ESTIMATION OF INCOME LEVELS PER BUILDING UNIT USING SATELLITE IMAGE AND INCOME STATISTICS

Ryota Yamanotera[1], Yuki Akiyama[2], Hiroyuki Miyazaki[3] and Satoshi Miyazawa[4]
[1]Graduate School of Integrative Science and Engineering, Tokyo City University
1-28-1, Tamazutsumi, Setagaya, Tokyo 158-0087, Japan,
Email: g2381644@tcu.ac.jp

[2]Associate Professor, Faculty of Architecture and Urban Design,Tokyo City University
1-28-1, Tamazutsumi, Setagaya, Tokyo 158-0087, Japan,
Email: akiyamay@tcu.ac.jp

[3] CEO, GLODAL, Inc.
7th Floor CrossGate Bldg. 1-101-1 Sakuragi-Cho Naka-ku Yokohama Kanagawa 2310062 Japan
Email: miyazaki@glodal-inc.com

[4]LocationMind Inc.,
#4F PMO Kanda-tsukasamachi, 2-8-1 Kanda-tsukasamachi, Chiyoda-ku, Tokyo 101-0048, Japan
Email: miyazawa@locationmind.com

KEY WORDS: Income Level, Population, Satellite Image, Developing Countries, Deep Learning

ABSTRACT: In many cities of developing countries, they are currently facing the problem of rapid population growth. However, due to insufficiently organized statistics on current population distribution, it has been difficult to establish effective urban plans to manage this problem. In order to grasp the detailed spatial distribution of the population, it is first necessary to understand the distribution of residential buildings. In recent years, the development of methods to understand the distribution of population by estimating the buildings where people reside from satellite images has been promoted. These studies have already confirmed a certain level of accuracy, and it has been found that it is entirely possible to grasp the distribution of buildings and the area of buildings from satellite images. However, the number of household members is related not only to the area of the building but also to the household income. In other words, the number of people living per unit area of a building greatly varies depending on the income class of the area. Therefore, to grasp the population more accurately, a method of estimating the income level of each building is required. In existing studies, there have been attempts to estimate household income using the results of field surveys as training data. However, verification of the results obtained in these studies has been limited to using information on income obtained through field surveys of very limited locations, and verification has not been conducted using information on income over a large area. Therefore, this study proposed and verified a method of estimating the income level of each building using satellite images. The target area is Japan where could use more detailed statistics related to estimated annual income, which led us to believe we could more purely evaluate our methodology. We first created economic level-based training data using Google map background images and estimated annual household income at each city block. Next, we created an economic level classification model using deep learning model widely used for image recognition, and the training data. Third, we sused them to classify the images, and classified the results by income level. Based on these results, we finally assigned income level to the residential map. This result allowed us to compare the distribution of economic levels with the correct data distribution.

1. INTRODUCTION

In many developing countries, they are facing the problem of rapid population growth in recent years. In such circumstances, in order to continue sustainable development and growth of the cities, it is necessary to quantitatively and accurately grasp the actual conditions of the cities. One of the pieces of information that makes this possible is population statistics. However, in developing countries, population statistics may cover a limited geographical range, be updated infrequently, or be irregular. Therefore, it is difficult to accurately grasp the current population distribution.

Therefore, in these countries, there is a demand for methods to grasp population distribution at a high resolution. To solve this problem, a method of estimating the population at the building unit using building data is mentioned. For example, Akiyama et al. (2019) proposed a method of extrapolating to developing countries the population estimation model based on building area and number of floors developed in advanced countries where detailed building data and population statistics are maintained.

However, this method presupposes that the building data of the target area is sufficiently maintained. Therefore, it is difficult to apply this method to cities and regions where building data is not sufficiently maintained. In addition, research is advancing on methods to grasp population distribution by estimating buildings where people are distributed from satellite images. Kaiser et al. (2017) have developed a method to grasp the distribution of buildings from satellite images using deep learning, and this method has enabled a rough understanding of where people are concentrated and living.
On the other hand, it is known that the number of household members is not only related to the size and distribution of buildings but also has a significant relationship with household income. Therefore, a method to estimate the income level of each building is also required. However, in developing countries, there are often no statistics on accurate income levels. Therefore, it is necessary to first develop a method to understand the income level at the building unit. As an initiative to grasp the income level at the building unit, there is an income level estimation at the building unit using satellite images and deep learning, creating teacher data using information on the income level at the building unit estimated from interviews conducted locally by Okuda et al. (2019). However, the interviews target a very limited area, and verification using detailed data on income has not been conducted.

1.1 Objective

To address the above problem, it is necessary to construct a method to estimate the income level at the building unit in regions where large-scale and reliable statistics on income are available, and to verify its reliability. Therefore, this study aims to propose and verify a method to estimate the income level at the building unit using satellite images, targeting Japan where reliable statistics on income are maintained on a nationwide scale.

2. METHOD

First, using two kinds of Japanese statistics: the Housing and Land Survey and the National Census, we created average household income data for individual city block (hereinafter called the "city block household income data"). Next, by spatial integration of the building polygon data obtained from residential map provided by Zenrin. Co. Ltd. And the city block household income data, we attributed the average household income of the city block where each building is located to each building. Furthermore, referring to the Japanese Household Survey, we classified the income into three ranks and used it as income level labels, which were used as labels for the training data. In addition, we cropped satellite images that these buildings would be centered and created training data by pairing them with the created income level labels. Subsequently, using a pretrained deep learning model called VGG-16 and the training data, we constructed a classification model for income level. The learning method in this study is the transfer learning. Finally, we verified the accuracy using the constructed model and data not used in the training.

2.1 Target area

The target area of this study was Shinjuku Ward, Tokyo prefecture (hereinafter called "Shinjuku Ward"). The reason is that the geographical condition where buildings of various heights are densely distributed is like the conditions seen in urban areas of developing countries. In addition, the availability of reliable statistics on income in Japan is also a reason for selecting Shinjuku Ward as the target area. The location of the target area and the overall situation are shown in Figure 1.

3. DATA

3.1 City block household income data

The city block household income data was used to attribute economic level attributes to each building. It was created using the National Census (2015) and the Housing and Land Statistics Survey (2018) by the method of Yamanaka et al. (2021). While Yamanaka et al. estimate household income at the 1 km and 500 m square mesh unit, this study applied the method of Yamanaka et al. to city blocks to create city block household income data. Figure 2 shows the city block household income data in Shinjuku Ward.

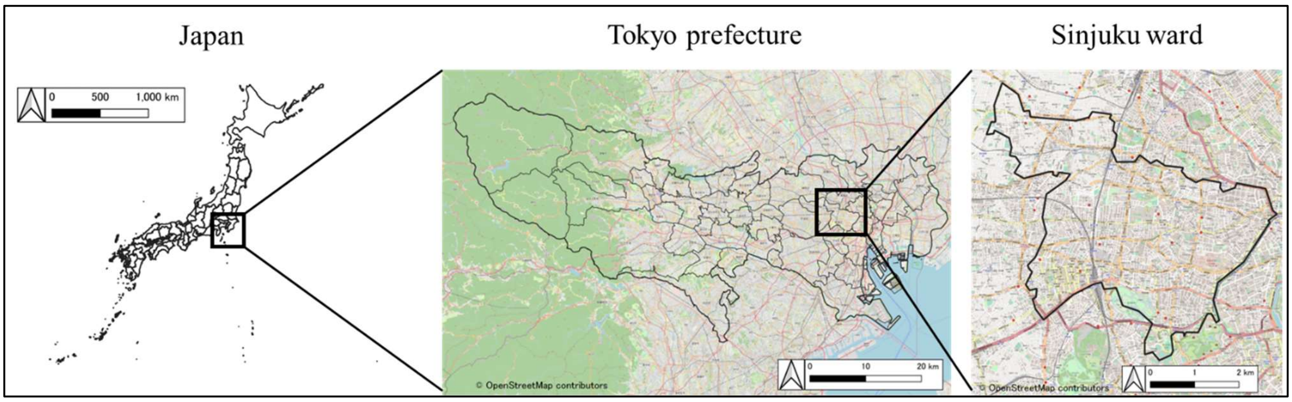

Figure 1. Location of Sinjuku ward

3.2 Building polygon data

We used the polygon data representing the shape of the buildings recorded in the residential map of Zenrin Co., Ltd. (Zmap TOWN II: 2020) as the building polygon data. The attribute of the polygon data contains building use. Therefore, in this study, since we are targeting residential buildings, we used only the buildings with the use shown in Table 1.

3.3 Satellite image

We used the background images from Google Map. The size is 256 pixel, and we used images with a resolution of 0.3 m per pixel.

4. CLASSIFICATION OF INCOME LEVEL AT THE BUILDING UNIT

4.1 Creation of training data

First, we created training data for constructing the deep learning model. Household incomes derived from the city block household income data were classified into income level categories ranging from 0 to 2. In this study, level 0 was set to less than 3.5 million yen, level 1 to 3.5 million yen or more however less than 7.5 million yen, and level 2 to 7.5 million yen or more. The number of buildings classified into each level is shown in Table 2. This classification was set by referring to the relationship between annual income and the number of household members in the household survey, adopting the average income that results in a change of one or more in the number of residents as the threshold.
Next, by spatial integration the city block household income data with the building polygon data, we integrated the average household income and income level to each building. In the case of buildings located across multiple city blocks, it was apportioned using the area of the intersecting city blocks as weights. Through this process, we first created the correct labels.
Finally, we created image data to pair with the correct labels. We cropped the satellite images in a region enclosing the central coordinates of each building in a 256pixel square and paired them with the buildings to create training data for model construction. Through this process, we were able to create training data for a total of 86,274 buildings.

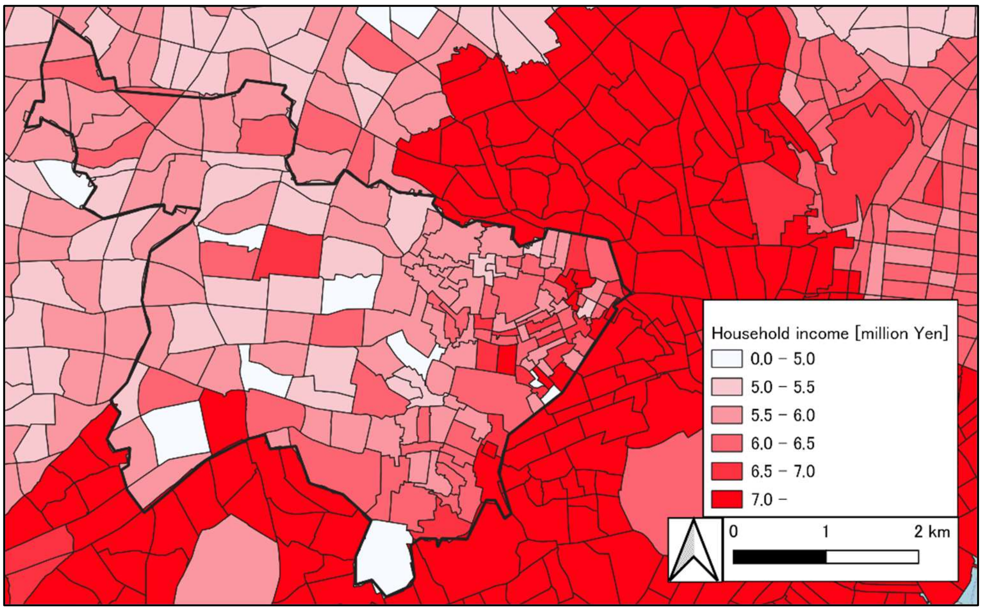

Figure 2. Household income data of city block unit

Table.1 List of building type and the number of buildings in Shinjuku Ward

Building use	Number of buildings
Multiuse and apartment buildings	14,283
Detached house	34,567
Detached office and commercial building	37,424

Table.2 List of Income Level and Buildings

Income level	Household income [Million yen]	Number of buildings
0	0 – 3.5	56,569
1	3.5 – 7.5	29,576
2	7.5 –	129

4.2 Construction of model using deep learning

Next, we proceeded to construct the deep learning model using the training data. In this study, we implemented a classification model for income level using a pretrained deep learning model called VGG-16 (Simonyan & Zisserman, 2014). VGG-16 is an algorithm for object detection and classification, capable of classifying 1,000 types of images with an accuracy of 92.7%, and it exhibits extremely excellent performance, especially in the field of image classification. In addition, because its architecture has a structure that is easy to understand, it is easy to adjust the model, and it is frequently used in transfer learning. Therefore, in this study, we constructed it using a method called transfer learning. For model construction, we used 69,019 images, which is 80% of the training data, and used the remaining 20%, or 17,255 images, for accuracy verification. The number of learning iterations was set to 20.

4.3 Output of Classification Results

We applied the income level classification model to the validation data to output the income level for each building. Figure 3 shows the results of visualizing the correct data and the prediction results by our model, as well as the distribution of misclassified buildings.

5. ACCURACY VALIDATION OF MODEL

5.1 Method of accuracy verification

In this study, we conducted accuracy verification for each income level using four indicators: accuracy, precision, recall, and F-measure. By using many indicators, we can understand the prediction quality of each class in detail, and clearly identify in which income level areas there are problems with classification accuracy. In addition, it becomes possible to check if there is any bias in the learning tendency. Furthermore, to check whether the model is not biased towards the training data, we created loss curves and accuracy curves using both the training data and the validation data.

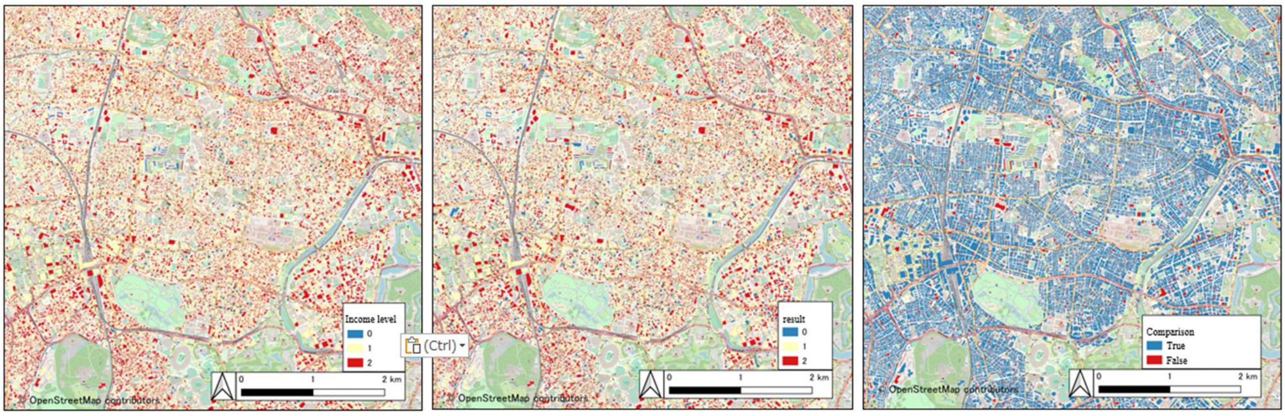

Figure3. Correct data (left), prediction results (center), and distribution of misclassified buildings (right)

Table-3 Results of each indicator for each income level

Income level	Accuracy (%)	Precision (%)	Recall (%)	F-measure (%)
0	98.39	92.00	62.16	74.19
1	98.39	99.09	98.48	98.79
2	98.39	97.13	98.45	97.78

5.2 Verification result

The results for each indicator at each income level are shown in Table 3, and the confusion matrix of the classification results, along with the loss curves and accuracy curves during model training, are shown in Figure 4. From the accuracy curve, classification can be achieved with a high accuracy rate of over 90% for both the training data and the validation data. In addition, since the trends in accuracy and loss changes for the training data and validation data are similar, it is understood that learning is being achieved while maintaining generalization performance. By checking the results of each indicator, classification could be performed at a very high level of over 90% in all indicators for income level 1 and 2. On the other hand, it was confirmed that the F-value was 74% for income level 0, which is lower in accuracy compared to other income levels. From these results, it was understood that while accurate classification is possible in middle and high-income areas, there are challenges in the accuracy for low-income areas.

6. DISCUSSION

6.1 Discussion Based on Visual Inspection of Images

According to the accuracy obtained in Chapter 5, it was confirmed that it is possible to classify economic levels from satellite images with quite high accuracy. However, it was also confirmed that there are challenges in classification accuracy in the low-income level.

Therefore, first, we sought to identify the trend of where misclassifications occurred most frequently across all regions through visual inspection of satellite images. As a result, it was confirmed that there were many misclassifications in areas with complex buildings having multiple functions, such as mixed-use buildings, and in high-income areas where there is a generous distance between neighbouring buildings. Figure 5 shows an example of buildings with complex shapes and shows an example of an area with high income and a large distance between neighbouring buildings. The reason for the decreased accuracy in these areas is believed to be that the number of samples meeting these conditions was small, preventing the model from making correct judgments. This is a common cause also for the low classification accuracy in the low-income bracket. In Shinjuku Ward, the target area of this study, over 95% of households had an average household income of over 3.5 million yen, showing a bias in the data itself. In other words, it was found that the classification accuracy of this study depends to some extent on the number of data points.

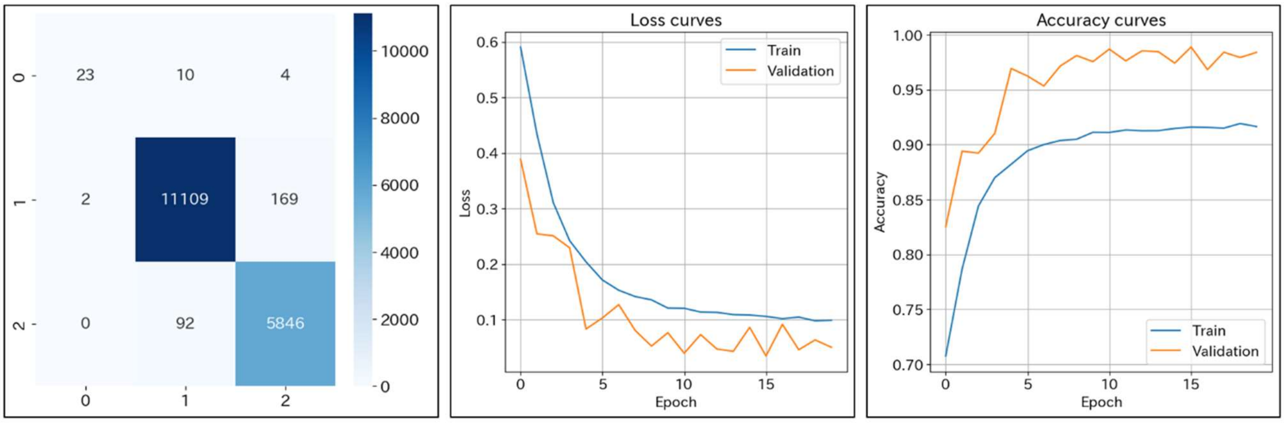

Figure4. Confusion matrix for validation data (left), loss curve (center), and accuracy curve (right)

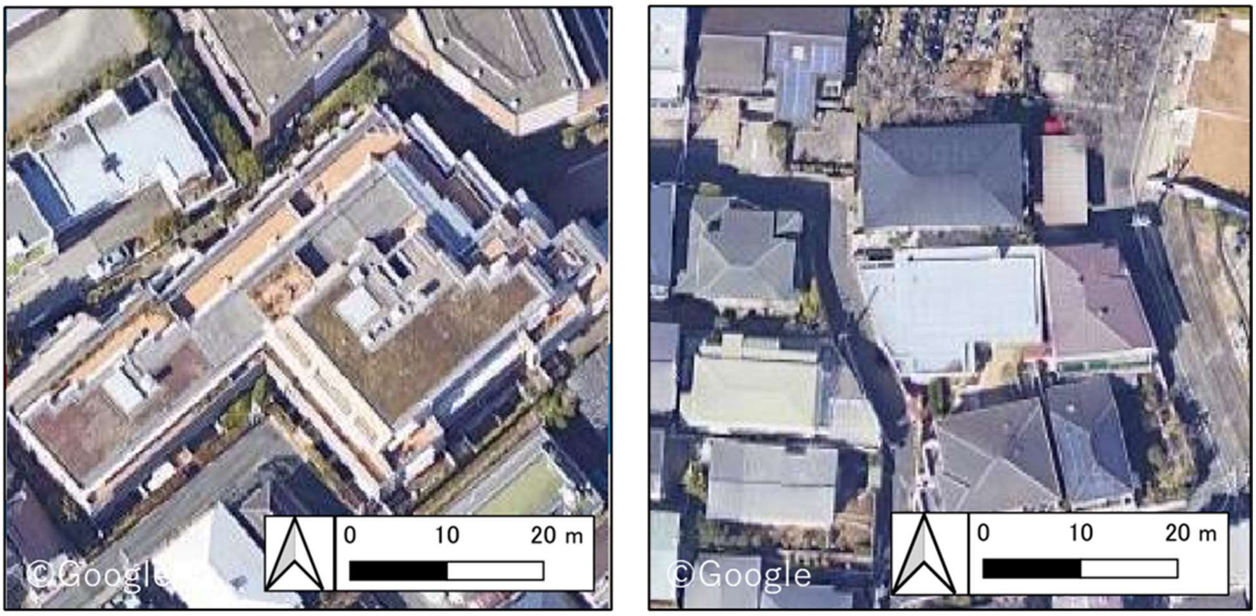

Figure5. Examples of buildings with complex shapes (left)
and areas with large distances between adjacent buildings (right)

6.2 Discussion Based on Visualization of Features Using Grad-CAM

Next, to explore what kind of features of buildings the model developed in this study is using to determine the income level, we performed learning visualization using a method called Grad-CAM (Selvaraju et al., 2017). Grad-CAM is a method for generating visual explanations when a Convolutional Neural Network (CNN) based model decides a class. It uses the gradients of any target concept flowing into the final convolutional layer to generate a localization map that highlights important areas in the image for predicting the concept, enabling visualization of the areas used for judgment and their importance. Using this method, we verified what the model created in this study emphasizes when classifying classes. Figure 6 shows the results of classification visualization using Grad-CAM. According to these results, it was found that the classification is not based on the buildings themselves, however on the gaps between the buildings. In fact, areas referred to as high-income tend to have densely packed buildings in Japan, so it is believed that this characteristic is being adequately captured. Additionally, it was confirmed that in the case of buildings with very large areas and complex shapes, such as large factories, it is possible that misclassifications occur because it cannot properly determine the boundary between the road and the building.

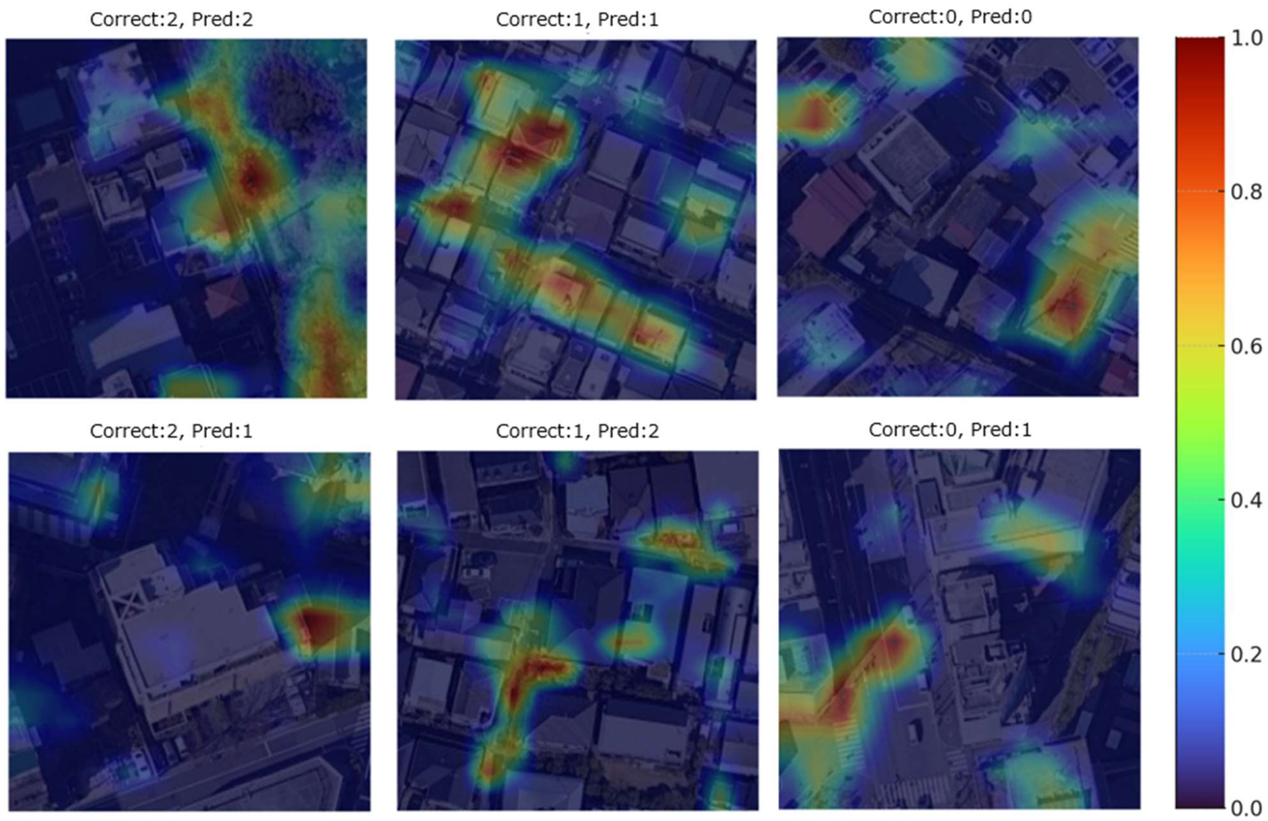

Figure-6 Results of visualizing the regions used during classification and their importance using Grad-CAM

6.3 Future works

First, it is believed that the improvement of classification accuracy in areas including low-income bands and complex buildings can be solved by expanding the target area. Shinjuku Ward is a region where the income level is quite high even in Japan, and it can be said to be a unique area in Japan where buildings of various heights and shapes are densely accumulated. Therefore, it is believed that by expanding the target area to the entire Tokyo area and further to the entire Kanto region in the future, the model can improve classification accuracy.

In addition, for large and complex-shaped buildings, a method of constructing a model in a state where they are temporarily excluded by setting a certain area as the threshold can be considered. In fact, the possibility of people settling in complex facilities and large-scale business establishments is small, so it is believed that excluding them will have little impact on the estimation accuracy of the model.

As for future developments, in addition to the above improvements, it is necessary to check what kind of impact it will have when deployed in developing countries. While it was observed in Shinjuku Ward that the higher the income area, the more densely buildings are located, it is necessary to verify whether correct classification can be done when implemented overseas.

Furthermore, towards the realization of the goal of this entire research, which is to estimate population distribution from satellite images, we are also working on automatic extraction of buildings from satellite images and estimation of building use. When we will integrate with these methods, we want to verify how much accuracy can be demonstrated in the next efforts.

7. CONCLUSION

In this study, we proposed and verified a method to estimate the income level of individual buildings using satellite images, targeting Japan where reliable income statistics are available. As a result, it was confirmed that we could classify the income levels of individual buildings with high accuracy. On the other hand, some challenges remained in the classification accuracy in low-income bands where the number of data was small, and in the accuracy for buildings with features such as complex buildings, which had a small number of cases overall. Going forward, while working to enhance the training data and improve the model so that accuracy can be demonstrated even for buildings with these features, we will proceed with efforts aimed at expanding the target areas to other regions in Japan and overseas and working towards population estimation at the building unit level.

Acknowledgements

This study was conducted with the support of a JSPS KAKENHI grant (JP20H01483). In addition, this study is a part of the results from collaborative research with CSIS, the university of Tokyo (No. 1170, used data: Zmap TOWN II in 2020). We would like to express our gratitude here to them.

References

Akiyama, Y., Miyazaki, H. and Sirikanjanaanan, S., (2019) Development of Micro Population Data for Each Building: Case Study in Tokyo and Bangkok, *2019 First International Conference on Smart Technology and Urban Development (STUD)*, pp.1-6.

Kaiser, P., Wegner, J., Lucchi, A., Jaggi, M., Hofmann, T. and Schindler, K. (2017) Learning Aerial Image Segmentation from Online Maps, *IEEE Transactions on Geoscience and Remote Sensing*, Vol.55, No.11, 201.

Okuda, K., Kawasaki, A., & Hamaguchi, R. (2019). Estimation of income levels for each building using satellite images and household survey data. *GIS - Theory and Applications*, Vol. 27, No.2, pp.9-18.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017), Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization, *Proceedings of the IEEE international conference on computer vision*, pp. 618-626.

Simonyan, K. and Zisserman, A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition, *arXiv preprint arXiv:1409.1556*.