DEVELOPMENT OF BUILDING DETECTION METHOD WITH BUILDING USE FROM SATELLITE IMAGES USING DEEP LEARNING

Keisuke OKADA[1], Ryota YAMANOTERA[1], Yuki AKIYAMA[2,3], Hiroyuki MIYAZAKI[3], and Satoshi MIYAZAWA[4]
[1]Graduate School of Integrative Science and Engineering, Tokyo City University,
1-28-1, Tamazutsumi, Setagaya-ku,Tokyo 158-8557, Japan,
(g2281611,g2381644)@tcu.ac.jp

[2] Department of Urban and Civil Engineering, Faculty of Architecture and Urban Design, Tokyo City University,
1-28-1, Tamazutsumi, Setagaya-ku,Tokyo 158-8557, Japan,
akiyamay@tcu.ac.jp

[3]Center for Spatial Information Science, The University of Tokyo,
5-1-5, Kashiwanoha, Kashiwa, Chiba 277-8568, Japan,
heromiya@csis.u-tokyo.ac.jp

[4] LocationMind Inc.,
2-8-1 Kanda Tsukasa-machi, Chiyoda-ku, Tokyo 101-0048, Japan
miyazawa@locationmind.com

KEY WORDS: Satellite Images, Deep Learning, Building Detection, Building Use

ABSTRACT:
In recent years, despite rapid population growth in many cities in developing countries, it has become difficult to determine the actual population distribution due to incomplete demographic statistics, which cover only some cities and regions, or are updated irregularly. Therefore, there is a need for a method that can provide a high-resolution information of the actual population distribution in these areas. In this study, we developed a method using deep learning to extract buildings from satellite images and to identify buildings where residents are distributed by estimating the building use. As a result, we can extract buildings from satellite images with an average IoU of 64.3% and an average extraction rate, in other words the number of extracted buildings divided by the number of buildings in the verification data, of 62.1%. We can also classify detached buildings and non-detached buildings with an average accuracy of 78.7% for estimating building use. Furthermore. The process for classifying detached buildings into residential and non-residential was able to be performed with an average accuracy of 62.4%, while the process for classifying non-detached buildings into apartment buildings and other buildings was able to be performed with an average accuracy of 75.2%.

1. INTRODUCTION

In almost all developed countries, it is possible to obtain a detailed understanding of population distribution by utilizing official demographics. By using these statistics, we can conduct effective medium- and long-term urban planning in various fields, including transportation management (Fuller et al., 2013), disaster management (Rumbach, 2016), urban climate change mitigation (Dulal et al., 2011), urban resource management (Agudelo-Vera et al., 2011), and public health (Niemelä et al., 1999). However, many developing countries face challenges that existing data cover only some cities or regions, or they are updated infrequently or irregularly (Robinson et al., 2017, Akiyama et al., 2019). This is due not only to the large budget and labor required to maintain large-scale demographics, but also to the existence of the informal sector, such as street dwellers, who are not included in official demographics, and the existence of districts that are not adequately covered by statistical surveys such as slums (Kumar, 2014). Therefore, there is a need for a method to determine detailed population distribution in developing countries.

1.1 Literature Review

To address this problem, there are some previous studies which have tried to estimate detailed population distribution in developing countries using satellite image, which is relatively readily available and can be accumulated at the same quality as in developed countries. To monitor detailed population distribution, it is first necessary to detect the distribution of buildings where the resident population is distributed. To address this problem, some studies have been conducted to identify buildings where the population may be distributed using satellite images. There are some old examples of satellite-image-based population estimation (Polie, 1984, Taragi et al., 1994), it was difficult to obtain a detailed population distribution due to the low resolution of satellite images available at the time. In recent years, Doupe et al. (2016) proposed a method to estimate population distribution by classifying the number of artifacts in patch images culled from U.S. satellite images into 14 classes and multiplying each class by a population factor. However, this method estimates the population without considering households in apartment buildings and non-residential areas, and thus deviates from the actual population distribution. Similarly, the estimated population based on a 1 km square grid developed by Balk and Yetman (2004), and "Estimated population based on 100m square grid" by Atem (2017) are known examples of the population distribution being estimated by grid. These data proportionally distribute the population by estimating the land use status of each grid, especially the build-up status. The actual population distribution varies greatly depending on the building use. Furthermore, Iino et al. (2018) extracted build-up areas from synthetic aperture radar satellite images for urban areas in Jakarta using deep learning to produce high-resolution urban distribution maps. In addition, urban distribution maps at multiple points in time have been used successfully to understand the progression of urbanization. However, this method is not applicable to areas where building statistics are not maintained because the buildings are extracted based on the number of buildings recorded in the statistics published by the Indonesian government. In addition, the building use was not estimated in this study.

1.2 Objective

As described above, previous studies mainly extracted build-up areas on a grid scale or relied on existing statistics to extract buildings. As mentioned above, to monitor detailed population distribution, it is important to not only extract building distribution but also know detailed building attributes such as building use and number of floors. If we can know the building use, it would be possible to identify buildings with residential use that serve as residential areas. Similarly, if we can know the number of floors on each building, the population can be proportionally distributed according to the volume of each building. Thus, it is expected that this information will enable an accurate estimation of the population distribution.

Therefore, the objective of this study is to develop a method for generating building data from satellite images, which are relatively easy to obtain even in developing countries, and to develop a method for estimating building use to identify buildings in which residents are distributed.

2. STUDY FLOW

Figure 1 shows the flow of this study. This study consists of two phases. Phase one is Building detection, and phase two is Estimation of building use. In phase one, we develop a method to automatically extract buildings from satellite images. In phase two, we estimate the building use of the extracted building images. Although the processing of phase one and phase two should be done in one step, we found that the processing of each phase is different in nature, as described below. Therefore, in this paper, phase one was developed independently in Section 3 and phase 2 in Section 4. Furthermore, while this study should be conducted in developing countries with incomplete demographics as target areas, developing countries often do not have sufficient data to verify the reliability of information on extracted buildings. Therefore, the target areas of this study are several areas in Tokyo, where sufficient data for verification are available.

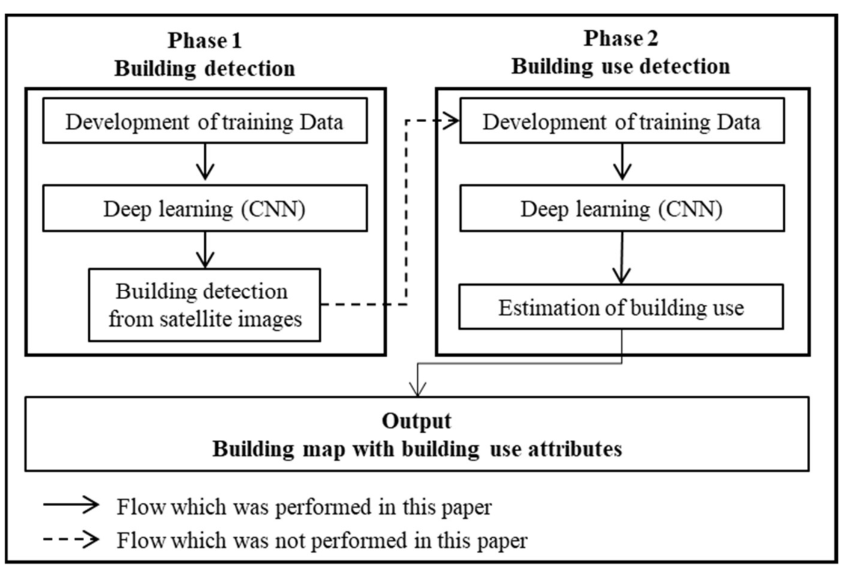


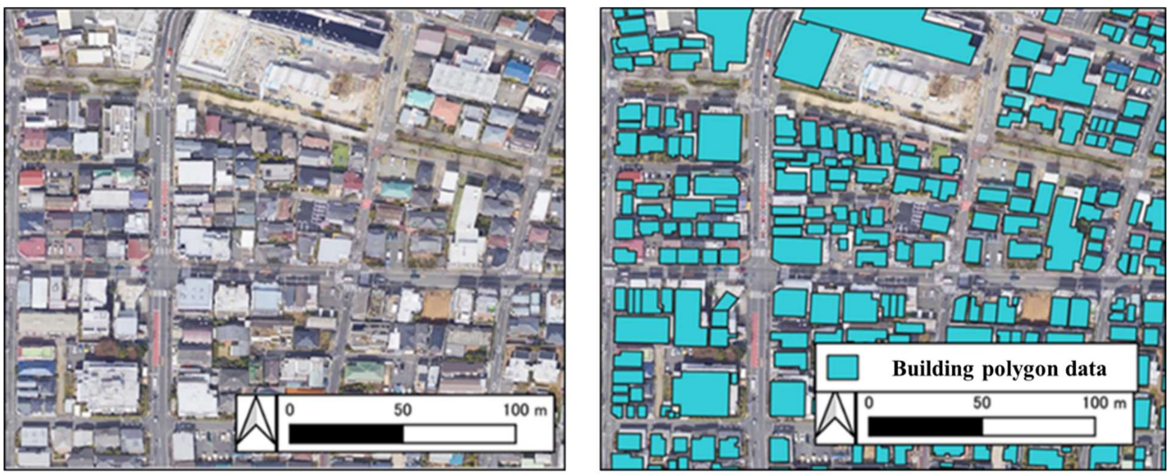
Figure 1. Study flow of this paper

3. BUILDING DETECTION USING DEEP LEARNING

In this study, Shinjuku-ku, Tokyo (hereinafter referred to as "Shinjuku-ku") was selected as the target area in Japan, which has the characteristic of high-density distribution of buildings with height variety. This feature was often seen in cities in developing countries. Satellite images of Shinjuku-ku was used to train the model. In addition, Hachioji City, Tokyo (hereinafter referred to as "Hachioji City") was also included in the validation data as an area where buildings are not as densely distributed as in Shinjuku-ku, assuming that the model is applicable to suburban areas in developing countries.

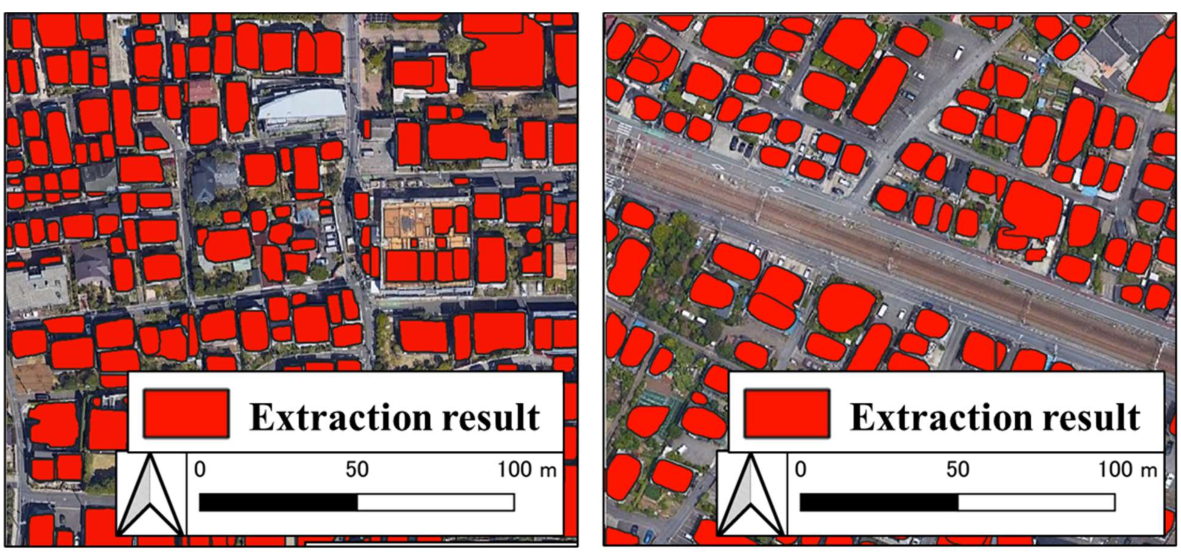
3.1 Method of Building Detection

The training data was satellite images which is background satellite images from Google Maps of Shinjuku-ku and corresponding polygon data of building shapes from residential map (2020) provided by ZENRIN CO.,LTD. (hereinafter referred to as "building polygon data"). Figure 2 shows an example of a satellite image and building polygon data in Shinjuku-ku. The data were spatially integrated to create training data for building the model. The model was constructed using Meta's detectron2 library(Facebookresearch), which is a CNN-based library that provides higher accuracy and shorter processing time than conventional methods. Finally, the extraction results were compared with building polygon data to verify the accuracy of the model.



Satellite images: © 2021 Google, Airbus, Digital Earth Technology, Maxar Technologies, Planet.com, The GeoInformation Group

Figure 2. Satellite image (left) and building polygon data (right) (Example of Shinjuku-ku)



Satellite image: © 2021 Google, Airbus, Digital Earth Technology, Maxar Technologies, Planet.com, The GeoInformation Group

Figure 3. Result of building extraction
(Left: Example of Shinjuku-ku Right: Example of Hachioji City)

3.2 Result of Building Detection

Accuracy of building detection is shown in Table 1. In this study, the IoU, which indicates the degree of overlap between the building polygon data and the building polygon data obtained from the building detection results, and the building extraction rate, which indicates the percentage of agreement between the number of extracted buildings and the number of buildings in the building polygon data, were used as evaluation indicators. In both Shinjuku-ku and Hachioji city, the IoU was more than 60% and the building extraction rate was more than 65%. The accuracy was particularly high in Hachioji city, and it was clear that extrapolation was possible with high accuracy for areas where the density of buildings was somewhat low. On the other hand, the value for Shinjuku-ku was lower than that for Hachioji city. This is because in areas where buildings are densely distributed, the extraction of narrow buildings failed, and multiple adjacent buildings were extracted as one building. To solve this problem, it is expected that the accuracy can be improved by increasing the amount of training data in areas where narrow buildings are densely distributed.

4. ESTIMATION OF BUILDING USE USING DEEP LEARNING

We will proceed to develop a method for the estimation of building use. In this study, the target areas are Shinjuku-ku, Hachioji city, and Setagaya-ku, Tokyo (hereinafter referred to as "Setagaya-ku"), which is adjacent to Tokyo's city center and predominantly residential throughout almost its entire area.

4.1. Method of Building Use Estimation

We performed deep learning using data obtained by spatially integrating a square-cut image of the area where buildings are located from satellite images of the target area (Figure 4) with the attributes of building use obtained from 2020 building polygon data as training data. The building use contains five types of use: detached houses, detached offices, multi-use buildings mainly residential use, multi-use buildings mainly business use, and apartment building. We used the MMPreTrain pre-training framework provided by OpenMMLab as the model(MMPreTrain Contributors. 2023), and fine-tuned the model with Resnet18.

To discriminate the five types of building uses, a five-class image classification task is performed at once, which is known to result in low accuracy due to the large number of classes. (Okada et al.,2022). Therefore, this study divided the building use estimation into two steps as shown in Figure 5. In the step 1, the five building uses were classified into two classes: detached buildings (detached houses and detached offices) and non-detached buildings (multi-use buildings mainly residential use, multi-use buildings mainly business use, and apartment buildings), which have similar building appearance characteristics. Next, in Step 2, buildings classified in Step 1 as detached buildings were classified into two classes: detached houses and detached offices, and buildings classified in Step 1 as non-detached buildings were classified into three classes: multi-use buildings mainly residential use, multi-use buildings mainly business use, and apartment buildings. Furthermore, in the target area, the number of buildings in the multi-use buildings mainly residential use and multi-use buildings mainly business use was small compared to the number of detached houses and apartment buildings, creating a bias in the number of building images for each building use. It has been pointed out that deep learning on such an unbalanced dataset reduces accuracy (Ali et al., 2013). Therefore, to reduce the bias in the number of images in each class, we corrected the imbalance in the number of images among classes by oversampling, which amplifies the number of images by duplicating image in classes with fewer images, and undersampling, which randomly reduces the number of images in classes with more images.

Table 1. Accuracy of building detection

	Shinjuku-ku	Hachioji city
IoU (%)	64.34	76.02
Building extraction rate (%)	62.08	75.09

Figure 4. Example of each building extraction from satellite image

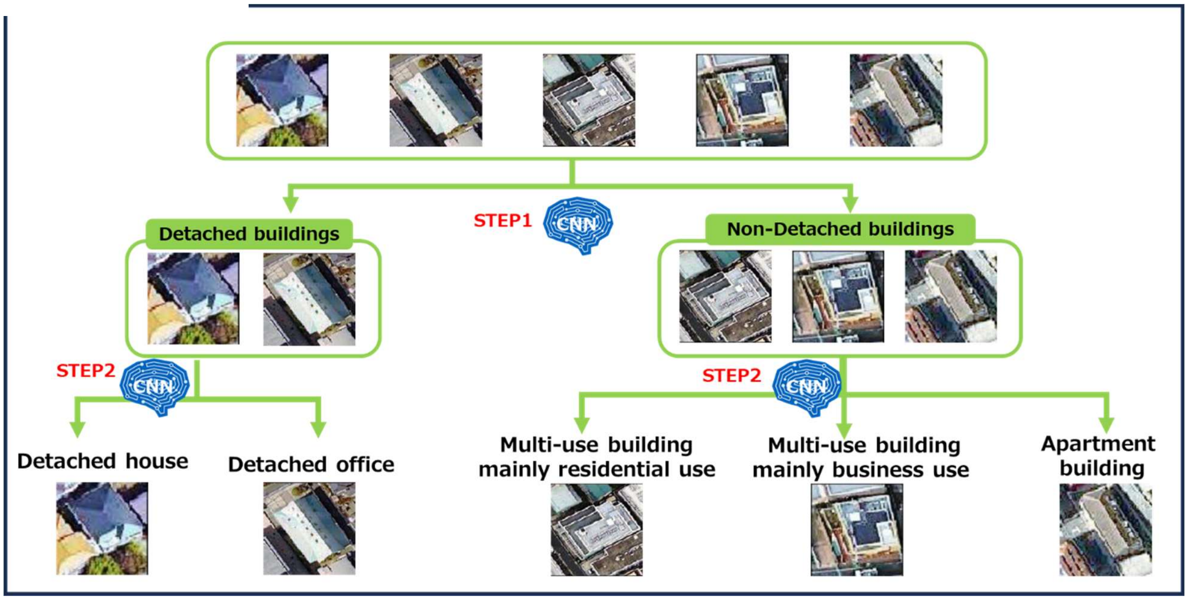


Figure 5. Flow of building use estimation

4.2. Result

4.2.1. Estimated Results for Building Use in Step1

Table 2 shows the results for building use estimation of Step 1 for each municipality in the target area. In each municipality, accuracy, precision, recall, and F1-score exceeded 70%. It indicate a high degree of accuracy in estimating whether a building is detached or non-detached.

4.2.2. Estimated Results for Building Use in Step2

Table 3 shows the results of building use estimation for detached buildings in Step 2 for each municipality in the target area, and Table 4 shows the results of building use estimation for non-detached buildings. The accuracy of the estimation of detached buildings was generally lower than that of Step 1. This is because many small detached offices locate in residential areas where detached houses are densely located, and the difference in appearance characteristics between detached houses and detached offices in such areas is the reason for the lower accuracy.

On the other hand, the estimation result of non-detached buildings is a high rate of correct responses. However, precision and recall are low. This is because many apartment houses were correctly estimated, however many multi-use buildings were not estimated correctly. This is because apartment houses include buildings that are similar in appearance to multi-use buildings. Therefore, it is considered that the multi-use buildings were estimated as apartment houses.

Table 2. Accuracy for each municipality (Step1)

	Shinjuku-ku	Setagaya-ku	Hachioji city
Accuracy (%)	74.29	81.71	89.33
Precision (%)	83.14	90.14	90.90
Recall (%)	74.24	86.49	97.10
F1 score (%)	78.43	88.28	93.90

Table 3. Accuracy for each municipality (Step2: Detached Buildings)

	Shinjuku-ku	Setagaya-ku	Hachioji city
Accuracy (%)	60.29	59.13	66.21
Precision (%)	65.91	58.72	65.32
Recall (%)	75.91	90.17	94.70
F1 score (%)	70.56	71.13	77.31

4.2.3. Discussion for Improvement of Accuracy

The accuracy of Step 2 was lower than that of Step 1. This may be due to the fact that the differences in appearance between detached and non-detached buildings were relatively easy to distinguish in Step 1, whereas the differences in appearance between buildings were not so apparent in Step 2. We think that the accuracy of estimating building use can be improved by learning not only the appearance of the building, but also information about the surrounding environment and the area of the building.

4.3. Estimation of building use using images that incorporate surrounding information

To learn information on surrounding buildings and building area, a 100 meter square grid around the center of gravity of a building was cut out from a satellite image of the target area, as shown in Figure 4. Then, deep learning was performed using the method described in Section 4.1 to estimate the building use.

4.3.1. Estimated Results for Building Use (Step1)

Table 5 shows the results of the building use estimation in Step 1 for each municipality in the target area. Setagaya-ku and Hachioji City were able to estimate the use of detached or non-detached buildings with high accuracy, while Shinjuku-ku showed low accuracy. This is due to the fact that in Shinjuku-ku, the estimation of detached buildings is not accurate. In Setagaya-ku, detached and non-detached buildings are distributed throughout the city , but non-detached buildings are distributed relatively more along trunk roads and railroad tracks.,. In Hachioji City, there are areas with a high density of detached houses and areas with a high density of apartment houses, so the differences in characteristics between the images of detached and detached buildings are clearly visible. The difference between the images of detached buildings and those of non-detached buildings is clear. In Shinjuku-ku, on the other hand, detached buildings and non-detached buildings are distributed mixed throughout the city, so the difference in features between the images of detached and non-detached buildings is not so apparent, which is considered to reduce the accuracy.

4.3.2. Estimated Results for Building Use (Step2)

Table 6 and Table 7 show the results of building use estimation for detached buildings and non-detached buildings for each municipality of Step 2. In the estimation of detached buildings, the accuracy rate was low in Setagaya-ku and Hachioji City, and slightly higher in Shinjuku-ku. This is because Setagaya-ku and Hachioji city are distributed throughout detached house and detached office, while detached offices are relatively densely distributed around stations in Shinjuku-ku. On the other hand, while the accuracy of the estimation of non-detached buildings was high in Hachioji City, it was low in Shinjuku-ku and Setagaya-ku. This is because Hachioji City is divided into areas with a high density of mixed mixed-use buildings, such as the area around Hachioji Station, and areas with a high density of apartment buildings, such as “Tama New Town” which is one of the Japan's largest large-scale residential complexes developed more than half a century ago. In Shinjuku-ku and Setagaya-ku, however, there is a mixture of residential and commercial multi-use buildings and apartment buildings, and there was little difference in the characteristics of the images by building use.

Table 4. Accuracy for each municipality (Step2: Non-detached Buildings)

	Shinjuku-ku	Setagaya-ku	Hachioji city
Accuracy (%)	64.55	79.76	74.66
Precision (%)	51.69	41.60	37.88
Recall (%)	52.55	48.60	43.86
F1 score (%)	52.12	44.83	40.65

Figure 4. Example of an image cropped in a 100m x 100m area around the center of gravity of a building

Table 5. Accuracy for each municipality (Step1)

	Shinjuku-ku	Setagaya-ku	Hachioji city
Accuracy (%)	66.16	76.07	86.37
Precision (%)	79.76	87.32	89.47
Recall (%)	66.74	82.45	95.16
F1 score (%)	72.67	84.82	92.22

Table 6. Accuracy for each municipality (Step2: Detached Buildings)

	Shinjuku-ku	Setagaya-ku	Hachioji city
Accuracy (%)	68.79	60.12	55.89
Precision (%)	70.02	61.16	53.43
Recall (%)	84.33	88.88	93.82
F1 score (%)	76.51	72.46	68.09

Table 7. Accuracy for each municipality (Step2: Detached Buildings)

	Shinjuku-ku	Setagaya-ku	Hachioji city
Accuracy (%)	62.63	69.52	77.50
Precision (%)	53.18	40.67	45.08
Recall (%)	57.11	59.89	60.44
F1 score (%)	55.08	48.44	51.64

4.4. Discussion for Improvement of Accuracy

The lower accuracy in Step 2 compared to the high accuracy in Step 1 can be attributed to the fact that in Step 1, the differences in appearance characteristics between detached and non-detached buildings were relatively clear. In contrast, in Step 2, the differences in appearance characteristics between individual buildings were not clear. This suggests that improving the efficiency of learning image features from satellite images could potentially address this problem. One approach could be to use models with more convolutional layers, such as Resnet50 or Resnet101, which may be more effective at learning subtle differences in image features than the Resnet18 model used in this study.

In addition, Section the model in Section 4.1 exhibited higher overall accuracy than the model in Section 4.3. However, in some municipalities, the model in Section 4.3 outperformed model in Section 4.1 on building use estimation. These results suggest that the choice of the region extracted from satellite images may influence accuracy. Therefore, adjusting the size of the region cropped from satellite images could potentially lead to improved accuracy.

Furthermore, this study created training data using only building images from three municipalities. However, there were fewer multi-use buildings compared to detached houses and apartment buildings. Consequently, it is likely that the features of multi-use buildings were not adequately learned. To resolve this problem, including regions with similar characteristics to the three target municipalities in the study area and creating more images of multi-use buildings could help improve accuracy by allowing for better learning of the features of multi-use buildings in such areas.

5. CONCLUSION

This study developed a method for identifying buildings where residents are distributed in developing countries. We used deep learning to extract buildings from satellite images and estimate their use, ultimately identifying buildings with resident distribution. As a result, building extraction from satellite images achieved an average IoU of 64.3% and an average extraction rate of 62.1%. In addition, the estimation of building use successfully classified detached buildings and non-detached buildings with an average accuracy of 78.7%. Additionally, the categorization of detached buildings into residential and non-residential achieved an average accuracy of 62.4%, while the classification of non-detached buildings into apartment housing and other buildings achieved an average accuracy of 75.2%.

In the future, we aim to improve the accuracy of the estimation of building use by first improving and expanding the training data. In addition, we aim to broaden the scope of this method by expanding the target area. Furthermore, the building detection (Section 3) and the building use estimation (Section 4) were developed independently and verified for accuracy in this paper. However, these processes should be performed in a single step to verify the accuracy of the results.

In addition, although this study was conducted in Tokyo due to the availability of highly institutionalized validation data, to achieve the objectives of this study, it is necessary to conduct in urban areas of developing countries to demonstrate that our method can be applied as well. In the future, we aim to develop detailed population data in developing countries by realizing a method to detect the distribution of buildings by use from satellite images and to estimate the number of residents in each building.

Acknowledgement

This work was supported by KAKENHI Grant Number JP20H01483.

References

Agudelo-Vera, C.M., Mels, A.R., Keesman, K.J., Rijnaarts, H.H., 2011: Resource management as a key factor for sustainable urban planning. J. *Environ, Manage*., 92(10), 2295-2303.

Akiyama, Y., Miyazaki, H, Sirikanjanaanan, S., 2019: Development of micro population data for each building: Case study in Tokyo and Bangkok. *2019 First International Conference on Smart Technology & Urban Development (STUD). IEEE,*. p. 1-6.

Ali, A., Shamsuddin, S.M., Ralescu, A.L., 2013: Classification with class imbalance problem. Int. J. *Advance Soft Compu*. Appl, 5.3.

Atem, A. WorldPop, open data for spatial demography, *Scientific data,* Vol. 4, 170004, 2017.

Balk, D., Yetman G., 2004: The Global Distribution of Population: Evaluating the gains in resolution refinement. *New York: Center for International Earth Science Information Network (CIESIN)*, Columbia University.

Doupe, P., Bruzelius, E., Faghmous, J, Ruchman, S. G., 2016: Equitable development through deep learning: The case of sub-nationa population density estimation, *Proceedings of the 7th Annual Symposium on Computing for Development*, 1-10.

Dulal, H.B., Brodnig, G., Onoriose, C.G., 2011: Climate change mitigation in the transport sector through urban planning: A review. *Habitat Int.*, 35(3), 494-500.

Facebookresearch:Detectron2,<https://github.com/facebookresearch/detectron2>, (accessed 2023.8.22)

Fuller, D., and Morency, P. 2013: A population approach to transportation planning: reducing exposure to motor-vehicles. *Journal of environmental and public health.*,2013

Iino, S., Ito, R., Doi, K., Imaizumi, T., Hikosaka, S., 2018: CNN-based generation of high-accuracy urban distribution maps utilising SAR satellite imagery for short-term change monitoring. *Int. J. Image Data Fusion*, 9(4), 302-318.

Kumar, J. Slums in India: a focus on metropolitan cities. *Int J Dev Res Full Length Res Artic*, 2014, 4: 388-393.

Niemelä, J., 1999: Ecology and urban planning. *Biodivers. Conserv.*, 8(1), 119-131.

MMPreTrain Contributors. 2023. OpenMMLab's Pre-training Toolbox and Benchmark. <https://github.com/open-mmlab/mmpretrain.> (accesed 2023.8.22)

Okada, K., Nishiyama, N., Akiyama, Y.,Miyazaki, H. and Miyazawa, S., Development of Detailed Building Distribution Map to Support Smart City Promotion -An Approach Using Satellite Image and Deep Learning-, *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, X-4/W3-2022, 189–196,2022

Polie V.F.L., 1984: Population estimation from aerial photographs for non-homogeneous urban residential area. *ITC Journal*, Vol. 1984–2, pp. 116–122.

Rumbach, A., 2016: Decentralization and small cities: Towards more effective urban disaster governance? *Habitat Int.*, 52, 35-42.

Robinson, C., Hohman, F., Dilkina, B., 2017: A deep learning approach for population estimation from satellite imagery. *In Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities* (pp. 47-54).

Taragi, R.C.S., Bisht, K.S., Sokhi, B.S., 1994: Generating population census data through aerial remote sensing. *J. Indian Soc. Remote Sens.* 22, 131-138. https://doi.org/10.1007/BF03024774