DEVELOPMENT OF DETAILED BUILDING DISTRIBUTION MAP TO SUPPORT SMART CITY PROMOTION -AN APPROACH USING SATELLITE IMAGE AND DEEP LEARNING–

K. Okada[1], N. Nishiyama[2], Y. Akiyama[1,*], H. Miyazaki[3], S. Miyazawa[4]

[1] Tokyo City University, 1-28-1, Tamazutsumi, Setagaya-ku,Tokyo, Japan - (g2281611,akiyamay)@tcu.ac.jp
[2] Yokosuka City Office, 11 Ogawa-cho, Yokosuka-shi, Kanagawa-ken,Japan - naojinishiyama@gmail.com
[3] Center for Spatial Information Science, The University of Tokyo, 5-1-5, Kashiwanoha, Kashiwa, Chiba, Japan - heromiya@csis.u-tokyo.ac.jp
[4] LocationMind Inc.,701, 3-5-2, Iwamoto-cho, Chiyoda-ku, Tokyo, 101-0032, Japan - miyazawa@locationmind.com

Commission IV, WG IV/9

KEY WORDS: Satellite Image, Building Detection, Building Use, Deep Learning, Population, Planning Support.

ABSTRACT:

Detailed demographics play an important role in the development of smart cities. However, especially in developing countries, the maintenance and management of this data is incomplete, which hinders the promotion of smart cities. The objective of this study is to develop a method to create detailed building distribution maps from satellite images, which will serve as a basis for developing detailed demographic data to support the promotion of smart cities around the world. The target area is several areas of Tokyo where validation data is available. We first developed a method for extracting buildings from satellite images and then estimating the building use to determine the buildings where residents are distributed. Both methods use deep learning. As a result, it was possible to extract buildings with an extraction rate (the number of buildings in the automatically extracted building data divided by the number of buildings in the data for verification) of up to 60.3% for the entire target area. In addition, in the estimation of building use, our method was able to classify detached and non-detached buildings with an average accuracy of 78.7% for the entire target area.

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). In Japan, for example, the results of the population census are used to forecast the future population (Kento et al., 2022), conduct transportation management (Kawasaki, 2015), and develop disaster prevention plans (Adu-Gyamfi, Shaw, 2021) and so on. In addition, in recent years, cities around the world have been accelerating their efforts to create smart cities, and it has been pointed that the development of detailed demographics is also important for creating resilient smart cities that can continue into the future, and specific applications that contribute to the realization of smart cities have been proposed (Bação et al., 2018, Akiyama et al., 2019;). Thus, demographic data is expected to play an important role not only in current urban planning but also in the realization of smart cities. 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).

* Corresponding author

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 imagery, 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-imagery-based population estimation (Polie, 1984; Taragi et al., 1994); although these studies provided useful information for the scheme of our study, 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 amount 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 condominiums 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.

However, these methods only monitor the distribution of build-up areas. 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, since the purpose of this study is to extract build-up areas, the building use is not analyzed.

Similarly, while there have been studies on extracting the spatial distribution of buildings from satellite images (Pan et al. , 2020), there are only a few methods that can estimate building use.

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 the building use is known, it would be possible to identify buildings with residential use that serve as residential areas. Similarly, if the number of floors of a building is known, 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.

1.2 Objective

The objective of this study is to develop a method to extract buildings from satellite images using deep learning and to estimate building use to identify buildings where residents are distributed from the extracted buildings, with the aim of overcoming the problems of existing research and contributing to the promotion of smart cities, especially in developing countries.

2. FLOW OF STUDY

Figure 1 shows the flow of this research. This study consists of two phases: Phase 1, Building detection, and Phase 2, Estimation of building use. In Phase 1, we develop a method to automatically extract buildings from satellite images using a convolutional neural network (CNN), a deep learning methodology. In Phase 2, we reuse the CNN to estimate the building use of the extracted building images. As a result, we can develop building maps with the attribute of building use from satellite images. Although the processing of Phase 1 and Phase 2 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 1 was developed independently in Section 3 and Phase 2 in Section 4.

2.1 Target area

Although 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 area of this study is Tokyo, where sufficient data for verification are available. The specific areas are Shinjuku-ku (ward), which has strong urban characteristics with skyscrapers districts and a large shopping district; Setagaya-ku, which is adjacent to Tokyo's City Center and contains mostly residential areas; and Hachioji city, which is a suburb of Tokyo and includes rural and mountainous areas. The method proposed in this study can be applied to different target regions, and it is expected that the usefulness of the method demonstrated in Tokyo will make it possible to apply it to cities in developing countries.

3. BUILDING DETECTION USING DEEP LEARNING

3.1 Method of Building Detection

To detect buildings in satellite images (background satellite images from Google Maps) using deep learning, we first created training data. First, satellite images of the entire target area were divided into a grid of 250 m squares, and building areas were extracted from each image by manually tracing the building perimeter lines using GIS, resulting in the extraction of 36,073 building areas from the entire target area. Next, using these data as training data, a model was built to automatically extract buildings from satellite images using deep learning. The deep learning used in this study was Faster-RCNN (Ren et al., 2015), an object detection method based on CNN (Krizhevsky et al., 2012). CNNs have been reported to significantly outperform existing methods for tasks such as image classification, object extraction, and region segmentation, and are also characterized by their ability to extract features directly from data and short processing time. Faster-RCNN uses a CNN structure called a region proposed network (RPN), which has improved speed and accuracy compared to conventional methods.

3.2 Result of Building Detection and Problems

Figure 2 shows an example of building detection results in the vicinity of Shinjuku Terminal, Shinjuku-ku. The results showed that the accuracy of the building data extracted from each satellite image differed. Therefore, we used 30 satellite images with different characteristics and compared them for each region with different building characteristics to verify whether a relationship exists between building characteristics and the accuracy of automatic building extraction.

Using the number of buildings in one image, the total area occupied by the buildings, the average area per building, and the area of the largest building as indicators of regional characteristics, we verified which regions have the highest

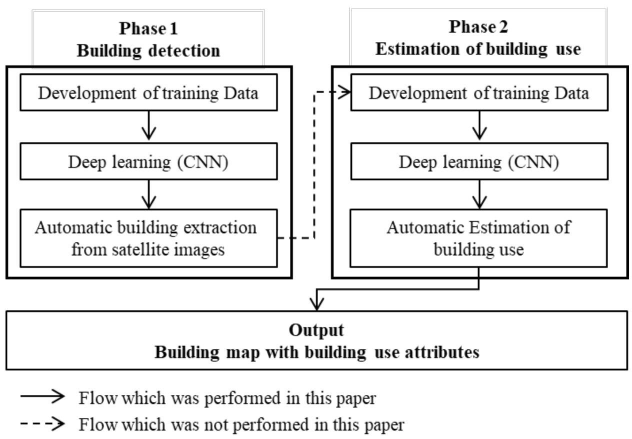

Figure 1. Flow of study

intersection over Union (IoU) and building extraction rate. IoU is the value obtained by dividing the intersection, which is the common part of building areas obtained from the map information of the automatically extracted building data and the verification data, by the sum set of the areas, and is one of the evaluation indices for object detection (Cai et al., 2022). The building extraction rate is the number of buildings in the automatically extracted building data divided by the number of buildings in the data for verification. The satellite images used were 30 images of Shinjuku Ward, Tokyo, which has diverse geographic features such as train stations, railroads, high-rise buildings, and residential areas. Building polygon data from a 2020 digital residential map was used for the validation data (correct values).

First, areas with high IoU were characterized by many buildings, a small average area per building, and a large total area occupied by buildings. In terms of the building extraction rate, it was found that the rate decreased as the number of buildings increased (Figure 3). The fact that IoU is particularly high in areas where the average area per building is small and the number of buildings is large (Figure 4) indicates that our method of automatic building extraction is useful in urban areas in developing countries, where uncontrolled urbanization and population influx have led to high building densities. As shown in Figure 5, the IoU was lower in areas with many buildings that have the opposite characteristics of those shown in Figure 3. In addition, as shown in Figure 6, in areas with a large number of buildings, the building extraction rate tends to be low because adjacent buildings cannot be distinguished from each other properly. This tendency can be regarded as an issue for automatic building extraction at the current stage.

3.3 Examination to improve extraction accuracy

One way to improve extraction accuracy is to change the numerical values of parameters related to deep learning. Therefore, this study examines changes in the accuracy index of automatic building extraction by varying the maximum number of iterations. In addition to IoU and the building extraction rate, we used the root mean square error (RMSE) of the average distance between the centroid of each building polygon in the extracted building data and the residential map as an evaluation index. In this study, the maximum number of iterations was increased to 900,000.

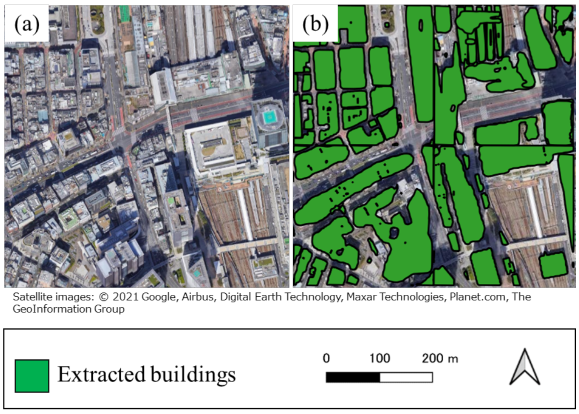

Figure 2. Example of building detection results （Vicinity of Shinjuku Terminal, Shinjuku-ku）
（a: Satellite image b: Extracted building data）

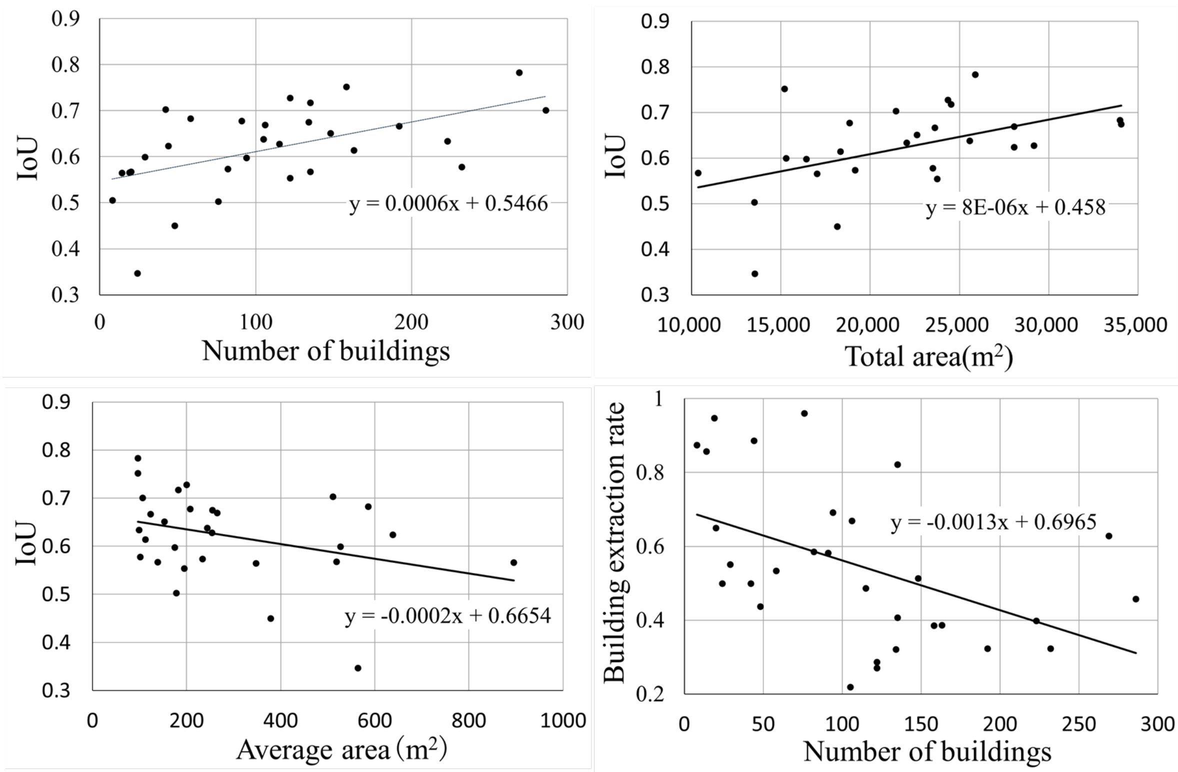

Figure 3. Relationship between indicator of regional characteristics and extraction accuracy

Figure 7 and Table 1 show the results. These results indicate that there is a positive correlation between the parameter values and IoU. While the building extraction rate tends to decrease as the parameter value increases, the RMSE also decreases, indicating that although the building extraction rate decreases, the location accuracy of each extracted building increases. The IoU and the building extraction rate decreased with the number of iterations being 400,000, and the RMSE of the centroid decreased with the number of iterations (600,000), indicating that the optimal parameter values for the CNN used in this study are between 400,000 and 600,000. However, considering that increasing the parameter value by 200,000 increases the time required to create a Geopackage file with automatically extracted building data using Colab Pro for deep learning by 48 hours, it is appropriate to set the parameter value at 400,000.

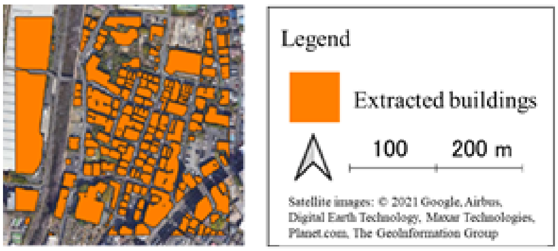

Figure 4. Example of region where the IoU is high

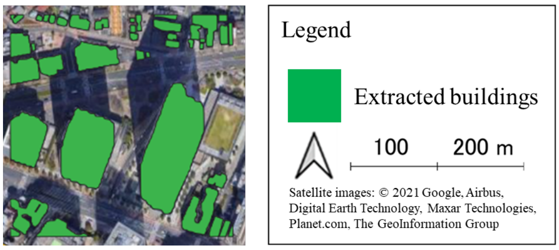

Figure 5. Example of region where the IoU is low

4. ESTIMATING BUILDING USE DURING DEEP LEARNING

4.1 Estimation of building use (5 class)

First, the data of building uses (detached houses, detached offices, multi-use buildings (mainly residential use), multi-use buildings (mainly business use), and condominiums) from the building polygon data of the residential map in 2020 were spatially integrated with the satellite image from which the areas of buildings were extracted from satellite images of the entire area under study (Figure 8). Next, to include the environment around the building, we defined the image size by the longer length of the vertical and horizontal sides of the building and cropped the building from the satellite image with

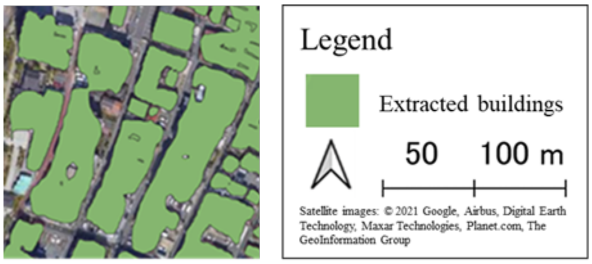

Figure 6. Example of region where the number of buildings is large

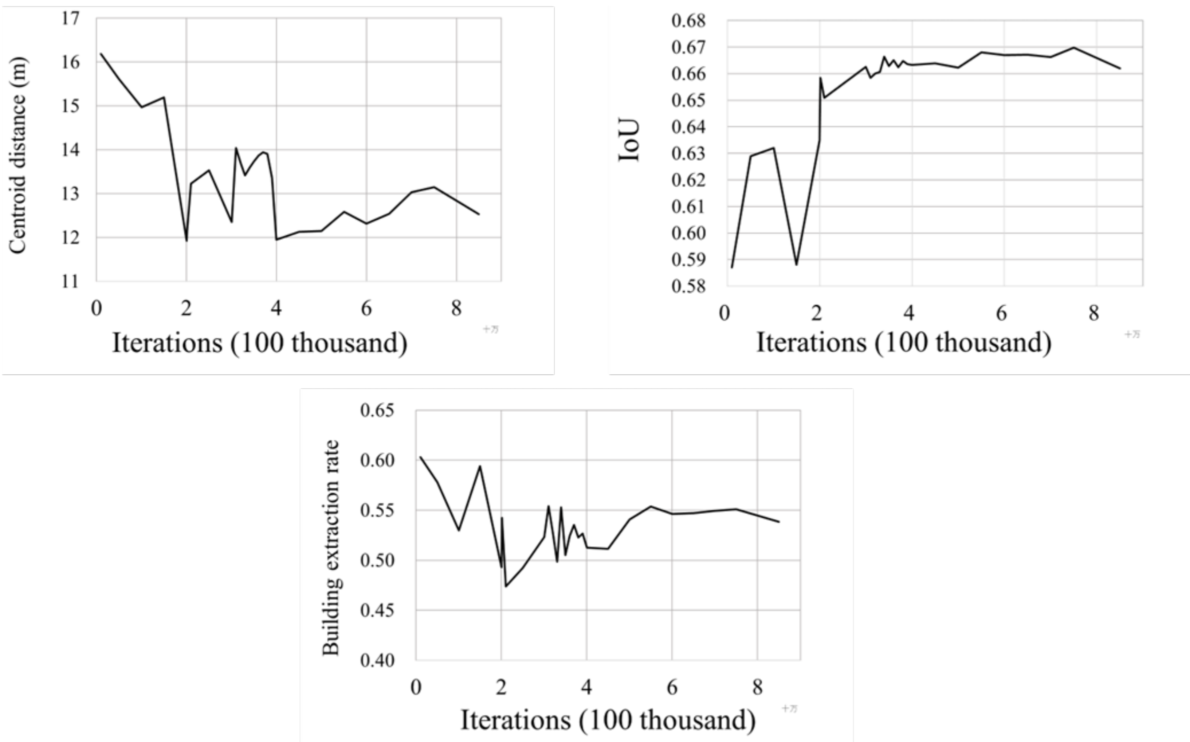

Figure 7. Relationship between iteration, centroid distance, IoU, and building extraction rate

a 10% margin. We performed deep learning using the spatially integrated data of the data as the training data. We used a fine-tuned ResNet 18 model as the learned model for deep learning. Then, 10% of the training data was diverted to test data for estimation of building use, and the reliability of the estimation was confirmed by comparing the estimation results of the test data with the building uses on the residential map.

4.2 Estimated results for building use (5 class)

Table 2 shows the estimation accuracy for each municipality and Table 3 shows the estimation precision per building use. Estimation accuracies were low for all municipalities. The results for each building use showed that detached houses and detached offices were estimated to be present in all municipalities, but there were some municipalities that estimated zero cases for other building use. Table 4 shows the breakdown of the training data for each municipality by building use. The number of detached houses and detached offices was larger than that of other building use, leading to a larger amount of training data, which is thought to have resulted in higher precision than that for other building uses. In other words, the lack of and bias in the training data greatly affected the accuracy of the building use estimation in this study.

4.3 Examination to improve accuracy

A solution to biased training data is the under-sampling and over-sampling methods (Ali et al., 2013). Under-sampling is a method to reduce the number of data entries in a large number of classes to match the number of entries in a smaller number of classes, while over-sampling increases the number of data entries in a small number of classes by duplicating the data in a large number of classes to match the number of entries in a large number of classes. Another method is to simplify the task of image recognition to improve accuracy (Cai et al., 2019). Simplifying the task here means reducing the number of image classification classes to be performed at a time. Building use estimation in Section 4.1 classified images into five building use classes; however, reducing the number of classes to be classified is expected to improve accuracy.

4.4 Estimation of building use (2 class)

When classifying building use, a classification method in which the number of residents in each building is expected to differ significantly is to divide buildings by whether they are detached buildings (detached houses and detached offices) or non-detached. The former is expected to have fewer residents per building, while the latter is expected to have more. It is also expected that the appearance of the buildings is different.

Therefore, we first determined whether the buildings were detached or non-detached buildings and then undersampled and oversampled the training data. As shown in section 4.1, the building images are assigned the five types of building use.

	Iterations [Thousand]				
	10	50	200	400	900
IoU	0.578	0.629	0.635	0.663	0.665
Building extraction rate	0.603	0.578	0.493	0.512	0.547
RMSE	16.179	15.613	11.923	11.948	12.687

Table 1. Relationship between iterations and accuracy

Therefore, detached houses and detached offices are classified as detached buildings, and the rest are classified as non-detached buildings. Second, we divided the training data into two building uses (detached buildings and non-detached buildings) and performed deep learning to estimate the building use. To eliminate unequal training data for each class, we set the number of training data for each class to 5,000. For classes with many training data, the number of entries was reduced to 5,000, and for classes with insufficient training data, the number of training data was replicated to increase the training data to 5,000 entries.

4.5 Estimated results for building use (2 class)

Figure 9 shows an examples of building image estimated as detached buildings, and Figure 10 shows an example estimated as non-detached buildings. Table 5 shows estimating accuracy for each municipality, and Table 6 shows recall for each of the five types of building attributes. In this estimation, accuracy exceeded 70% in all municipalities. It indicates that the building

Figure 8. Example of building extraction from satellite image

Municipality	Accuracy [%]
Shinjuku-ku	6,37
Setagaya-ku	6,37
Hachiouji city	1,08

Table 2. Accuracy for each municipality

		Precision [%]		
		Shinjuku-ku	Setagaya-ku	Hachiouji city
Building use	Detached house	10,96	12,41	16,17
	Detached office	19,16	12,23	2,99
	Multi-use building (mainly residential use)	10,01	-	3,15
	Multi-use building (mainly business use)	0,25	0,22	-
	Apartment house	0,86	0,07	-

Table 3. Precision per building use

use could be estimated with relatively high accuracy. In all municipalities, the accuracy of estimating detached houses as detached buildings and that of estimating condominiums as non-detached buildings were high. However, the accuracy of estimating detached office buildings as detached buildings was not as high. This is because many of the buildings included in the detached offices have the same appearance as multi-use buildings.the detached offices have the same appearance as multi-use buildings.

4.6 Building use estimation for detached buildings

Third, we verified the accuracy of estimating whether a building is a detached house or a detached office. The number of training data for each building use was set to 5,000 for both. Table 7 shows estimating accuracy for each municipality, and Table 8 shows recall for each building use. The result shows that the accuracy of estimating a detached building as either a detached house or a detached office was higher for detached houses in Setagaya-ku and detached offices in Hachioji city. On the other hand, accuracies for the building use for other municipalities are about 60%. Figure 11 shows an example of an image of a building that is a detached house building which was estimated as detached offices, and Figure 12 shows an example of an image of a building that is a detached office which was estimated as a detached house. The results suggest that buildings with a small area tend to be estimated as detached

		Number of training data		
		Shinjuku-ku	Setagaya-ku	Hachiouji city
Building use	Detached house	11,304	78,774	75,460
	Detached office	4,424	13658	10,706
	Multi-use building (mainly residential use)	2,209	2,209	2,209
	Multi-use building (mainly business use)	1,239	1,241	1,241
	Apartment house	7,520	24,469	8,237

Table 4. Breakdown of training data by building use for each municipality

Figure 9. Example of building images estimated as detached buildings

Figure 10. Example of building images estimated as non-detached buildings

Municipality	Accuracy[%]
Shinjuku-ku	74,83
Setagaya-ku	75,07
Hachiouji city	86,05

Table 5. Accuracy for each municipality

		Recall[%]		
		Shinjuku-ku	Setagaya-ku	Hachiouji city
Building use	Detached house	84,15	80,11	89,55
	Detached office	50,00	51,93	66,17
	Multi-use building (mainly residential use)	84,84	84,84	93,14
	Multi-use building (mainly business use)	94,87	95,51	98,72
	Apartment house	69,26	69,84	76,04

Table 6. Recall per building use

Municipality	Accuracy[%]
Shinjuku-ku	68.68
Setagaya-ku	77.27
Hachiouji city	60.05

Table 7. Accuracy for each municipality

		Recall[%]		
		Shinjuku-ku	Setagaya-ku	Hachiouji-city
Building use	Detached house	69,36	83.10	58.39
	Detached office	66.98	56.38	75.35

Table 8. Recall per building use

Figure 11. Example of detached house images estimated as detached houses

Figure 12. Example of detached office images estimated as detached offices

Municipality	Accuracy [%]
Shinjuku-ku	64,31
Setagaya-ku	66,19
Hachiouji city	81,08

Table 9 Accuracy for each municipality

houses, while buildings with a large area tend to be estimated as detached offices. Therefore, the inclusion of detached houses with relatively large building areas and detached offices with relatively small building areas is considered to have affected the accuracy.

4.7 Building use estimation for non-detached buildings

Finally, we tested the accuracy of the estimation of whether a non-detached building is a multi-use building (mainly residential), a multi-use building (mainly business), or a condominium. Table 9 shows estimating accuracy for each municipality, and Table 10 shows recall for each building use. As a result, accuracy for Hachioji City was about 80%. As in section 4.6, the accuracy was affected by the fact that many buildings in Shinjuku and Setagaya wards have a relatively large area in relation to the building use.

5. DISCUSSION TO IMPROVE ACCURACY

The results of building use estimation showed that the accuracyof the estimation for detached offices was low. Therefore, we discuss methods to improve the accuracy of building use estimation.

First, hyperparameters such as batch size and number of epochs need to be set to appropriate values when performing deep learning (Feurer et al., 2019). The batch size is the number of data contained in each subset when the dataset is divided into

		Recall[%]		
		Shinjuku-ku	Setagaya-ku	Hachiouji city
Building use	Apartment building (mainly residential)	32,49	44,40	53,07
	Apartment building (mainly office buildings)	63,46	41,16	66,03
	Apartment house	73,83	67,81	90,88

Table 10. Recall per building use

several subsets for learning. The number of epochs is a value that determines the number of iterations of training on a single set of training data. In this study, the optimal hyperparameters for building use classification have not been determined. Therefore, it is expected that the accuracy will be improved by finding the appropriate hyperparameter values in the future. A possible method to find the optimal hyperparameters is to perform deep learning while changing the values of hyperparameters to find the parameter values with the highest accuracy. Another possible method is to use a tool such as Optuna (Preferred Networks, Inc., 2022) that can automatically tune hyperparameters.

Another possible method of improving accuracy is to increase the number of training data (Krizhevsky et al., 2012). In this study, only the building data in each municipality were used as training data to estimate the building use of each municipality. In the future, it is expected to be possible to improve the accuracy by learning more features of each building attribute by learning the data of all three municipalities as training data. In addition, the buildings' surroundings (e.g., commercial, residential) could be used as input for the deep learning model. In this deep learning model, buildings visible in satellite images are learned by labeling the use of the buildings. Therefore, it would be possible to improve the accuracy by adding information such as land use type.

6. CONCLUSION

The method in this study made it possible to detect buildings from satellite images with a certain degree of reliability. This method is useful in urban areas in developing countries because it can detect buildings with high accuracy even in areas where the number of buildings is large and the average area per building is small. On the other hand, while the method used was able to accurately estimate the use of detached houses or condominiums, the accuracy was low for some types of building use in some municipalities.

Our future endeavors are to improve the accuracy of building use estimation by setting optimal parameters and expanding training data. In addition, we aim to improve the accuracy of building extraction and building attribute estimation by expanding the target area. These improvements will make our method applicable over a wide area. Moreover, building detection (Section 3) and building use estimation (Section 4) were developed independently of each other and verified for accuracy. Ideally, however, these processes should be performed in a single step and the accuracy of the final results should be verified. To ensure sufficient accuracy in this series of processes, it is necessary to further improve the accuracy of building detection. Furthermore, in the future, we aim to develop detailed demographic data in developing countries by realizing a method to detect the distribution of buildings with their uses from satellite images and to estimate the number of residents living in those buildings.

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant Number JP20H01483.

REFERENCES

Adu-Gyamfi, B., Shaw, R., 2021: Utilizing population distribution patterns for disaster vulnerability assessment: Case of foreign residents in the Tokyo metropolitan area of Japan. *Int. J. Environ. Res. Public Health,* 18(8), 4061.

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, S*cientific data*, Vol. 4, 170004, 2017.

Bação, F. J. F. L., Henriques, R.,Antunes, J. 2018: Contribution Towards Smart Cities: Exploring Block Level Census Data for the Characterization of Change in Lisbon. Trends in Spatial Analysis and Modelling, 59-73.

Balk, D., Yetman G., 2004: The Global Distribution of Population: Evaluating the gains in resolution refinement. [Online]. https://www.researchgate.net/publication/228735948_The_Global_Distribution_of_Population_Evaluating_the_Gains_in_Resolution_Refinement (10 November 2019).

Cai, D., Chen, K., Qian, Y., Kämäräinen, J.K., 2019: Convolutional low-resolution fine-grained classification. *Pattern Recognit. Lett.*, 119: 166-171.
Cai, Z., Vasconcelos, N., 2019: Cascade R-CNN: high quality object detection and instance segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 43.5: 1483-1498.

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.

Feurer, M., Hutter, F., 2019: Hyperparameter optimization. In *Automated machine learning* (pp. 3-33). Springer, Cham.

Fuller, D., & Morency, P. 2013: Research Article A Population Approach to Transportation Planning: Reducing Exposure to Motor-Vehicles.

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.

Kawasaki, S., 2015: The challenges of transportation/traffic statistics in Japan and directions for the future. *IATSS Res.*, 39(1), 1-8.

Kento, K., Ma, J., Seto, T., Sekimoto, Y., Ogawa, Y., Omata, H., 2022: Development of current estimated household data and agent-based simulation of the future population distribution of households in Japan. *arXiv preprint arXiv:*2204.00198.

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

Krizhevsky, A., Sutskever, I., Hinton, G.E, Geoffrey E., 2012: Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process Syst.*, 25.

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

Pan, Z., Xu, J., Guo, Y., Hu, Y., Wang, G., 2020: Deep learning segmentation and classification for urban village using a worldview satellite image based on U-Net. *Remote Sens.*, *12*(10), 1574.

Preferred Networks, Inc. Optuna, ＜https://www.preferred.jp/ja/projects/optuna/＞, （15 February 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.

Ren, S., He, K., Girshick, R., Sun, J., 2015: Faster R-CNN: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process Syst.*, 28.

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 popul ation census data through aerial remote sensing. *J. Indian Soc. R emote Sens.* 22, 131-138. https://doi.org/10.1007/BF03024774