Automatic Generation Method of Horizontal Building Mask Images by Using a 3D Model with Aerial Photographs for Deep Learning

Improvement of training accuracy by removing thin clouds in aerial photographs using Generative Adversarial Network

○ Kazunosuke Ikeno[*1], Tomohiro Fukuda[*2] and Nobuyoshi Yabuki[*3]

*1 Div. of Sustainable Energy and Environmental Engineering, Graduate School of Engineering, Osaka University
*2 Assoc. Prof., Div. of Sustainable Energy and Environmental Engineering, Graduate School of Engineering, Osaka University, Ph.D.
*3 Prof., Div. of Sustainable Energy and Environmental Engineering, Graduate School of Engineering, Osaka University, Ph.D.

Keywords: Urban planning and design; Deep learning; Generative Adversarial Network (GAN); Semantic segmentation; Mask image; Training data.

1. Introduction

1.1. BACKGROUND

Information extracted from aerial photographs is widely used in urban planning and design. For example, green coverage rate and sky view factor can be measured and building location and exterior can be confirmed from photographs. As the use of unmanned aerial vehicle (UAV) technology has become more widespread, aerial photographs have become easier to take. Information that requires real-time properties, such as damage to buildings during a disaster, can be grasped using aerial photographs taken by UAVs. To obtain highly accurate information, it is necessary to capture many photographs in a short period of time. An effective method for detecting buildings in aerial photographs is to use artificial intelligence for understanding the current state of a target region.

Recently, methods have been proposed for object detection and segmentation by deep learning. These methods can quickly and automatically detect the target objects in an image. It is also possible to detect buildings in aerial photographs by using this method. The accuracy of building detection is greatly influenced by the quantity and features of the dataset used to train the model, and it is necessary to train the model adequately for each target area. However, the building mask images used to train the model are generated manually in many cases. Considerable time is required to generate mask images from aerial photographs for model training because one aerial photograph can contain many buildings and many sets of aerial photographs and mask images are needed to train the model. A lot of image editing software is available, such as Adobe Illustrator [1)] and GIMP [2)]. This editing software has a function for automatically clipping target objects. However, functions are not effective for generating specific mask images, such as buildings.

A method is proposed for automatically generating mask images of buildings, roads, and other objects by using virtual reality (VR) 3D models for deep learning [3)]. Since the appearance of a normal virtual model is similar to but not the same as a photograph, it is difficult to obtain highly accurate detection results in the real world even if the image is used for deep learning training. Texture mapping is a method for defining surface texture, or color information on a 3D virtual model. We can also use photographs as textures in this technology [4)]. To use photographs as texture improve representation of 3D virtual models. However, photographs may contain obstacles other than the object. When using an aerial photograph with thin clouds as a training dataset, the training accuracy is reduced. These thin clouds need to be removed when using as textures.

1.2. PREVIOUS RESEARCH

In recent years, many object detection and segmentation methods have been proposed that use deep learning. By providing training data, features are automatically calculated and objects are detected based on the calculated features. Methods that detect objective areas as rectangles in images by using a convolutional neural network (CNN) such as AlexNet [5)] and You Only Look Once (YOLO) [6)]. Semantic segmentation [7)] classifies each pixel into several categories and segments the objects in images by the silhouette. A system of automatically calculating green coverage rate and sky factor by semantic segmentation [8)] has also been developed. A deep CNN-based method for automatically detecting suburban buildings from high-resolution Google Earth images has been proposed [9)] as a building detection method using deep learning. A fused fully convolutional network model has been proposed to perform building

segmentation [10]. Some research is being done to improve the accuracy of Mask-R-CNN for detecting building footprint boundaries. A method combining Mask R-CNN with building boundary regularization [11] has been presented. A method for detecting different scales of building and segmenting buildings to have accurately segmented edges [12] has been proposed. However, the building mask images for training the model are generated manually in many cases, which requires considerable time and expense to build.

To solve this challenge, a method has been proposed for automatically generating mask images of buildings, roads, and other objects by using VR 3D models for deep learning [3]. By using the 3D virtual model, we can create datasets that include mask images easily and rapidly. Since normal virtual models do not have the realism of a photograph, it is difficult to obtain highly accurate detection results in the real world even if the image is used for deep learning training. High-precision rendering methods have been developed, but it is generally difficult to use such methods because many computers do not have high enough specs. Using textured 3D virtual models with photographs can solve this challenge [4]. Photographs may contain obstacles other than the object. To remove these obstacles in photograph, the image generation methods by using Generative Adversarial Network (GAN) [13] are used.

1.3. OBJECTIVE

The objective of this research is to propose an automatic generation method for horizontal building mask images by using 3D models with textured aerial photographs for deep learning. Specifically, we aim to improve the representation of the VR models by using textured aerial photographs on 3D models. Some aerial photographs include thin clouds, which degrade the image quality. The thin clouds on these aerial photographs are removed by using GAN for improving training accuracy. The proposed method can automatically generate mask images by using these 3D models and GAN.

2. Proposed method

Our proposed method automatically generates building mask images and aerial photographs. The generated mask images are used to train the deep learning model for semantic segmentation. The proposed method loads 3D models that include terrain and building objects, classifies the building class and others class, switches between a model with all objects and a model with only buildings, and generates two upper view images of the models from multiple viewpoints. Aerial photographs which include thin clouds are regenerated as images without thin clouds by using GAN. This system can generate multiple sets of mask images and aerial photographs without thin clouds from one 3D model. The flowchart and the conceptual diagram for generating the dataset are shown in Figures 1 and 2, respectively.

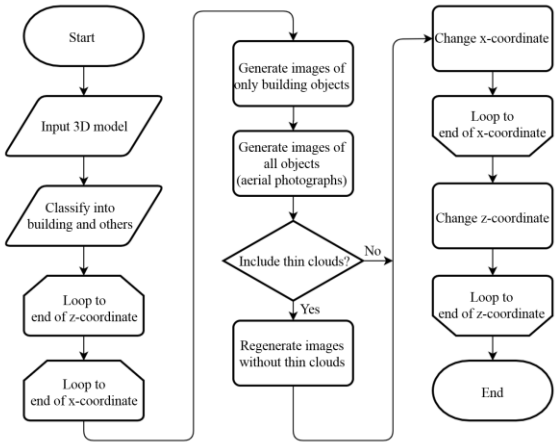


Figure 1. Flowchart of our proposed method

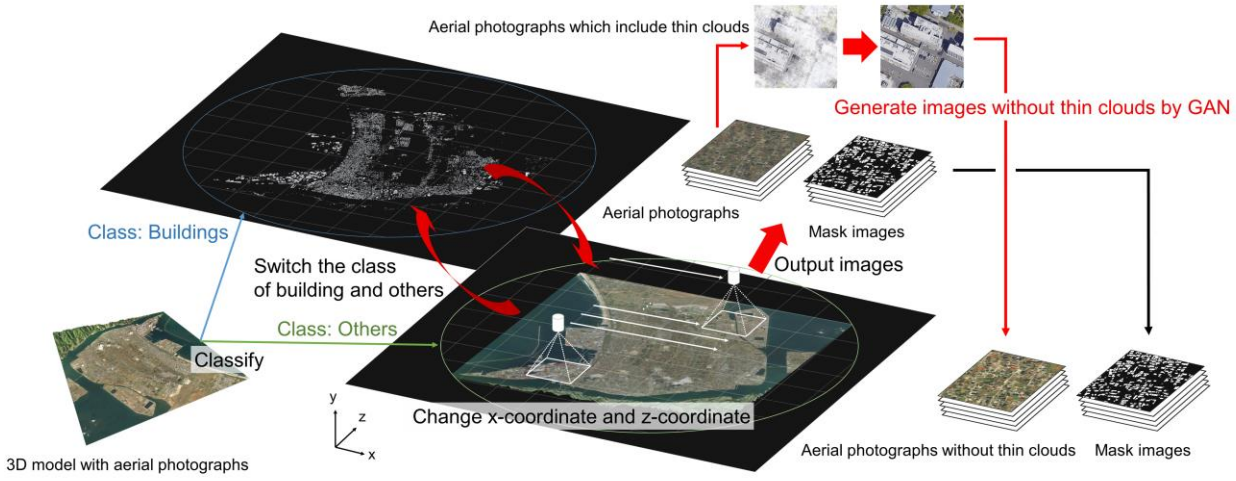


Figure 2. Conceptual diagram of our proposed method

3. Prototype system

A prototype system was constructed to generate sets of mask images and aerial photographs without thin clouds by our proposed method. The automatic mask image generation system by using a 3D model with aerial photographs is developed in the game engine [4]. To build systems to generate datasets, Unity was used as a game engine that can load 3D models.

To generate thin cloud removal images, we use spatial attention generative adversarial networks (SpA GAN) [14], which use SPatial Attention Network (SPANet) [15] as a generator. The architecture of SpA GAN is shown on Figure 3. The SpA GAN model trained by using the open source RICE dataset [16] is used to generate thin cloud removal images from aerial photographs that include thin. The color tone of the generated image is corrected to match the original aerial photograph. The specification of the PC is shown on Table 1.

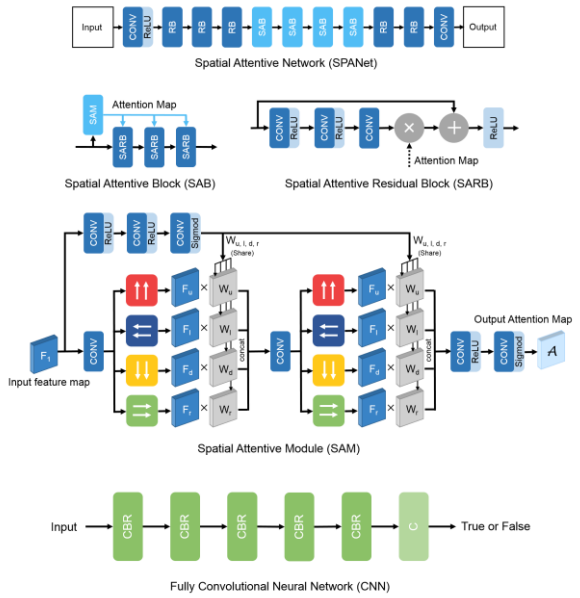


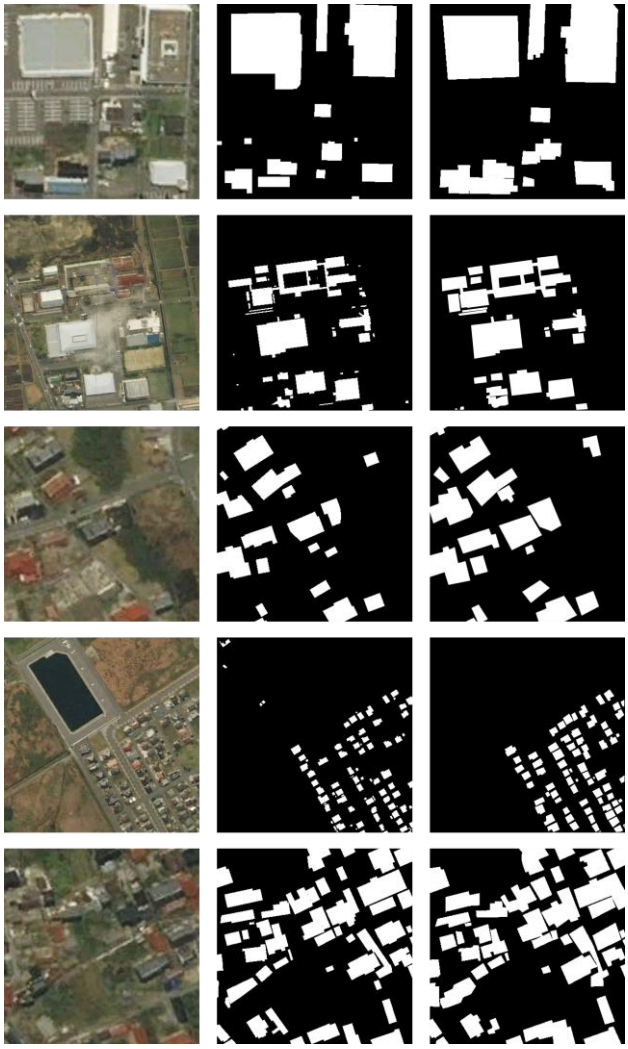
Figure 3. Generator and discriminator
(Created by the author with reference to the literature)

Table 1. The specification of PC

OS	Ubuntu 16.04 LTS
CPU	Intel(R) Core(TM) i7-3770K CPU @ 3.50GHz
GPU	Geforce GTX 1060
RAM	28.0 GB

4. Results

The sets of aerial photographs and mask images automatically generated by our proposed system are shown in Figure 4. The middle column shows automatically generated mask images and aerial photographs in which thin clouds are removed by the prototype system and the right column shows manually generated mask images. The white areas show the building mask. Our prototype system could generate 2912 sets in 219 seconds. The time required for the generation of 91 thin cloud removal images by GAN was 29 seconds in this time.



Photographs | Mask images Automatically | Mask images Manually

Figure 4. Generated aerial photographs and mask images

5. Discussion

Our prototype system generates 2912 sets of mask images and aerial photographs without thin clouds in 219 seconds. The time to generate the mask images was reduced by automatically generating them from 3D models in comparison to the manual generating method. The mask images generated by our prototype system are almost the same as the mask images generated manually. This system can generate mask images with detailed shapes. However, it cannot generate mask images of small warehouses. It is necessary to prescreen the generated mask images.

The aerial photographs before and after thin clouds removal by GAN are shown in Figure 5. The buildings that were covered by the thin cloud cover on the thin cloud removal image generated by the GAN are clearly visible. However, it was covered with thin clouds, and the buildings that were completely invisible remained hidden in the clouds.

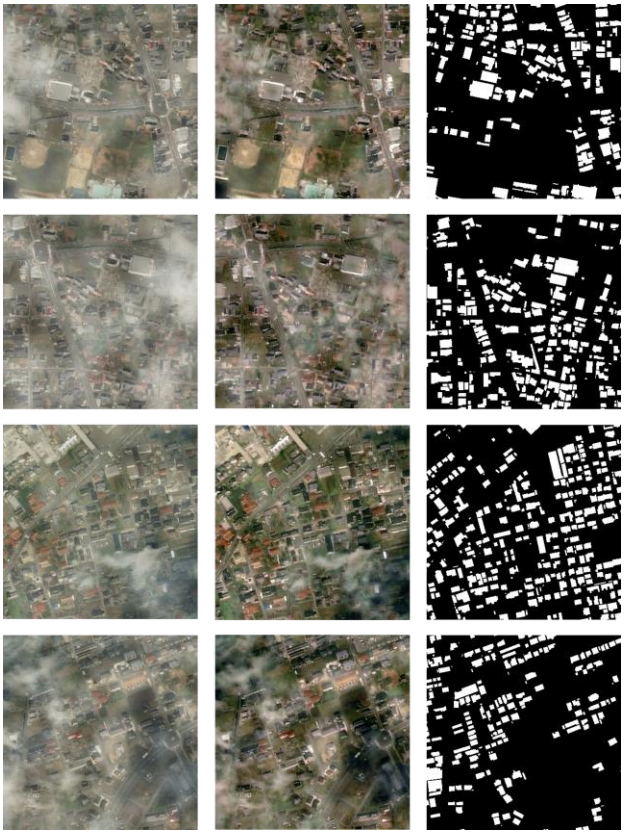


Figure 5. Generated aerial photographs and mask images

6. Conclusion

The conclusions of the present study are shown below.

- Our prototype system can generate sets of aerial photographs in which thin clouds are removed by GAN and mask images from a 3D model.
- The aerial photographs before and after thin clouds removal by GAN were compared.

Work left for the future includes training deep learning model by using the datasets generated by our prototype system, and evaluating the accuracy of the trained model.

References

1) Adobe: 2020, Photoshop <https://www.adobe.com/products/photoshopfamily.html> (accessed 13 September 2020).
2) The GIMP Team: 2020, GNU Image Manipulation Program (GIMP) <https://www.gimp.org/> (accessed 13 September 2020).
3) Fukuda, T., Novak, M., Fujii, H. and Pencreach, Y.: 2020, Virtual reality rendering methods for training deep learning, analysing landscapes, and preventing virtual reality sickness, International Journal of Architectural Computing. <https://doi.org/10.1177/1478077120957544>
4) Ikeno, K., Fukuda, T. and Yabuki, N.: 2020, Automatic Generation of Horizontal Building Mask Images by Using a 3D Model with Aerial Photographs for Deep Learning, Proceedings of eCAADe 2020, 2, 271–278.
5) Krizhevsky, A., Sutskever, I. and Hinton, G. E.: 2012, ImageNet classification with deep convolutional neural networks, Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS 2012), 1097–1105.
6) Redmon, J., Divvala, S., Girshick, R. and Farhadi, A.: 2016, You Only Look Once: Unified, Real-Time Object Detection, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779–788.
7) Long, J., Shelhamer, E. and Darrell, T.: 2015, Fully Convolutional Networks for Semantic Segmentation, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3431–3440.
8) Cao, R., Fukuda, T. and Yabuki, N., 2019, Quantifying Visual Environment by Semantic Segmentation Using Deep Learning, Proceedings of the 24th International Conference on Computer-Aided Architectural Design Research in Asia (CAADRIA 2019), 623–632.
9) Zhang, Q., Wang, Y., Liu, Q., Liu, X. and Wang, W.: 2016, CNN based suburban building detection using monocular high resolution Google Earth images, Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS), 661–664.
10) Bittner, K., Adam, F., Cui, S., Körner, M. and Reinartz, P.: 2018, Building Footprint Extraction From VHR Remote Sensing Images Combined With Normalized DSMs Using Fused Fully Convolutional Networks, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11, 2615–2629.
11) Zhao, K., Kang, J., Jung, J. and Sohn, G.: 2018, Building Extraction from Satellite Images Using Mask R-CNN with Building Boundary Regularization, Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 247–251.
12) Zhou, K., Chen, Y., Smal, I. and Lindenbergh, R.: 2019, Building segmentation from Airborne VHR Images Using mask R-CNN', Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., 155–161.
13) Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y.: 2014, Generative adversarial nets, NIPS, 2672–2680.
14) Pan, H. (Penn000): 2020, SpA-GAN_for_cloud_removal <https://github.com/Penn000/SpA-GAN_for_cloud_removal> (accessed 19 September 2020).
15) Wang, T., Yang, X., Xu, K., Chen, S., Zhang, Q. and Lau, R.: 2019, Spatial Attentive Single-Image Deraining with a High Quality Real Rain Dataset, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 12262–12271.
16) Liu, D. (BUPTLdy): 2019, RICE_DATASET <https://github.com/BUPTLdy/RICE_DATASET> (accessed 19 September 2020).