Research Paper

A basic analysis on urban landscape continuity in a lowland urban heritage using deep learning-based method

M.R. Derbel [1], J. Makhlouf [2] and N. Mishima [3]

ARTICLE INFORMATION

Article history:

Received: 03 July, 2019
Received in revised form: 14 April, 2020
Accepted: 20 May, 2020
Publish on: 06 September, 2020

Keywords:

Machine learning
Urban heritage area
Façade analysis
Urban façade
Deep learning

ABSTRACT

Architectural intervention in urban heritage area is subject to numerous parameters making it a time-consuming process. Urban façade analyses are also one of the required long-term tasks held by the architect, especially in urban heritage area where pressure concerning the neighborhood harmony is often faced.

To address this issue, A computer vision method for an automatic evaluation of the urban façade is used to compare a set of façade's pictures. The target area is Hizenhamashuku, in a "preservation area of traditional buildings" located in Kashima city, which is a typical lowland city in Saga prefecture.

This project aims to explore possibilities to boost the performance of urban facades study using a deep learning method. The used algorithm is able analyze pictures of buildings from different historic eras with different historic styles, regarding any selected feature.

First, an objective feature, such as the orientation of the building which, having a unique parameter, prevent from bias and thus its results can be used as reference. Next in order, a more subjective parameter such as the quality of insertion is tested, results are quantified and compared in order to evaluate the algorithm performance and enhance it in further research.

1. Introduction

Through the last decades, our view on urban heritage preservation has drastically changed. This increase of interest is strongly related to social, environmental, and economic variables, and aims to sustain preservation areas regarding those criteria. But as the human perception of an urban environment is inherently incomplete, discontinuous, and distorted (Lynch 1960), it makes the analysis of Historic Urban Landscape more challenging for the planner.

The actual method is human skill-based observation and monitoring (Brownson, Hoehner, et al. 2009), which is quite limited since its manual nature makes it inherently time-consuming and derive few economies of scale (Harvey 2014).

Recently, the availability of new computer vision techniques, made the automatic evaluation of several Urban landscapes features possible (Doersch, Singh et al. 2012; Salesses, Schechtner, et al. 2013; Naik, Philipoom, et al. 2014; Ordonez and Berg 2014; Quercia, O'Hare, et al. 2014; Lee, Maisonneuve, et al. 2015).

Our goal in this paper is thus to explore this possibility in terms of urban heritage area preservation quality. We choose two visual features. First, the building orientation as a more objective feature will undergo the algorithm

[1] Phd candidate & IALT member, Saga University, Saga 840-8502, JAPAN, rami.derbel@gmail.com
[2] Phd candidate, Kyushu University, Fukuoka 819-0395, JAPAN, makhlouf_jihed@yahoo.fr
[3] Professor & IALT member, Saga University, Saga 840-8502, JAPAN, mishiman@cc.saga-u.ac.jp
Note: Discussion on this paper is open until March 2021

judgment as a starting point in this study. Insertion quality in the building will be studied in a second time. The first feature is studied as a forgoing step toward setting a comparative point for further researches as higher results are expected due to the inherent objective nature of the feature. The second feature is influential to the historic landscape analysis.

This research paper will focus on the preservation of urban heritage areas in Japan and takes as experiment area Hizen Hamashuku, Kashima city, and Saga city in Saga prefecture. After collecting data in both areas, we submitted the dataset to a panel of experts for rating. Then we divided the data set on two entities; the training set was submitted to the model for training, then the test set was submitted to the trained model to test its ability of prevision. The discussion of the first feature results showed us satisfying performances. The model was able to set a good pattern to recognize the orientation of the building in the test set. The second feature's result, however, was not enough relevant, therefore we needed to investigate more using data analysis. This led us to better understand points to correct or to improve in our experience protocol.

2. Features selection

2.1 Building orientation regarding the street

A continuous façade is formed when buildings are lined in a row without significant interruptions caused by vacant lots or setbacks. This urban façade offers a sense of enclosure (Ewing and Clemente 2013), majesty, and controlled uniformity (Milroy 2010), and draws pedestrians and activities. As early as the 15th century, relevant rules had appeared in street design codes in Nuremberg, Germany, which required buildings to be lined up to create an "undeviating building line" (Kostof 1992). Nowadays, it is addressed in numerous planning codes and guidelines, especially in urban heritage areas.

In the paper, we will study the orientation of the façade regarding the camera position as the first step. In further research we will address the question of its continuity using the same methodology we are using now to classify the pictures regarding their orientation.

Furthermore, this feature prediction results will be used as a used standard to compare further experiment results.

2.2 Insertion quality in an Urban Heritage Area

The architecture façade is a highly influential component of the urban space that concentrates visual attention and 'radiates' onto the urban space (Von Meiss 2013). The urban landscape is also shaped by the building façades, creating a physical limit to human sight (Buchanan 1988). In specific areas such as urban heritage area, these factors are emphasized and thus the visual quality of façade insertion is a combined effect of various factors inherent to the building as well as factors related to the preservation area. It includes non-exhaustively:

- Composition which creates visual rhythms and holds the attention (Buchanan 1988). It is formed by the repetition of constituent parts (e.g. windows, doors, and bays), the ratio of solid to void, the articulation of vertical and horizontal elements, etc. (Carmona 2010).
- Material which gives texture and pattern to the surface and applies certain visual friction to slow the eye and space (Buchanan 1988).
- Detail which holds the eye and provides interest. Space can feel harsh and inhuman if its surfaces lack fine details and visual interest, while finely detailed, space can be delicate, airy, and inviting (Carmona 2010). However, overloaded details can also have a counter effect, since too much complexity is tested to be negatively correlated with people's preference (Devlin and Nasar 1989).
- Color which evokes feelings and emotions. According to Wassily Kandinsky, each color is linked to a certain feeling, such as red to alive, restless, blue to deep, inner, supernatural, peaceful, etc. (Kandinsky 2012).
- Compatibility with neighboring is an issue for most urban fabrics but it is emphasized in urban heritage areas where usually guidelines are set up to protect this specific criterion.
- We have also to pay particular attention to the complexity of the façade (its richness), as well as the identifiability of the architectural style (its historicity).

In this paper, we will question the quality of the architectural façade in the urban heritage areas of Saga prefecture using an expert panel rating which will be based on the previously cited criteria. (Ulrich 1983)

Fig. 1. Saga Prefecture in Japan

3. Data collection and preparation

3.1 Data collection

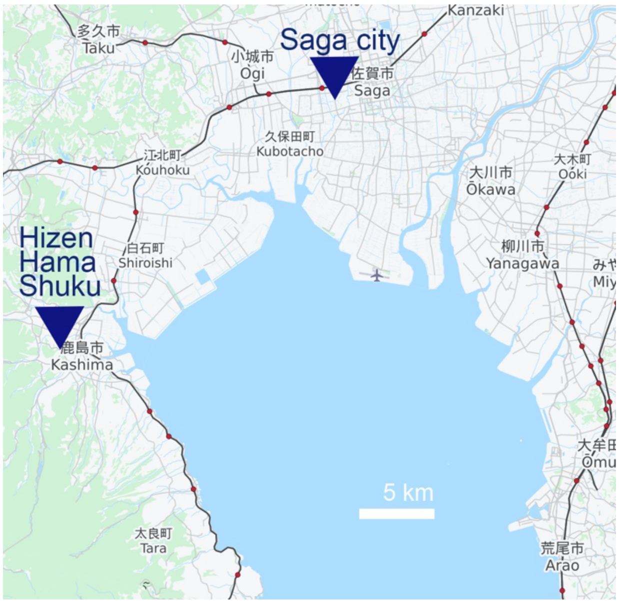

Fig. 2. Research areas in Saga prefecture

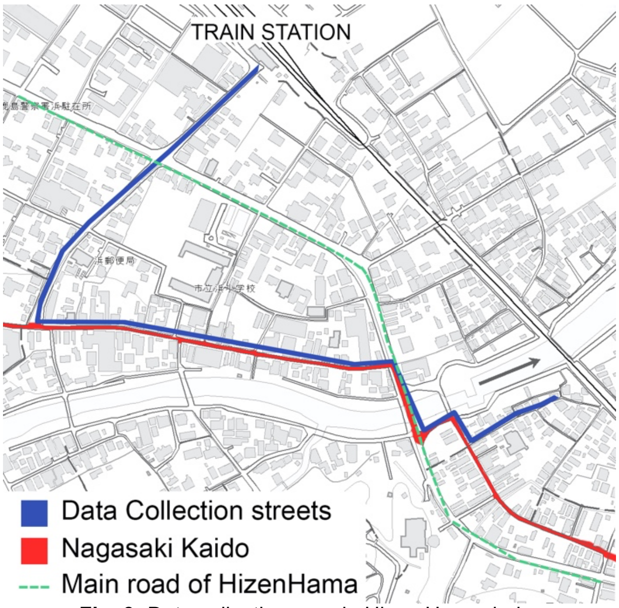

Fig. 3. Data collection area in Hizen Hamashuku

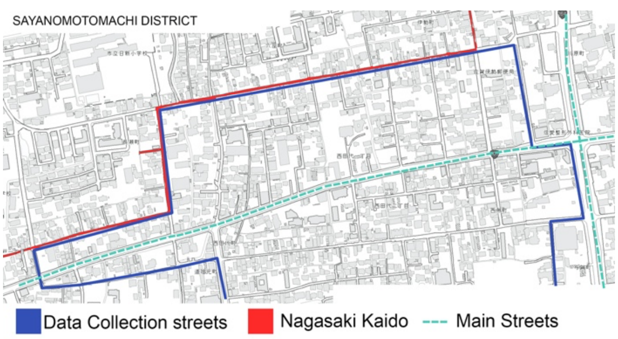

Fig. 4. Data collection area in Saga City

The dataset for image training and prediction is composed of 726 pictures. Those pictures were taken in two areas, using 2 cameras.

Our case of study is composed of two areas in Saga prefecture Japan (fig. 1 and 2) which is a lowland area.

The first area is Hizen Hamashuku (fig. 3) in Kashima city, Saga prefecture, where 161 pictures were shot using a Nikon D60 camera, and 218 pictures were shot using a Redmi Note 5 phone. The second area is Saga city Nagasemachi (fig. 4), where 166 pictures were shot using a Nikon D60 camera, and 192 pictures were shot using Redmi Note 5 phone.

Both areas have urban heritage preservation districts and feature a segment of the old Nagasaki road (Nagasaki Kaido see figures 3 and 4). However, the pictures tried to capture a variety of building styles in preservation districts from not only historic eras such as from Edo to Meiji era but also contemporary buildings. Having a variety of building styles is important for the training process as the program we are planning to train should be able to distinguish traditional architecture from non-traditional architecture.

Different from most existing studies that focused on the entire streetscape and used images taken with the camera facing the street, we put more emphasis on the architectural façade and set the camera facing the buildings so that the architecture takes a larger proportion of the image (Fig. 5 and 6).

However, due to the narrowness of the streets, we had to take our pictures from different angles while keeping the building in the center of the picture.

All the pictures were resized to normalize the entry input in the algorithm and to optimize the training process time. For this purpose, we developed an algorithm that takes as input all the pictures we will use in our dataset and gives as output the same pictures all with the same size of 512x352.

3.2 Data labelling using a questionnaire

All the pictures are labeled regarding different features we intend to test. In this paper, we will present the result of our research regarding two features, the orientation of the building in the pictures and the quality of insertion in a historic landscape.

The pictures were manually rated on the quality of insertion feature by architectural students, having a wide knowledge as much of the topic as of the research areas. In fact, they have studied architecture and urban planning for more than three years and participated in workshops in the designated areas. We then calculated the average score and labeled each picture consequently.

Fig. 5. Example of pictures used in the data set

Fig. 6. Example of pictures used in the data set

According to the reviewed factors, earlier the insertion quality of an architectural façade in an urban heritage area is contributed by fine-textured materials, good quality of details, appropriate coloring, and rhythmic composition. Also, as studied in previous research, ordering, familiar and historical elements contribute to the insertion quality, as well as moderate complexity and artificial nuisances and easy accessibility to the common users.

In the case of preservation areas of Saga prefecture, we rated the visual quality into six classes from zero points to five points. Five points are given to façades that meet almost all the standards above, which usually appear on a well-renovated façade, or well-maintained traditional architecture, etc. Four points are given to less well-maintained buildings, designed with fewer details, using less adequate materials, and have such imperfections as hanging wires and iron window fences. Nevertheless, this group of building façades generally present a neat and clean look. Three points rated buildings present an average aspect regarding the maintenance and used materials. Those rated two points are built with hardly any attention to the details, historicity, or general atmosphere of the historic urban landscape. Besides, they are usually subject to inadequate maintenance, resulting in messy hanging wires, stained walls, rusty iron fences, etc. One point is given to those in a quite dilapidated condition, featured by ramshackle roofs, temporary building material such as metal roof sheets, etc., which usually happen at the urban fringe. Zero point is the score of the images picturing buildings that are absolutely obsolete in such an urban heritage environment.

The survey returns 42 five points images (5.78%), 98 four points images (7.8%), 172 three points images (23.69%), 228 two points images (31.4%), 128 one-point images (17.63%) and 58 zero-point images (7.98%) (Fig.7).

Fig. 7. Example of labeled pictures with

4. Methodology

In the field of computer vision, there are quite a few approaches for image recognition. One of the most promising methods is based on Convolutional Neural Networks (CNN).

Convolutional neural networks are at the core of most state-of-the-art computer vision methodology for a wide range of possibilities. Since 2014 very deep convolutional networks became mainstream, showing substantial gains in various accuracy tests. Moreover, the exponential increase of computational efficiency allows the bigger labeled dataset to be trained, thus, a quality gain in the results (Szegedy 2016).

For this research, it was chosen to work with an inception model for image classification, a pre-trained model using Tensorflow dedicated to computer vision. Compared with conventional image techniques, which are dominated by low-level features like edges and corners, the deep convolutional networks can capture both local and high-level image characteristics.

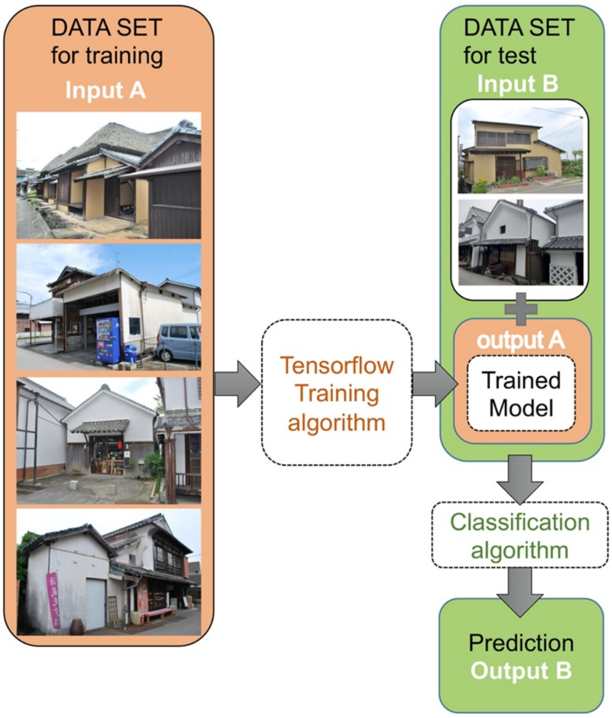

Fig. 8. Training and testing phases diagram

The labeled data set was divided into two subsets, the training set, and the test set. The training set and the test set were randomly sampled in each labeled class, on a, respectively one third and two third proportion. For example, for the quality of the insertion feature, 484 images were randomly sampled in each of the six labels groups for the training set and 242 images each for the test set.

The first set is used as input for the training code. The result will be a trained model that will be used as an input for the classification algorithm. The final output will be a table of prediction scores (fig.8). For the first feature, the orientation of the building, the output will be a prediction of the orientation using one of the suggested labels. The second feature testing result is a ranking of the quality of insertion on a score scale from 0 to 5 for each tested picture.

In terms of the evaluation of model performance, we used the F1-score for the building orientation model and the mean squared error (MSE) for the quality of insertion rating model, which are calculated as below:

$$\mathrm{Re}call = \frac{TP}{TP+FN} \quad [1]$$

$$\mathrm{Pr}ecision = \frac{TP}{TP+FP} \quad [2]$$

$$F1 = \frac{2TP}{2TP+FN+FP} \quad [3]$$

Where, TP = True Positive, FP = False Positive, and FN = False Negative.

While precision refers to the percentage of relevant results, recall refers to the percentage of total relevant results correctly classified by the algorithm and as it can be deducted from its formula, F1 is a general performance indicator that provides a performance indication balanced between Precision and Recall.

Theoretically, A perfect algorithm will provide answers whose precision and recall are equal to 1 (the algorithm finds all of the relevant answers - recall - and makes no errors - precision). In reality, the search algorithms are more or less precise and more or less relevant. It is possible to obtain a very precise system (for example a precision score of 0.99), but poorly performing (for example, with a recall of 0.10, which will mean that it has found only 10% of the possible answers). As far as that goes, an algorithm with a strong recall (for example, 0.99, almost all of the relevant documents), but low precision (e.g. 0.10) will provide many erroneous answers beside the relevant ones: it will, therefore, be difficult to exploit.

Thus, in borderline cases, an algorithm that returns all of the answers in its database will have a recall of 1 but poor precision, whereas a search system that returns only the user's query will have a precision of 1, however, a very weak recall. The performance of an algorithm cannot be reduced to a good score in precision or recall; therefore, it is necessary to add a third performance indicator F1 which combines recall and precision.

5. Results and discussion

5.1 Building orientation feature

For this first feature, we calculated the Recall, precision and F1 scores for each class and obtained the following results:

A significant differentiation of the classes with a high proficiency was established, especially regarding the fact that it is our first simulation and regarding the relatively small number of used images.

These results even if promising, need to be narrowed. Therefore, pictures with false negative pictures should be checked manually to detect if there are some abnormalities that can be enhanced while preparing the data set. For example, the change in the color of the street or some special shapes in the building can induce the model to a False Positive. (fig 10)

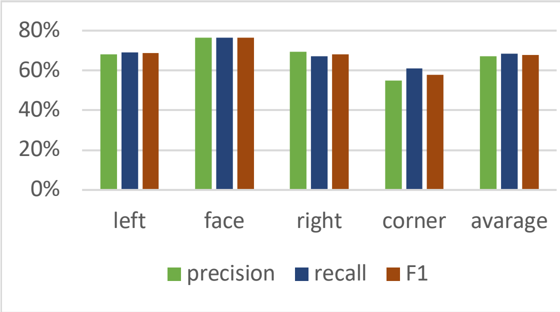

Fig. 9. Comparison of performance model regarding each class

Table 1. Performance of building orientation

	left	Face	right	corner
Precision(%)	68.24	76.36	69.51	55.00
Recall (%)	69.05	76.36	67.06	61.11
F1 (%)	68.64	76.36	68.26	57.89

Fig. 10. The model didn't classify this picture as in a corner probably because of the color of the ground

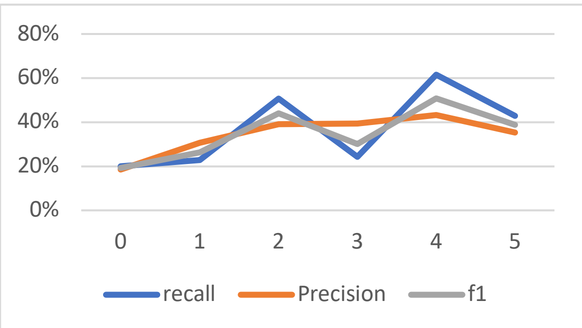

Fig. 11. Comparison of performance model regarding each class

Table 2. Quality of insertion performance

	recall	Precision	F1
0	20.00%	18.52%	19.23%
1	22.92%	30.56%	26.19%
2	50.75%	39.08%	44.16%
3	24.19%	39.47%	30.00%
4	61.54%	43.24%	50.79%
5	42.86%	35.29%	38.71%
average	37.04%	34.36%	34.85%

5.2 Quality of insertion feature

When evaluating the quality of insertion criteria, the lower scores in the model performance indicators were encountered (Table 2). In fact, the performances decreased by half compared to the previous feature. This tells that those issues aren't necessarily related to the algorithm itself but more to other criteria that should be investigated.

A better analysis of the evolution of the performance indicators regarding the quality of the insertion in urban heritage area (Table 2; Fig. 11) shows that the higher is the insertion quality, the better the indicator performances, where the numbers 0 to 5 in Table 10 and Figure 11 are correlated with the extent to which the quality of insertion is perceived. It means also that there is no clear pattern to distinguish between the classes, especially the first ones. This can be related to a big variety of building styles in these classes with fewer pictures.

In order to refine the results, a mathematic formula was developed to calculate the proportions of correct answers and to classify the Negative answers from the closest to the further answer.

Figure 12 helps us to better the overall accuracy of the algorithm. In fact, this graph abscissa values are calculated as follow:

$$X_n = y_i - t_i \qquad [4]$$

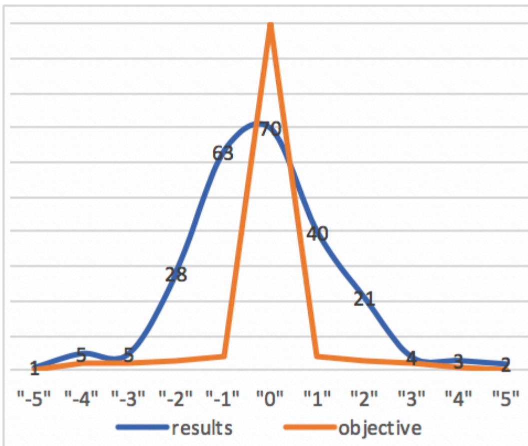

Fig. 12. Evolution of the precision of the result compared to the aim to achieve. X in Abscissa are the difference between the expert panel answers and Y in Ordinate is the number of corresponding answers. It shows that the results present a high number of correct answers, as well as answers with one point difference, and a low rate of answers of 3 points difference and more

Where Yi is the expert rating value and Ti is the prevision suggested by the model. Xn tends to 0 means the results are accurate. However, the farther Xn goes from 0 the less precise are the results. It is visible then that the majority of the model prediction was accurate, with a

1-point difference or in a smaller extent with 2 points difference. The number of predictions with more than two points' difference is negligible. This means that at a certain level, the model could have a general pattern to distinguish between classes, but it is not clear enough at a certain level of accuracy. In this case, the accuracy of the model can be enhanced by exploring three axes:

- As the feature may have been divided into too many classes, some of them should be overlapped.
- A bigger number of experts can also increase the labeling phase precision
- Also, a better explanation of the methodology of scoring to the panel of experts can help to improve the results.

6. Conclusion

The paper aims to select and test a machine learning method to automatically evaluate building insertion quality in urban heritage areas regarding historical landscape. Two features were chosen as the starting point of this research, building orientation regarding the urban façade and the quality of insertion in an urban heritage area. The method can be further extended to evaluate other features such as colors, material, textures, proportions, etc.

Through applying the state-of-art deep convolutional networks, we were able to achieve a machine learning performance on the orientation of building feature with satisfying Recall of 68.40%, Precision of 67.28%, and F1 of 67.79%. This means the model was able to find a precise pattern to differentiate between the classes. Though, the results can be improved using a bigger data set with a bigger variety of buildings.

However, the second feature's quality of building insertion performances was lower. Even though, these results can be interesting for the next steps of research as by analyzing it became possible to see relevant points to enhance. The fact the same model was used for both features, as well as the same data set, proves that the problem resides in other parts of the research. As the difference of rating statistics revealed that even though the average precision was 34.36%, if close classes in the classification are taken into consideration, a higher performance can be achieved. That means a reduction of number of classes have to be made to get better results. Also, previous results can be overlapped to test this hypothesis.

Besides, this indicates an eventual issue in the expert rating protocol. While analyzing False Negative images from the test set, it was found out that some pictures presenting the same building with just a different orientation were rated differently. The difference in rating was mostly of one point.

Also, the number of experts should increase to have a more accurate average score. This point seems though more problematic as increasing the quality of rating protocol and the number of images in the data set will make the process excessively time-consuming.

The convolutional neural network (CNN) was able to capture more general features but it may still not able to grasp all the visual cues that contribute to the judgments as mentioned in factors review which is still an open problem in the field of deep learning. An exploration of other deep learning algorithms can be useful in that case.

It is important to highlight that the computer algorithm is not meant to take decisions but moreover give direction for decision makers in an urban heritage area (Quercia et al. 2014) as the algorithm does not always suggest the best condition. For instance, although a high quality of insertion in urban heritage contributes to the calmness and harmony of the street, interruptions at certain points are also necessary to provide variety, as well as a rest for the eyes.

This paper serves as a first step toward the insertion of new buildings in urban heritage areas with computational methods help. The proposition is that this line of research can be extended in several ways. First, as mentioned before, more features can be included into the machine learning algorithm to produce a more comprehensive profile of the urban visual environment, such as material, colors, the building scale, the area, etc.

Second, a bigger dataset providing a better-quality image with more accurate labeling can have a positive effect on the result.

Third, more general use of the model extended to other areas than urban heritage area could be done, as this research is meant to be used more generally in any urban area with a specific design guideline.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 19H02315.

I would like to express my very great appreciation to Professor Hayashida for his valuable and constructive suggestions during the planning and development of this research work. His willingness to give his time so generously has been very much appreciated.

I would also like to extend my thanks to students from Mishima Laboratory, Saga University and Bilel Daoud from Kyushu University for their help in offering me the resources in running the research.

References

Brownson, R. C., C. M. Hoehner, et al. (2009). "Measuring the built environment for physical activity: state of the science." American journal of preventive medicine 36(4): S99-S123. e112.

Buchanan, P. (1988). A report from the front. Urban Design Reader.

Carmona, M. (2010). Public places, urban spaces: the dimensions of urban design, Routledge.

Devlin, K. and J. L. Nasar (1989). "The beauty and the beast: Some preliminary comparisons of 'high' versus 'popular 'residential architecture and public versus architect judgments of same." Journal of Environmental Psychology 9(4): 333-344.

Doersch, C., S. Singh, et al. (2012). "What makes Paris look like Paris?" ACM Transactions on Graphics 31(4).

Ewing, R. and O. Clemente (2013). Measuring urban design: Metrics for livable places, Island Press.

Kandinsky, W. (2012). Concerning the spiritual in art, Courier Corporation.

Kostof, S. (1992). The city assembled, London.

Harvey, C. W. (2014). "Measuring Streetscape Design for Livability Using Spatial Data and Methods."

Lee, S., N. Maisonneuve, et al. (2015). Linking Past to Present: Discovering Style in Two Centuries of Architecture. IEEE International Conference on Computational Photography.

Liu L., Wang H., Wu C. (20xx) A machine learning method for the large-scale evaluation of urban visual environment. Computers, Environment and Urban Systems Volume 65, September 2017, Pages 113-125

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

Milroy, B. M. (2010). Thinking planning and urbanism, UBC Press.

Naik, N., J. Philipoom, et al. (2014). Streetscore--predicting the perceived safety of one million streetscapes. Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE.

Ordonez, V. and T. L. Berg (2014). Learning high-level judgments of urban perception. Computer Vision–ECCV 2014, Springer: 494-510.

Quercia, D., N. K. O'Hare, et al. (2014). "Aesthetic capital." 945-955.

Salesses, P., K. Schechtner, et al. (2013). "The collaborative image of the city: mapping the inequality of urban perception." PLoS One 8(7): e68400.

Szegedy C., Vanhoucke V., Ioffe S, Shlens J, Wojna Z; The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2818-2826

Ulrich, R. S., (1983) "Aesthetic and affective response to natural environment", In I.Altman, 1983

Von Meiss, P. (2013). Elements of architecture: from form to place, Routledge. Routledge: 204-207.