceedings of the 36th Annual Conference of the Association for Computer Aided Design in Architecture (ACADIA) ISBN 978-0-692-77095-5] Ann Arbor 27-29 October, 2016, Pp. 106-115. CUMINCAD. http://papers.cumincad.org/cgi-bin/works/paper/acadia16_106.
19. Deleuze, Gilles. 1993. *The Deleuze Reader*. Edited by Constantin V. Boundas. 1st Edition edition. New York: Columbia Univ Pr.
20. Donath, Dirk, Danny Lobos, and Bauhaus-University Weimar. 2008. "Massing Study Support, A New Tool for Early Stages of Architectural Design." *Prediction and Evaluation - ECAADe 26*, 8.
21. Eastman, Charles M. 1971. "HEURISTIC ALGORITHMS FOR AUTOMATED SPACE PLANNING," 13.
22. Eisenman, Peter. 1999. *Diagram Diaries*. New York: Universe.
23. Ena, Valeria. 2018. "De-Coding Rio de Janeiro's Favelas - Shape Grammar Application as a Contribution to the Debate over the Regularisation of Favelas. The Case of Parque Royal." In *Kepeczynska-Walczak, A, Bialkowski, S (Eds.), Computing for a Better Tomorrow - Proceedings of the 36th ECAADe Conference - Volume 2*, Lodz University of Technology, Lodz, Poland, 19-21 September 2018, Pp. 429-438. CUMINCAD. http://papers.cumincad.org/cgi-bin/works/paper/ecaade2018_336.
24. Evans, Henry Oliver. 1941. "Notes on Pittsburgh Transportation Prior to 1890." *Western Pennsylvania History: 1918 - 2018* 24 (3): 161-182-182. <https://journals.psu.edu/wph/article/view/2198>.
25. Fernando, Ruwan. 2014. "Space Planning and Preliminary Design Using Artificial Life." In *Rethinking Comprehensive Design: Speculative Counterculture: Proceedings of the 19th CAADRIA Conference*, edited by Ning Gu, Shun Watanabe, Halil Erhan, Matthias Hank Haeusler, Weixin Huang, and Ricardo Sosa, 657-66. Hong Kong: CAADRIA.
26. Flemming, Ulrich. 1989. "More On the Representation and Generation of Loosely Packed Arrangements of Rectangles" 16: 359. <https://doi.org/10.1068/b160327>.
27. Frew, Margaret A. 1929. "Squirrel Hill." *Western Pennsylvania History: 1918 - 2018* 12 (4): 242-256-256. <https://journals.psu.edu/wph/article/view/1498>.
28. Garcia, Mark. 2010. *The Diagrams of Architecture: AD Reader*. Chichester: Wiley.
29. Gero, John S., and Vladimir A. Kazakov. 1997. "Learning and Re-Using Information in Space Layout Planning Problems Using Genetic Engineering." *Artificial Intelligence in Engineering* 11 (3): 329-34. [https://doi.org/10.1016/S0954-1810\(96\)00051-9](https://doi.org/10.1016/S0954-1810(96)00051-9).
30. Harding, John. 2016. "Evolving Parametric Models Using Genetic Programming with Artificial Selection," 10.
31. Hillier, Bill, and Julienne Hanson. 1984. *The Social Logic of Space*. Cambridge [Cambridgeshire]: Cambridge University Press. <http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=711679&site=ehost-live>.
32. Hua, Hao, and Ting-Li Jia. 2010. "An Agent-Based System for Layout Planning," 9.
33. Ireland, Tim. 2015. "A Cell Inspired Model of Configuration." In *Computational Ecologies: Design in the Anthropocene: Proceedings of 35th ACADIA Conference*, edited by Lohns Combs and Chris Perry, 136-47. New York: ACADIA.
34. Jiang, Shan, Joseph Ferreira, and Marta C. Gonzalez. 2012. "Discovering Urban Spatial-Temporal Structure from Human Activity Patterns." *Other University Web Domain*, August. <http://dspace.mit.edu/handle/1721.1/77151>.
35. Jolliffe, I. T. 1986. *Principal Component Analysis*. Springer Series in Statistics. New York, NY:

Springer New York. <https://doi.org/10.1007/978-1-4757-1904-8>.

36. Kingma, Diederik P., and Max Welling. 2013. "Auto-Encoding Variational Bayes." ArXiv:1312.6114 [Cs, Stat], December. <http://arxiv.org/abs/1312.6114>.
37. Koenig, Reinhard., and Katja Knecht. 2010. "Generating Floor Plan Layouts with K-d Trees and Evolutionary Algorithms." <https://www.researchgate.net/publication/256471356>.
38. Koenig, Reinhard, and Katja Knecht. 2014a. "Comparing Two Evolutionary Algorithm Based Methods for Layout Generation: Dense Packing versus Subdivision." *AI EDAM* 28 (3): 285–299.
39. Koenig, Reinhard, and Katja Knecht. 2014b. "Comparing Two Evolutionary Algorithm Based Methods for Layout Generation: Dense Packing versus Subdivision." *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 28 (03): 285–99. <https://doi.org/10.1017/S0890060414000237>.
40. Kropf, Karl. 1996. "Urban Tissue and the Character of Towns." *URBAN DESIGN International* 1 (3): 247–63. <https://doi.org/10.1057/udi.1996.32>.
41. KUO, CHUNG-JEN. 2003. "SPATIAL ANALYSIS OF CHINESE GARDEN DESIGNS WITH MACHINE LEARNING."
42. Lee, Jae-Beom, and Suk-Chan Ko. 2009. "An Analysis of the Factors Affecting Apartment Rent/ Price Ratio in Seoul Area." *한국지역개발학회지* 21 (1): 113–28. <http://www.dbpia.co.kr/Article/NODE01879514>.
43. Mathias, M., A. Martinovic, J. Weissenberg, S. Haegler, and L. Van Gool. 2012. "AUTOMATIC ARCHITECTURAL STYLE RECOGNITION." *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII-5/W16 (September)*: 171–76. <https://doi.org/10.5194/isprsarchives-XXXVIII-5-W16-171-2011>.
44. Merrell, Paul, Eric Schkufza, and Vladlen Koltun. 2010. "Computer-Generated Residential Building Layouts," 11.
45. Michalek, Jeremy, Ruchi Choudhary, and Panos Papalambros. 2002. "Architectural Layout Design Optimization." *Engineering Optimization* 34 (5): 461–84. <https://doi.org/10.1080/03052150214016>.
46. Michalek, Jeremy J. 2001. "Interactive Layout Design Optimization." Master of Science, University of Michigan.
47. Michalek, Jeremy, and Panos Papalambros. 2002. "Interactive Design Optimization of Architectural Layouts." *Engineering Optimization* 34 (5): 485–501. <https://doi.org/10.1080/03052150214021>.
48. Mirahmadi, Maysam, and Abdallah Shami. 2012. "A Novel Algorithm for Real-Time Procedural Generation of Building Floor Plans." ArXiv:1211.5842 [Cs], November. <http://arxiv.org/abs/1211.5842>.
49. Moon, Junsik. 2007. "Shape Grammar for Mies van Der Rohe's High-Rise Apartment." Master of Science in Architecture Studies, Cambridge: MASSACHUSETTS INSTITUTE OF TECHNOLOGY.
50. Narahara, Taro. 2017. "Collective Construction Modeling and Machine Learning: Potential for Architectural Design." In Fioravanti, A, Cursi, S, Elahmar, S, Gargaro, S, Loffreda, G, Novembri, G, Trento, A (Eds.), *ShoCK! - Sharing Computational Knowledge! - Proceedings of the 35th ECAADe Conference - Volume 2*, Sapienza University of Rome, Rome, Italy, 20-22 September 2017, Pp. 341-348. CUMINCAD. http://papers.cumincad.org/cgi-bin/works/paper/ecaade2017_271.
51. Negroponte, Nicholas. 1970. *The Architecture Machine*. Cambridge, Mass: MITPress.
52. Pai, Hyungmin. 2002. *The Portfolio and the Diagram: Architecture, Discourse, and Modernity in America*. Cambridge, Mass: MIT Press.
53. Parish, Yoav I H, and Pascal Müller. n.d. "Procedural Modeling of Cities," 8.
54. Pasternak, Agata. 2016. "Optimization of the Building in Relation to the Insolation Conditions of Premises in Adjacent Buildings." *SHAPE, FORM AND GEOMETRY | Grammars and Concepts* 2: 8.
55. Pellitteri, Giuseppe, Raimondo Lattuca, and Giuseppe Conti. 2010. "A Generative Design System

to Interactively Explore Different Urban Scenarios.” FUTURE CITIES 28th ECAADe Conference-City Modelling, 10. https://www.researchgate.net/publication/323245610_A_Generative_Design_System_to_Interactively_Explore_Different_Urban_Scenarios.

56. “Price | Economics.” n.d. Encyclopedia Britannica. Accessed April 10, 2019. <https://www.britannica.com/topic/price-economics>.

57. Puusepp, Renee. 2014. “Agent-Based Models for Computing Circulation.” In *Design Agency: Proceedings of the 34th ACADIA Conference*, edited by David Gerber, Alvin Huang, and Jose Sanchez, 43–52. Cambridge: Riverside Architectural Press.

58. RIBA Journal. n.d. “Interview with RIBA, Q&A: Florian Beigel.” Accessed April 4, 2019. <https://www.ribaj.com/intelligence/q-a-florian-beigel>.

59. Rossi, Aldo. 1982. *The Architecture of the City*. Cambridge, Massachusetts, and London, England: MIT Press. <https://mitpress.mit.edu/books/architecture-city>.

60. Roweis, Sam T., and Lawrence K. Saul. 2000. “Nonlinear Dimensionality Reduction by Locally Linear Embedding.” *Science* 290 (5500): 2323–26. <https://doi.org/10.1126/science.290.5500.2323>.

61. S Liggett, Robin. 2000. *Automated Facilities Layout: Past, Present and Future*. Vol. 9. [https://doi.org/10.1016/S0926-5805\(99\)00005-9](https://doi.org/10.1016/S0926-5805(99)00005-9).

62. Schmitt, Gerhard. 1988. “ARCHPLAN: An Architectural Planning Front End to Engineering Design Expert Systems.” In *Expert Systems for Engineering Design*, 257–78. Elsevier. <https://doi.org/10.1016/B978-0-12-605110-0.50017-3>.

63. Schusler, William Kenneth. 1960. “The Railroad Comes To Pittsburgh.” *Western Pennsylvania History: 1918 - 2018* 43 (3): 251-266–266. <https://journals.psu.edu/wph/article/view/2704>.

64. Sevtsuk, Andres. 2010. “Path and Place : A Study of Urban Geometry and Retail Activity in Cambridge and Somerville, MA.” Thesis, Massachusetts Institute of Technology. <http://dspace.mit.edu/handle/1721.1/62034>.

65. Shane, David Grahame. 2005. *Recombinant Urbanism: Conceptual Modeling in Architecture, Urban Design, and City Theory*. Chichester, England ; Wiley.

66. Silvestre, Joaquim, Yasushi Ikeda, and FRANÇOIS GUÉNA. 2016. “ARTIFICIAL IMAGINATION OF ARCHITECTURE WITH DEEP CONVOLUTIONAL NEURAL NETWORK,” 10.

67. Simonyan, Karen, and Andrew Zisserman. 2014. “Very Deep Convolutional Networks for Large-Scale Image Recognition.” *ArXiv:1409.1556 [Cs]*, September. <http://arxiv.org/abs/1409.1556>.

68. Stiny, G, and W J Mitchell. 1978. “The Palladian Grammar.” *Environment and Planning B: Planning and Design* 5 (1): 5–18. <https://doi.org/10.1068/b050005>.

69. Strobbe, Tiemen, Francis wyffels, Ruben Verstraeten, Ronald De Meyer, and Jan Van Campenhout. 2016. “Automatic Architectural Style Detection Using One-Class Support Vector Machines and Graph Kernels.” *Automation in Construction* 69 (September): 1–10. <https://doi.org/10.1016/j.autcon.2016.05.024>.

70. Ulrich Flemming, Robert Coyne, and Raymond Gindroz. 1985. *A Pattern Book For Shadyside*. 1st ed. Vol. 1. 1 vols. Pittsburgh: Department of Architecture, Carnegie-Mellon University.

71. Vardouli, Theodora. 2000. ““Architecture-by-Yourself”: Early Studies in Computer-Aided Participatory Design,” 10.

72. Veloso, Pedro. 2014. “Exploring the Bubble Diagram.” In *Design in Freedom: Proceedings of 18th SIGraDi Conference*, edited by Fernando G. Amen, 1:115–119. São Paulo: Blucher.

73. Vidler, Anthony. 2000. “Diagrams of Diagrams: Architectural Abstraction and Modern Representation.” *Representations*, no. 72: 1–20. <https://doi.org/10.2307/2902906>.

74. Vitor, Oliveira. 2016. *Urban Morphology*. New York, NY: Springer Berlin Heidelberg.

75. Weinzapfel, Guy, Timothy E. Johnson, and John Perkins. 1971. *IMAGE: An Interactive Computer*

System for Multi-Constrained Spatial Synthesis. <https://doi.org/10.1145/800158.805064>.

76. William, J Mitchell. 1977. Computer-Aided Architectural Design.

77. Zhou, Bolei, Liu Liu, Aude Oliva, and Antonio Torralba. 2014. "Recognizing City Identity via Attribute Analysis of Geo-Tagged Images." MIT Web Domain. <http://dspace.mit.edu/handle/1721.1/90999>.

List of Figures

- Fig. 1. Example of Urban tissue shown at increasing levels of resolution, Kropf, Karl. 1996. “Urban Tissue and the Character of Towns.”
- Fig. 2. The Nolli map of Rome, Giambattista Nolli (Left), Imperial Rome, views of model at the Museo della Civiltà Romana, Collage City, Colin Rowe and Fred Koetter (Middle), The matrix of 17th century Rome – plan after Bufalini (Right), https://en.wikipedia.org/wiki/Giambattista_Nolli (Left), Koetter, Fred and Rowe, Colin, 1979, *Collage City* (Middle and Right).
- Fig. 3. Definition of Urban Morphological Type with the Computational Aspect, Jinmo Rhee. 2019
- Fig. 4. Research A, Reach to built volume within a 2,000ft network radius from each building in Cambridge & Somerville, MA, Sevtsuk, Andres. 2010. “Path and Place : A Study of Urban Geometry and Retail Activity in Cambridge and Somerville, MA.”
- Fig. 5. Research B, Time-cumulative spatial densities of shopping/errand & recreation/entertainment activities in the Chicago sub-region, Jiang, Shan, Joseph Ferreira, and Marta C. Gonzalez. 2012. “Discovering Urban Spatial-Temporal Structure from Human Activity Patterns.”
- Fig. 6. Research C, Result of Cluster analysis for seven urban typologies, Colaninno, Nicola, Josep Roca Cladera, and Karin Pfeffer. 2011. “An Automatic Classification of Urban Texture: Form and Compactness of Morphological Homogeneous Structures in Barcelona.”
- Fig. 7. Research D, The map of attribute diversity and the map of sidewalk density in Boston, Zhou, Bolei, Liu Liu, Aude Oliva, and Antonio Torralba. 2014. “Recognizing City Identity via Attribute Analysis of Geo-Tagged Images.”
- Fig. 8. Research E, Barcelona: ground truth (top) and predicted probabilities (bottom), Albert, Adrian, Jasleen Kaur, and Marta C. Gonzalez. 2017. “Using Convolutional Networks and Satellite Imagery to Identify Patterns in Urban Environments at a Large Scale.”
- Fig. 9. Research F, Map of distribution and length of permit approval process in the center of Lisbon, Brandão, Filipe, and Alexandra Paio Ricardo Correia. 2017. “Rhythms of Renewal of the City.”
- Fig. 10. The Palladian Grammar, Rule for the generation of exterior walls (Left), The underlying wall pattern for the Villa Malcontenta (Right), Stiny, G, and W J Mitchell. 1978. “The Palladian Grammar.”
- Fig. 11. Architectural Layout Design Optimization, Sample design topology and final geometry generated by the automated design tool, Michalek, Jeremy, Ruchi Choudhary, and Panos Papalambros. 2002. “Architectural Layout Design Optimization.”
- Fig. 12. Computer-generated Building Layout, architectural programs illustrated by a bubble diagram (Left) and a set of floor plans (Right), Merrell, Paul, Eric Schkufza, and Vladlen Koltun. 2010. “Computer-Generated Residential Building Layouts,”
- Fig. 13. URBAN5, A moment of a sequence of frames that depict travel through an environment constructed jointly by the architect and the machine, Negroponte, Nicholas. 1970. *The Architecture Machine*.
- Fig. 14. STUNI, The Sequence of layout modifications, Cross, Nigel. 1977. *The Automated Architect*.
- Fig. 15. Evolution-Based Generative Design System, Best solutions for Oporto using EUI as fitness function, Caldas, Luisa. 2001. “An Evolution-Based Generative Design System: Using Adaptation to Shape Architectural Form.”
- Fig. 16. Contextuality of Architectural Design with AI techniques, Jinmo Rhee. 2019
- Fig. 17. Relationships among Architect, Computer, and Diagrams, Jinmo Rhee. 2019
- Fig. 18. Diagrams as a Generative Device, examples of Staten Island diagrams, Eisenman Architects, Eisenman, Peter. 1999. *Diagram Diaries*.

- Fig. 19. Example of Vectorized Value from Raster Images, Jinmo Rhee. 2019
- Fig. 20. Example of Vectorized Value from Vector Geometries, Jinmo Rhee. 2019
- Fig. 21. Three Types of Urban Information Image: Satellite Images, Map Images, and Diagrammatic Images, Jinmo Rhee. 2019
- Fig. 22. An Example of Diagrammatic Image, Jinmo Rhee. 2019
- Fig. 23. Custom Components for Generating DID (Top) and The Grasshopper Algorithm for DID Generation (Bottom), Jinmo Rhee. 2019
- Fig. 24. Four Color Gradients for representing location information, Jinmo Rhee. 2019
- Fig. 25. Apartment complexes in Seoul, Google Maps, Pedro, Veloso, Jinmo Rhee, Eddy Man Kim. 2018, "Custom Generative Systems with Adaptation Mechanisms an Experiment with South Korean Apartment complexes"
- Fig. 26. Overall Process: Recognition of DID Seoul, Jinmo Rhee. 2019
- Fig. 27. Average Price per square meter of a Building in an Apartment Complex, Jinmo Rhee. 2019
- Fig. 28. Example Process of Sampling the data and Global-Local Prices, Jinmo Rhee. 2019
- Fig. 29. Price classification with 'k-means-clustering', Jinmo Rhee. 2019
- Fig. 30. The process of Generating DID of Seoul, Jinmo Rhee. 2019
- Fig. 31. Structures of the five different models, Jinmo Rhee. 2019
- Fig. 32. Learning Accuracy of different models, Jinmo Rhee. 2019
- Fig. 33. Sample image, focus, heatmap, and gradients, Jinmo Rhee. 2019
- Fig. 34. Visualization of Weights in Convolution Layers, Jinmo Rhee. 2019
- Fig. 35. Overall Process: Usability of DID Shadyside, Jinmo Rhee. 2019
- Fig. 36. Sample Site Range, satellite images from Google Earth, Jinmo Rhee. 2019
- Fig. 37. The Composition of the DID and Geometry Approximation for Constant Bin Size, Jinmo Rhee. 2019
- Fig. 38. Test Results of Five Different Geometry Approximation Algorithms, Jinmo Rhee. 2019
- Fig. 39. Examples of geometrical approximation by Delaunay triangulation, Jinmo Rhee. 2019
- Fig. 40. Similarity of Predicted Footprints, 98% (Left), 88% (Middle), 48% (Right), Jinmo Rhee. 2019
- Fig. 41. Training and Validation Accuracy and Loss, Jinmo Rhee. 2019
- Fig. 42. Perspective, Example of urban fabrics in Pittsburgh, Jinmo Rhee. 2019
- Fig. 43. Overall Process: a data-driven urban fabric map using DID-PGH, Jinmo Rhee. 2019
- Fig. 44. Perspective, Example of urban fabrics in Pittsburgh, Jinmo Rhee. 2019
- Fig. 45. Decreased Range of the Data, Building Footprint, Parcel, Neighborhoods, Streets, Jinmo Rhee. 2019
- Fig. 46. LAS point clouds of Pittsburgh, DSM of Pittsburgh, DEM of Pittsburgh, Jinmo Rhee. 2019
- Fig. 47. Process of Building Height Calculation through DSM and DEM, Jinmo Rhee. 2019
- Fig. 48. Grasshopper Algorithms to Import the Shape File of Building Footprints, Jinmo Rhee. 2019

- Fig. 49. A Custom Grasshopper Component to Read CSV, Jinmo Rhee. 2019
- Fig. 50. Imported Shape Files in Rhinoceros Using Grasshopper (Left) and All Imported Geometries in Rhinoceros (Left), Jinmo Rhee. 2019
- Fig. 51. Process of Setting the Image Range and Sample of Image Range, Jinmo Rhee. 2019
- Fig. 52. An Example of Diagrammatic Image Generation with Customized Drawing Style, Jinmo Rhee. 2019
- Fig. 53. Pittsburgh Neighborhood Map and Index, Jinmo Rhee. 2019
- Fig. 54. Four Different Remapping and Arrangements of DID -PGH, By Neighborhood Density, Height, West to East, South to North, Jinmo Rhee. 2019
- Fig. 55. Pittsburgh Neighborhood Map and Index, Jinmo Rhee. 2019
- Fig. 56. The latent space of VAE from DID-PGH, Jinmo Rhee. 2019
- Fig. 57. t-SNE visualization of DID-PGH based on the Similarity of Architectural Context Conditions by Neighborhoods, Jinmo Rhee. 2019
- Fig. 58. t-SNE visualization clustered by DBSCAN, Jinmo Rhee. 2019
- Fig. 59. Process of Remapping and Coloring the Building Footprints, Jinmo Rhee. 2019
- Fig. 60. Data-driven Urban Fabrics Map, Building Footprints, Jinmo Rhee. 2019
- Fig. 61. Data-driven Urban Fabric Map, Boundaries, Jinmo Rhee. 2019
- Fig. 62. Statistics on morphological characteristics by cluster, Jinmo Rhee. 2019
- Fig. 63. Typical Conditions of 21 Different Urban Fabrics from Clusters, Jinmo Rhee. 2019
- Fig. 64. Site Survey, Places Where City Fabrics Conflict, Jinmo Rhee. 2019
- Fig. 65. Pembroke Place and St. James Street, Facing South (Left) and North (Right), Pittsburgh, Jinmo Rhee. 2019
- Fig. 66. Façade of Vernacular of Shadyside (top, bottom-left), Mansion in Shadyside, Ulrich Flemming, Robert Coyne, and Raymond Gindroz. 1985. A Pattern Book For Shadyside.
- Fig. 67. Fragmented Urban Fabric (Left) and Holy Spirit Byzantine Catholic, Fifth Avenue and Clyde Street, Jinmo Rhee. 2019
- Fig. 68. Hobart Street and Wendover Street (Left) and Wightman Street and Marlborough Road (Right), Jinmo Rhee. 2019
- Fig. 69. Urban Fabrics Map with Railroad 1872 (Left) and 2017 (Right), Jinmo Rhee. 2019
- Fig. 70. 1850s Canals in Pittsburgh Evans, Henry Oliver. 1941. "Notes on Pittsburgh Transportation Prior to 1890."
- Fig. 71. The Concept of Reverse Analysis for City, Jinmo Rhee. 2019
- Fig. 72. The Concept of Urban Synthesis, Jinmo Rhee. 2019

List of Tables

Table. 1 . Analysis of precedents' Granularity and Collectivity

Table. 2 . Five Different Types of Models

Table. 3 . Learning Accuracy by Models

Table. 4 . Sources of Data

Table. 5 . Obtainable Information through DID by Scale

Table. 6 . A Matrix of the Obtainable Information through DID