

An Image Augmentation Method for Detecting Construction Resources Using Convolutional Neural Network and UAV Images

S. Bang^a, F. Baek^b, S. Park^a, W. Kim^a, and H. Kim^a

^aDepartment of Civil and Environmental Engineering, Yonsei University, South Korea

^bIndustrial Technical Laboratory, Yonsei University, South Korea

E-mail: bangdeok@yonsei.ac.kr, fbeak@yonsei.ac.kr, somin109@yonsei.ac.kr, wontkim@hotmail.com, hyoungkwan@yonsei.ac.kr

Abstract –

Images acquired by UAV can be analyzed for resource management on construction sites. However, analyzing the construction site images acquired by UAV is difficult due to the characteristics of UAV images and construction site images. This paper proposes an image augmentation method to improve the performance of an object detection model for construction site images acquired by UAV. The method consists of three techniques: intensity variation, image smoothing, and scale transformation. Experimental results show that the method can improve the performance of the detection model (Faster R-CNN) by achieving a recall and a precision of 53.08% and 66.76%, respectively. With future studies, the method is expected to contribute to UAV-based resource management on construction sites.

Keywords –

Image augmentation method; Faster R-CNN; UAV; Resource management

1 Introduction

Previous studies have been conducted to analyze the images acquired by Unmanned Aerial Vehicle (UAV) for on-site construction management [1,2]. UAV is advantageous in acquiring images on a wide area in a short time, unlike closed-circuit television (CCTV). The use of UAV is more effective in large-scale construction sites where daily monitoring is difficult to perform. For generating managerial information on construction sites, it is necessary to recognize construction resources such as workers, construction equipment, and materials. The information on construction resources can be used for safety assessment, productivity analysis, and material tracking.

The object detection methods using image

processing and machine learning algorithms have been proposed for construction resources [3-5]. Although previous studies have shown excellent performance in resource detection on construction sites, it is difficult to directly apply them to analyzing construction site images acquired by UAV, as shown in Figure 1. First, the quality of the images varies due to the diverse conditions such as the flight situation of UAV, external illuminations, and weather conditions. Second, vibration in UAVs makes the acquisition of blurred images frequent. Third, the same object is recognized as a different size based on the UAV position.

Image detection models should be trained from datasets with a large number of labeled data for improving their generalization capability. In general, a dataset composed of construction site images acquired by UAV is extremely inconsistent because it has both characteristics of construction site images and UAV images. Therefore, the dataset requires a large number of labeled data reflecting various situations. However, the publicly available image data in the construction industry are scarce, unlike datasets consisting of millions of images with labels, i.e., ImageNet Large Scale Visual Detection Challenge (ILSVRC) dataset [6], COCO dataset [7], and Open Images dataset [8]. Kim et al. [3] released the dataset of 2,920 construction equipment images including dump truck, excavator, loader, mixer truck, and roller; however, its size is not big enough compared to the datasets in computer vision domain. Furthermore, it is difficult to find publically available datasets of construction site images acquired by UAV.

We propose an image augmentation method that increases the number of training images while preserving the labels of construction site images acquired by UAV. For solving the aforementioned challenges regarding the UAV images, three image augmentation techniques (intensity variation, image smoothing, and scale transformation) are applied to each



Figure 1. Examples of construction site images acquired by UAV

training image. Our datasets consist of 10 classes: workers, four types of materials (tarpaulin, rebar, H-beam, concrete pipe), and five types of equipment (drilling equipment, crane truck, excavator, concrete truck, and dump truck). The faster region-based convolutional neural network (Faster R-CNN) proposed by Ren et al. [9] is used for construction resource detection. To validate the effectiveness of the proposed method, we compare the detection performances between the Faster R-CNN trained with the 593 images and the network trained with the augmented 11,800 images. The experimental results show that the proposed method is effective in augmenting the training dataset for significantly improved detection results in the UAV images.

2 Related Works

Many researchers have proposed vision-based object detection methods for construction resources such as workers, materials, equipment, and buildings at construction sites. Golparvar-Fard et al. [10] presented a computer vision-based method for equipment action detection. A Support Vector Machine (SVM) classifier used in the method recognized and localized equipment actions using the features of Histogram of Oriented Gradients (HOG). Park and Brilakis [11] proposed a detection method for construction workers in video frames. They used background subtraction to detect moving objects and used HOG features and an SVM

classifier to extract objects like human-shape and used a hue, saturation, and value (HSV) color histogram and k-Nearest Neighbors (k-NN) to detect workers. Kim and Chi [4] presented a method for tracking construction equipment using functional integration of a detector and a tracker. They focused on solving the problem that the target equipment could be occluded due to its dynamic movements and site conditions. Hamledari et al. [12] suggested a computer vision-based algorithm for detecting the components of an interior partition. The partition including studs, electrical outlets, insulation, and three states for drywall sheets (installed, plastered, and painted) was detected by using image processing and machine-learning algorithms such as Otsu algorithm, k-mean clustering, and SVM. Kim et al. [13] proposed a data-driven scene parsing method to recognize the whole area of construction site images. The methods classified not only the construction resources such as workers, equipment, material, and structures, but also the background such as ground, sky, and tree.

The object detection methods using a convolutional neural network (CNN) have shown promising performances in computer vision tasks. These methods use the feature-maps extracted from the convolutional layer for finding candidate regions in which objects may exist. Girshick et al. [14] proposed the regions with convolutional neural network (R-CNN), which was known for the first attempt at implementing a CNN for object detection. In this network, the candidate regions extracted by using a selective search method [15] was input into the pre-trained AlexNet [16]. Girshick [17]

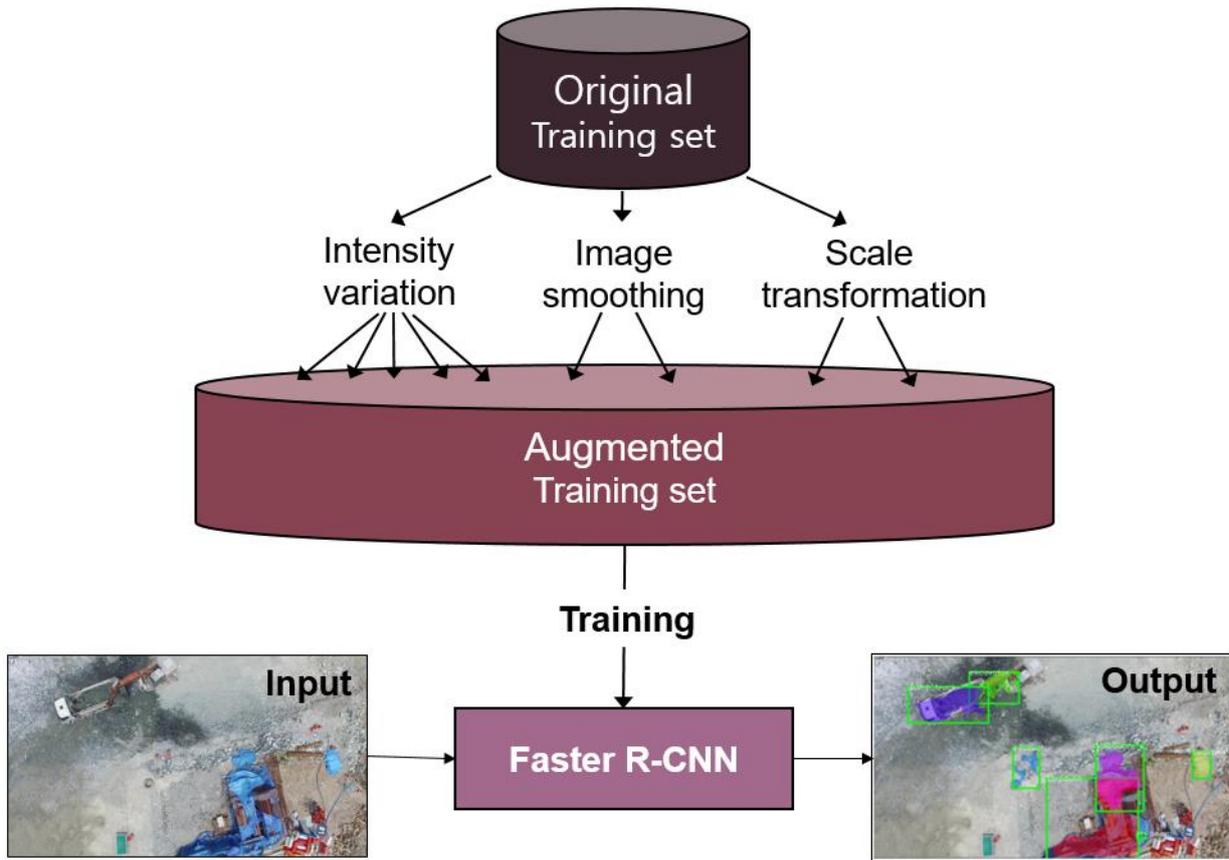


Figure 2. Overview of the image augmentation method

proposed Fast R-CNN to directly input the target image into the pre-trained AlexNet, unlike R-CNN. However, two networks (R-CNN, Fast R-CNN) took tens of seconds for processing an image because these networks used the selective search method to exhaustively generate bounding boxes regardless of aspect ratio and scale. To solve this problem, Ren et al. [9] proposed Faster R-CNN to perform the region proposal by adding a region proposal network (RPN) instead of using selective search methods. The RPN moved nine different size sliding windows on the feature-map obtained from the CNN and generated several bounding boxes with the expectation values indicating the probability that an object exists. This attempt has led Faster R-CNN to achieve promising performance in object detection.

CNN has been recently used successfully in construction resource detection. Kim et al. [3] proposed a method to detect construction site equipment using a region-based fully convolutional network (R-FCN). The R-FCN showed outstanding performance, achieving 96.33% mean average precision in identifying construction equipment. Kim et al. [18] proposed an integrated method of simulation and CNN-based

monitoring for an earthmoving process in a tunnel. The CNN was successfully used for detection of excavator and dump truck with a 99.09% average precision.

3 Image Augmentation Method with Faster R-CNN

An UAV-based resource management strategy on construction sites can be developed using a CNN, which has shown outstanding performances in object detection. Although UAV can efficiently acquire a large amount of data, the quality of construction site images acquired by UAV varies greatly depending on UAV flight situations and outdoor environments. Furthermore, it is difficult to obtain a sufficient number of labeled images due to the lack of publicly available dataset in the construction industry. In this context, image augmentation techniques are required to generate training images with preserving label information.

In this paper, we address three variables of construction site images acquired by UAV. The first variable is illumination changes due to ambient light exposure. The time of day, weather condition, and light exposure can affect the intensity of images. The second

variable is image quality change due to the vibration of UAV. Unlike other image collection devices such as CCTV, hand-held camera, and smartphone camera, UAV is likely to generate blurred images due to its continuous moving. The third variable is scale change in images due to the difference in flight height of UAV. The same visual data can be recorded as different sizes in different images depending on the distance between the construction site and the UAV.

Considering the three variables, we apply three image augmentation techniques: intensity variation, image smoothing, and scale transformation. Intensity variation is a technique to add or subtract the same value to all pixel in original images. Image smoothing is a technique to generate blurred images by applying a Gaussian filter to original images. In the case of the two techniques, we use the same label information as that of original images because the position of objects in augmented images does not change. Scale transformation is a technique to generate images using 2D scale transformation while preserving the aspect ratio of original images. In the case of third technique, the label information indicating the bounding box coordinates of objects are adjusted in accordance with the scale change of images. The following image augmentation techniques allow generating a sufficient number of training images in reflecting the characteristics of the construction site image acquired by the UAV. Figure 2 shows the framework of the proposed method.

4 Experimental Study

The effectiveness of the proposed methodology was validated through experiments. All experiments were conducted on the Ubuntu 16.04 operating system with GTX 1080 Ti GPU and an Intel Core i7-7700 processor,

using the Python programming language and Tensorflow. The experimental dataset consists of 703 images collected at five construction sites. Among them, 593 were used as training images, and the remaining 110 were used as test images. We defined the construction resources as the 10 classes. The proposed method was used to generate 11,860 additional training images based on the 593 original images. Table 1 shows the parameters of three techniques. Figure 3 shows examples of the prediction results, and Table 2 summarizes a comparison of the performances. Because both methods learned the same network (Faster R-CNN), the computational speed at which the network processes images and detects construction resources is the same. The experimental results show that the image augmentation method improve the detection accuracy of the Faster R-CNN on construction site resources.

Table 1. The parameters of three techniques

Technique	Parameters
Intensity variation	(adding value) -60, -30, 0, 30, 60
Image smoothing	(σ value of Gaussian filter) 0, 1
Scale transformation	(scale factor) 0.8, 1.0

Table 2. Result evaluation on the dataset and computational speed

Method	Precision	Recall
Faster R-CNN trained with 593 images	39.10%	68.65%
Faster R-CNN trained with 11,860 augmented images	53.08%	66.76%

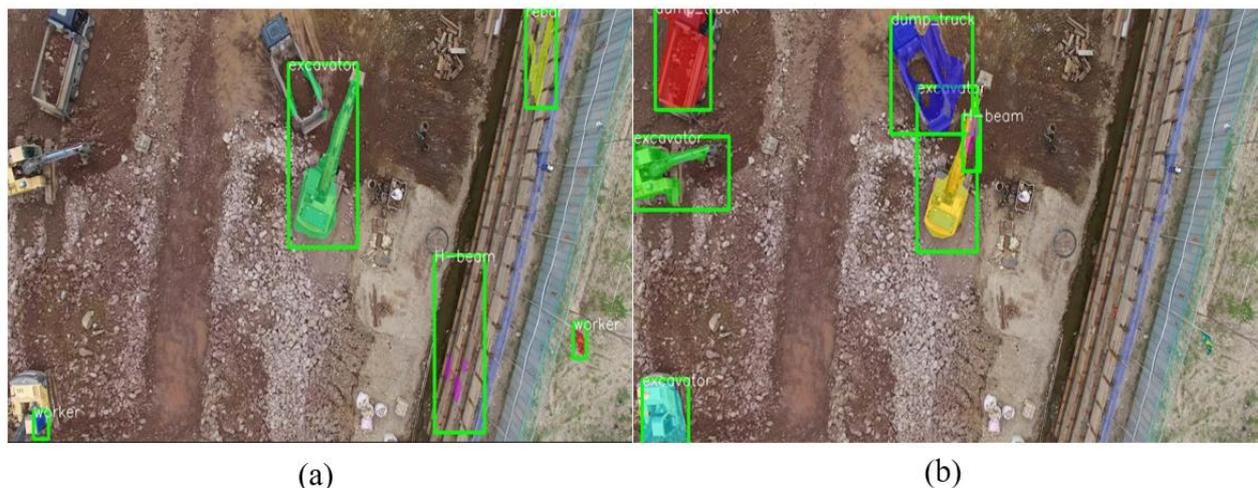


Figure 3. Comparison of construction resource detection results using Faster R-CNN: (a) a detection result without our method; (b) a detection result with our method

5 Conclusion

This paper proposes an image augmentation method for UAV-based resource management on construction sites. The proposed method consists of three image augmentation techniques: intensity variation, image smoothing, and scale transformation. Faster R-CNN, one of the CNNs achieving promising results in object detection, was used for validating the method. Of the 703 construction site images acquired by UAV, 593 were used to train Faster R-CNN, and the remaining 110 were used to test the network. The Faster R-CNN trained with 11,860 training images augmented by the method showed a recall of 53.08% and a precision of 66.76%, which is superior to the performances of the trained network without this method. However, this study has the following limitations. First, it is difficult to improve the ability to detect new objects in terms of shape or color, which do not exist in the training dataset. Second, since contextual information between objects and backgrounds are not generated, it is difficult to successfully detect objects in the images having a different context from training dataset. Future researches are needed to improve the detection performance for practical implications in construction sites.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science and ICT (No. 2018R1A2B2008600) and the Ministry of Education (No. 2018R1A6A1A08025348).

References

- [1] S. Bang, H. Kim, H. Kim, UAV-based automatic generation of high-resolution panorama at a construction site with a focus on preprocessing for image stitching, *Automation in Construction* 84 (2017) 70-80.
- [2] S. Siebert, J. Teizer, Mobile 3D mapping for surveying earthwork projects using an Unmanned Aerial Vehicle (UAV) system, *Automation in Construction* 41 (2014) 1-14.
- [3] H. Kim, H. Kim, Y.W. Hong, H. Byun, Detecting Construction Equipment Using a Region-Based Fully Convolutional Network and Transfer Learning, *Journal of Computing in Civil Engineering* 32 (2) (2017) 04017082.
- [4] J. Kim, S. Chi, Adaptive detector and tracker on construction sites using functional integration and online learning, *Journal of Computing in Civil Engineering* 31 (5) (2017) 04017026.
- [5] E.R. Azar, Construction equipment identification using marker-based recognition and an active zoom camera, *Journal of Computing in Civil Engineering* 30 (3) (2015) 04015033.
- [6] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, Imagenet large scale visual recognition challenge, *International journal of computer vision* 115 (3) (2015) 211-252.
- [7] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, Microsoft coco: Common objects in context, *European conference on computer vision*, Springer, 2014, pp. 740-755.
- [8] A. Kuznetsova, H. Rom, N. Alldrin, J. Uijlings, I. Krasin, J. Pont-Tuset, S. Kamali, S. Popov, M. Mallocci, T. Duerig, The Open Images Dataset V4: Unified image classification, object detection, and visual relationship detection at scale, *arXiv preprint arXiv:1811.00982* (2018).
- [9] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, *IEEE Transactions on Pattern Analysis & Machine Intelligence* (6) (2017) 1137-1149.
- [10] M. Golparvar-Fard, A. Heydarian, J.C. Niebles, Vision-based action recognition of earthmoving equipment using spatio-temporal features and support vector machine classifiers, *Advanced Engineering Informatics* 27 (4) (2013) 652-663.
- [11] M.-W. Park, I. Brilakis, Construction worker detection in video frames for initializing vision trackers, *Automation in Construction* 28 (2012) 15-25.
- [12] H. Hamledari, B. McCabe, S. Davari, Automated computer vision-based detection of components of under-construction indoor partitions, *Automation in Construction* 74 (2017) 78-94.
- [13] H. Kim, K. Kim, H. Kim, Data-driven scene parsing method for recognizing construction site objects in the whole image, *Automation in Construction* 71 (2016) 271-282.
- [14] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580-587.
- [15] J.R. Uijlings, K.E. Van De Sande, T. Gevers, A.W. Smeulders, Selective search for object recognition, *International journal of computer vision* 104 (2) (2013) 154-171.
- [16] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, 2012, pp. 1097-1105.

- [17] R. Girshick, Fast r-cnn, Proceedings of the IEEE international conference on computer vision, 2015, pp. 1440-1448.
- [18] H. Kim, S. Bang, H. Jeong, Y. Ham, H. Kim, Analyzing context and productivity of tunnel earthmoving processes using imaging and simulation, Automation in Construction 92 (2018) 188-198.