Automation in Construction 115 (2020) 103198

Contents lists available at ScienceDirect

Automation in Construction

journal homepage: www.elsevier.com/locate/autcon

Image augmentation to improve construction resource detection using generative adversarial networks, cut-and-paste, and image transformation techniques

Seongdeok Bang, Francis Baek, Somin Park, Wontae Kim, Hyoungkwan Kim*

School of Civil and Environmental Engineering, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea

ARTICLE INFO

Keywords:
Data augmentation
Construction resource detection
On-site management
GAN
UAV

ABSTRACT

The paper proposes an image augmentation method to construct a large-size dataset for improving construction resource detection. The method consists of three techniques: removing-and-inpainting, cut-and-paste, and image-variation. The removing-and-inpainting technique arbitrarily removes objects from images and reconstructs the removed regions via generative adversarial networks (GAN). The cut-and-paste technique extracts objects from the original dataset and places them into the reconstructed images via the previous technique. The image-variation technique applies three image transformation techniques, intensity-, blur- and scale-variation, to the images. To evaluate the method, 656 unmanned aerial vehicle (UAV)-acquired construction site images were used as the original dataset. A faster region-based convolutional neural network (Faster R-CNN) trained with the augmented training dataset achieves better performance, which is higher than that of a network trained with the original dataset. These results prove that the method is optimal for improving construction resource detection in UAV-acquired images.

1. Introduction

Information retrieved from images has been widely used in on-site construction management applications such as safety assessment [1–8], productivity analysis [9–13], and progress monitoring [8,14–19]. These applications can only be developed with a high recognition rate for construction resources. Vision-based construction site monitoring methods have used image processing and machine learning algorithms to identify workers, equipment, materials, and structures [20–24]. These methods mainly rely on supervised learning, a technique of training computers to find a specific pattern from given data and their labels. Thus, a sufficient amount of labeled training data is required to recognize construction resources well.

The construction industry lacks publicly available datasets such as the ImageNet Large Scale Visual Detection Challenge (ILSVRC) dataset [25], COCO dataset [26], and Open Images dataset [27]. Researchers in the construction industry have had to collect data and make their labels manually for vision-based monitoring studies [28–30]. As these tasks are labor-intensive, their studies were able to identify construction resources only in limited circumstances. However, to help managers grasp the status of the site on time, the resources should be understood in the context of a wide range of construction sites under various circumstances.

Advances in unmanned aerial vehicle (UAV) technology allow for more efficient image collection at construction sites compared to other devices such as smartphone cameras, closed-circuit television (CCTV), and camcorders [31–35]. Despite the advantages of UAVs, it is difficult to use them for on-site construction management, because object detection is challenging in construction site images acquired by UAVs. First, the images are highly dependent on external conditions, such as the time of the day and the weather conditions because construction sites are exposed to the natural environment. Second, the images may be blurred due to the vibration of the UAV itself. Third, the same objects can exhibit different sizes and shapes depending on the altitude and the location of the UAV. Finally, the number of objects in different classes is not constant in the acquired images. At construction sites, the number of workers or materials is generally larger than the amount of equipment. This imbalance of class distribution makes it difficult to train a detection model properly. Improving the performance of the model requires a balanced, large-size dataset with class-distribution that contains multiple patterns of construction sites with label information.

Increasing the number of training images does not always improve

* Corresponding author.
E-mail addresses: bangdeok@yonsei.ac.kr (S. Bang), fbaek@yonsei.ac.kr (F. Baek), somin109@yonsei.ac.kr (S. Park), wontkim@hotmail.com (W. Kim), hyoungkwan@yonsei.ac.kr (H. Kim).

https://doi.org/10.1016/j.autcon.2020.103198
Received 23 September 2019; Received in revised form 20 March 2020; Accepted 21 March 2020
0926-5805/ © 2020 Published by Elsevier B.V.

the performance of a detection model [36]. The training images should be created so that the model can learn the features of the construction resources correctly. Thus, herein, an image augmentation method to overcome the difficulties of analyzing construction site images and create a large-size dataset using generative adversarial networks (GAN), cut-and-paste, and image transformation is proposed. There are three techniques in the proposed method: removing-and-inpainting, cut-and-paste, and image-variation. These three techniques aim to address the class-distribution imbalance in the dataset, and to diversify the pattern of the construction resources in the images of the dataset. To evaluate the effectiveness of the method, the dataset consisting of UAV-acquired construction site images was created for 10 classes: construction workers, four kinds of materials (tarpaulin, rebar, H-beam, and concrete pipe), and five types of equipment (drilling, crane truck, excavator, concrete truck, and dump truck). The augmented dataset was created by the three techniques of the method, reflecting the various situations of construction sites and UAV flights. The experimental results show that the method can improve the performance of a detection model for construction resources.

2. Related works

2.1. Vision-based construction resource detection

Image processing has been used to recognize construction resources for monitoring construction sites in numerous studies. Some studies have focused on identifying one or two construction resources to develop construction management applications [4,6,23,24]. Park et al. [6] proposed a safety monitoring method that detects whether construction workers are wearing hard-hats. In their study, a hard-hat and worker's body were identified and matched using the Histogram of Oriented Gradients (HOG) features and a Support Vector Machine (SVM) classifier. Kim et al. [4] proposed a fuzzy inference-based safety assessment system for monitoring struck-by accidents on construction sites. The system detected workers and equipment using a Gaussian mixture model, and tracked the entities using a Kalman filter algorithm. Kim and Chi [24] presented a construction equipment tracking method consisting of functional integration of a detector and a tracker and online learning where training data are automatically collected. They attempted to solve the occlusion problem of construction equipment under its dynamic movements and various site conditions. Golparvar-Fard et al. [23] presented an equipment action recognition method using spatiotemporal features and SVM classifiers. Their method used a real-world video dataset of an earthmoving process for construction activity analysis.

Some other studies have analyzed construction site images with three or more recognizable objects [37–41]. Brilakis et al. [37] identified the most effective 2D tracker among point-based, contour-based, and kernel-based methods for 12 construction resources including vehicles, materials, and workers. Dimitrov and Golparvar-Fard [38] and Han and Golparvar-Fard [40] classified construction materials in over 20 different categories. Son et al. [39] detected concrete, steel, and wood in construction environments using ensemble classifiers. Hamledari et al. [41] identified the components of an interior partition such as studs, electrical outlets, insulation, and three states for drywall sheets (installed, plastered, and painted) in indoor construction site images.

A few studies have attempted to recognize all identifiable objects in construction site images [42,43]. Kim et al. [42] proposed a global recognition system for identifying objects in the whole area of images using a nonparametric scene-parsing method. This system can recognize the background such as grounds, grasses, trees, and skies, as well as construction resources such as workers, vehicles, materials, and structures in the images. Yang [43] suggested a data-driven, scene-parsing method for action recognition of construction workers. They showed that the semantic information of all recognizable objects can improve the action recognition performance.

Attempts have been made to use a convolutional neural network (CNN) for detecting construction resources. Fang et al. [2] developed a vision-based system using a monocular camera to determine whether construction workers are wearing harnesses. Two CNNs were used in the system; a faster region-based convolutional neural network (Faster R-CNN) for detecting workers, and a CNN based on the research of Krizhevsky et al. [44] for identifying their harnesses. The precision and recall of the Faster R-CNN were 99% and 95%, and those of the CNN model for harness detection were 80% and 98%, respectively. Kim et al. [12] proposed an integrated method of a construction process simulation, and vision-based monitoring using a region-based, fully convolutional network (R-FCN) and CCTV video. The method detected excavators and dump trucks, and analyzed the productivity of an earthmoving process in a tunnel. The detection model for two equipment exhibited high performance with a mean average precision of 99.09%. Kim et al. [45] proposed a vision-based method to detect five types of construction equipment using the R-FCN, transfer learning, and construction equipment images in the ImageNet dataset. In their study, transfer learning was effectively used to overcome the limitation of the small amount of training data. The R-FCN trained using transfer learning achieved 96.33% mean average precision. Fang et al. [3] proposed an improved Faster R-CNN to detect construction workers and excavators in real-time using surveillance videos. They adjusted some parameters of the Faster R-CNN, which are the bounding boxes' scale and aspect ratio, for detecting small-size workers. The precision and recall of the improved Faster R-CNN were both above 90%. Son et al. [46] developed a CNN-based model for detecting construction workers, robust to their pose and background change in images. The model with a deep residual network and a Faster R-CNN achieves a detection performance with precision and recall of 94.3% and 96.03%, respectively. Kim et al. [47] proposed a method for action recognition of excavators with sequential pattern analysis. With a CNN and a double-layer long short-term memory network, the sequential pattern of excavator action was recognized with an accuracy of 93.8%. Despite these successes, the existing methods faced difficulty in analyzing construction site images acquired by UAV owing to their unique characteristics. Such analysis requires a satisfactory amount of training data to reflect the diverse conditions of construction sites and UAV flights. Thus, a methodology is needed to obtain a sufficient number of construction site images acquired by UAV.

2.2. Image augmentation methods for improving the performance of an object detection model

Image augmentation methods that artificially create additional training data are used to improve the generalization capability of machine learning models [36,48–51]. This method is valid for CNNs with a large number of parameters. Krizhevsky et al. [44] used image augmentation techniques, such as patch extracting and image intensity altering, for reducing the overfitting problem on their CNN. Wong et al. [51] used a data warping technique that creates new images from the original images by applying transformations such as translation, shearing, and rotation. Inoue [50] synthesized new images by overlaying two randomly selected images from a training dataset with the same class. This intuitive and simple method reduced the error rate of GoogleNet from 8.22% to 6.93% in the classification task of the Canadian institute for advanced research (CIFAR-10) dataset. When performing image augmentation for detection, it is essential to preserve label information, including the location and the identification of objects. Dwibedi et al. [49] proposed an image generation method that cuts objects from the training dataset and pastes them into background images. Annadani and Jawahar [48] proposed an image augmentation method using crop, copy, and paste of objects for CNN-based image tampering detection. These cut-and-paste techniques can create a new synthetic image, preserving the label information of the original data without modification of a detection model. Although the cut-and-paste technique has performed very well on image augmentation, there is still some room for improvement. As this technique places new objects on the images of the original dataset, the objects present in the original dataset remain in an augmented dataset. To avoid duplication of the training data, and to diversify the placement pattern of objects, a pre-processing technique is required to remove the existing objects from the original images.

GAN, a type of CNN that relies on unsupervised learning, has been recently used for image augmentation [52]. GAN has succeeded in creating image datasets such as tumor images [53], liver images [54], and human face images [55]. While the study herein attempted to create fake construction site images using a GAN, it failed to create valid construction site images that could improve the detection performance of the construction resources. Compared to images in other fields where the GAN was successfully used, construction site images contain not only various types of objects with complicated backgrounds, but are also of high resolution. Thus, instead of using a GAN to create a fake construction site image, a GAN was used to restore the removed area after removing the objects from the image as a pre-processing of the cut-and-paste technique. In this context, this study aims to develop an image augmentation method for construction resource detection using a GAN, cut-and-paste, and image transformation techniques.

3. Methodology

The method, as previously mentioned, consists of three techniques: removing-and-inpainting, cut-and-paste, and image-variation. First, the removing-and-inpainting technique diversifies the placement pattern of objects in images as a pre-processing of the cut-and-paste technique. The technique removes objects in the image using the label information of the objects, and restores the removed regions of the images considering the context with the surrounding pixels. The GAN proposed by Yu et al. [49] is used in this technique for image inpainting. Second, the cut-and-paste technique creates a new image using the objects and their label information extracted from the original training dataset. In a dataset of construction site images, the number of objects in one type of construction resource is different from that in another type of resource. The techniques of removing-and-inpainting and cut-and-paste can address the class-distribution imbalance issue of the dataset. The image-variation technique applies variations of intensity, scale, and blur to an image: the three image transformation techniques. The quality of the UAV-acquired construction site image is likely to change according to the UAV flight situation (i.e., height, speed, and vibration), and the outdoor construction environment. The various conditions are incorporated into the UAV images by the image-variation technique. Fig. 1 describes the research flow for the UAV-acquired image augmentation method. The augmented image dataset is constructed by applying the image-variation technique to the original training dataset and the synthetic image dataset, for which the removing-and-inpainting and the cut-and-paste techniques are employed. Finally, the augmented image dataset is used to train a Faster R-CNN for detecting construction resources on UAV-acquired construction site images.

3.1. Removing-and-inpainting using a GAN

A well-learned model has an excellent generalization capability to detect target objects in unfamiliar backgrounds. One way for improving the generalization capability of a detection model is to let new images have a new relationship between an object and a background not present in the original dataset. Based on this idea, the cut-and-paste technique places objects into empty spaces in an image. However, with this technique, the area where no new object is placed overlaps with that of the original images because the technique does not change the unpasted area on which new objects are not pasted. This overlap between the original dataset and the newly generated dataset can result in a bias of class-distribution. Thus, the removing-and-inpainting technique arbitrarily removes objects in the original image. This object removal process widens the candidate area where the cut-and-paste technique can place new objects, and diversifies the pattern of object placement. The area where an object is removed should be restored with ambient pixels to have visual features to prevent a detection model from misjudging them as an object.

The removing-and-inpainting technique consists of two modules, object-removing and image-inpainting. The object-removing module selects a random number of objects from the original image, and then converts the selected objects into white mask with all red, green, and blue (RGB) values of 255 by utilizing the coordinates of the polygon in their label information, as shown in Fig. 2. The number of removed objects is set to any integer number between 1 and the total number of objects in the image. Using the GAN [49], the image-inpainting module restores the pixels converted via the object-removing module to be similar to the surrounding background. Unlike a typical GAN, which trains real-world images to create fake images, the GAN used in this technique is designed to find a vector close to the original of the damaged image. The GAN trained by undamaged images is input to the damaged image generated by the object-removing module. Then, the GAN generates the output image to be most similar to the damaged image, but fills the damaged region of the output image such that the resultant image can show a realistic UAV-acquired construction scene. In this module, a pre-trained GAN with a Place2 dataset [56] consisting of natural scene images was utilized. This is because the goal of this study is similar to the aim of the Place2 dataset, which is restoring the damaged natural scene.

3.2. Cut-and-paste

Each object class requires a sufficient amount of training data to perform an object detection task successfully. However, datasets consisting of construction site images have a non-uniform distribution in the number of objects per class. In the case of a certain piece of equipment larger than other construction resources, only one or even none may exist in an image, whereas dozens of materials of relatively small sizes can exist in an image. As the size of the dataset becomes larger, the class-distribution imbalance of these datasets is intensified. To address the imbalance, the cut-and-paste technique should be able to generate images for the under-represented types of construction resources; the generated images should contain a large number of objects for the particular resource classes with relatively small amounts of training data. Based on this idea, the cut-and-paste technique is developed to create realistic construction site images while maintaining their context. This technique is input to the image generated by the removing-and-inpainting technique.

Fig. 3 shows the procedure of the cut-and-paste technique. The object extraction module extracts a portion of the image containing the object using its label information from the original image dataset. The mask generation module segments the extracted object at the pixel level, and creates a binary mask with a foreground value of 1 and a background value of 0. The random padding module places the mask created from the mask generation module at a position chosen randomly within a binary image; the binary image is the blank image that has the same size as the original image. A background image (I_{back}) is randomly selected from the original image dataset, and converted to another mask image ($Mask_{back}$) again with a foreground value of 1 and a background value of 0; the masks with the value of 1 represent the labeled objects in the selected background image. Then, the occlusion checking module determines whether there is occlusion between the background mask ($Mask_{back}$) and the object mask ($Mask_{obj}$) generated by the random padding module. If the occlusion occurs between the two images, the random padding module creates a new binary mask repeatedly until the occlusion does not occur. When there is no occlusion, the module puts the object mask ($Mask_{obj}$) on the background mask

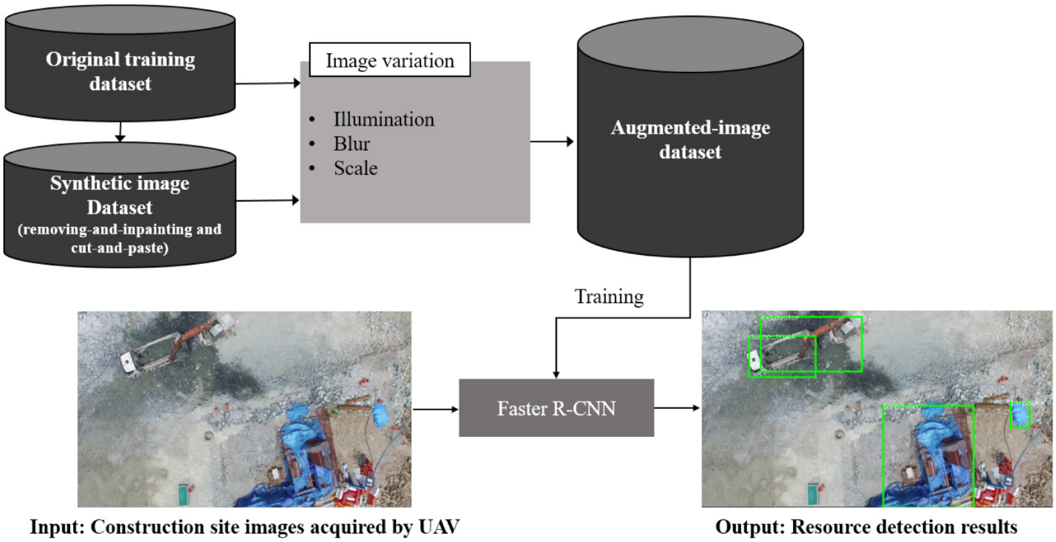

Original training dataset
Synthetic image Dataset (removing-and-inpainting and cut-and-paste)
Image variation
• Illumination
• Blur
• Scale
Augmented-image dataset
Training
Faster R-CNN
Input: Construction site images acquired by UAV
Output: Resource detection results

Fig. 1. Overview of the UAV-acquired image augmentation method.

($Mask_{back}$), creating a new combined mask image; here, the combined mask image becomes a new background mask image ($Mask_{back}$) that may be compared with a newly generated object mask image ($Mask_{obj}$) for their occlusion checking. For those objects that passed the occlusion checking module, the foreground replacing module replaces all foreground pixels of the object mask ($Mask_{obj}$) with the corresponding pixel values of the original image. Finally, the image-blending module combines the image (I_{obj}) generated in the foreground-replacing module with the image (I_{back}) selected from the background image selection module using Eq. (1).

$$I_{new} = \{Mask_{obj} * I_{obj}\} + \{(1 - Mask_{obj}) * I_{back}\} \tag{1}$$

where $*$ denotes pixel-wise multiplication. The image (I_{new}) created from the image blending module is stored in a synthetic image dataset as the final output of the technique. Fig. 2 shows an example of an image from the original dataset, and the resultant images of applying the original image to the object-removing module, the image-inpainting module, and the cut-and-paste technique, respectively.

The random padding module and the occlusion-checking module are performed according to a few conditions for realistic object placement, and the resolution of class-distribution imbalance. The detailed procedure of the two modules is as follows (Fig. 4):

(1) Initialize n as 1.

(2) Randomly set the number of objects placed per one image (N) to any integer number between 3 and 9.

(3) Calculate the probability of being placed on the background image per class using Eq. (2).

$$p_i = {}^{c_i}\!\big/\!\textstyle\sum_i^M c_i \left(where\ c_i = {}^{1}\!\big/_{n_i} \right) \tag{2}$$

where p_i is the probability that an object belonging to the ith class will be placed on background image, M is the total number of classes in the original dataset, and n_i is the total number of objects belonging to the ith class in the original training dataset. Eq. (2) is designed to have a high probability for an object with a small number of training images in the original training dataset.

(4) Place the selected object based on the probability (p_i) in step (3) in the image (I_{back}). To determine the appropriate size of the object, the image data of the original dataset are analyzed in advance. Using this process, the average size ratios of different classes are estimated and used for the size determination.

(5) Perform the occlusion checking using the object mask ($Mask_{obj}$) and the background mask ($Mask_{back}$).

(6) When the object mask ($Mask_{obj}$) passes the occlusion checking in step (5), increment n by 1 to choose the next object in the image until n reaches N.

3.3. Image-variation

The selected dataset should reflect characteristics of an image taken on construction sites exposed to a unique environment, as well as an image taken by a continuously moving UAV. The ideal data with multiple patterns of construction sites and UAV flights would enable a detection model to avoid overfitting, and improve its generalization capability. However, it is difficult to get a sufficient amount of diverse image data due to the labor-intensive task of labeling, and the lack of publicly available datasets in the construction industry. Thus, the image-variation technique constructs a new dataset by changing the values of three variables (illumination, blur, and scale) of the images. As aforementioned, the selection of the three variables and image transformation techniques is based on the characteristics of the data, and leads to the formation of three processing modules: intensity variation, image smoothing, and scale transformation. The three modules of the image-variation technique are separately applied to the dataset.

3.3.1. Intensity variation

The first variable of the dataset used is illumination because of the ambient light at construction sites. The intensity of the images depends on the time of day, weather condition, and the shadows generated by the structures. For example, the images taken at construction sites with no structures on a sunny day are likely to have a relatively high level of intensity. The images taken at sites with structures, on a cloudy day, or at night are likely to have a relatively low level of intensity. To consider the illumination changes, the intensity variation module creates a new image by increasing or decreasing the intensity values of all the pixels in the original image. As this module does not change the position of the objects in an image, the label information for the new images is the same as that for the original image.

3.3.2. Image smoothing

The second variable is the blur caused by the vibration of UAV, and the continuous movement of some construction resources. The wind along with an inexperienced UAV pilot lacking requisite control is likely to generate blurred images. The movement of a construction worker and equipment is likely to generate blurry regions on an image. To consider the blur, the image-smoothing module artificially creates blurry images by applying a Gaussian filter to the original image. This

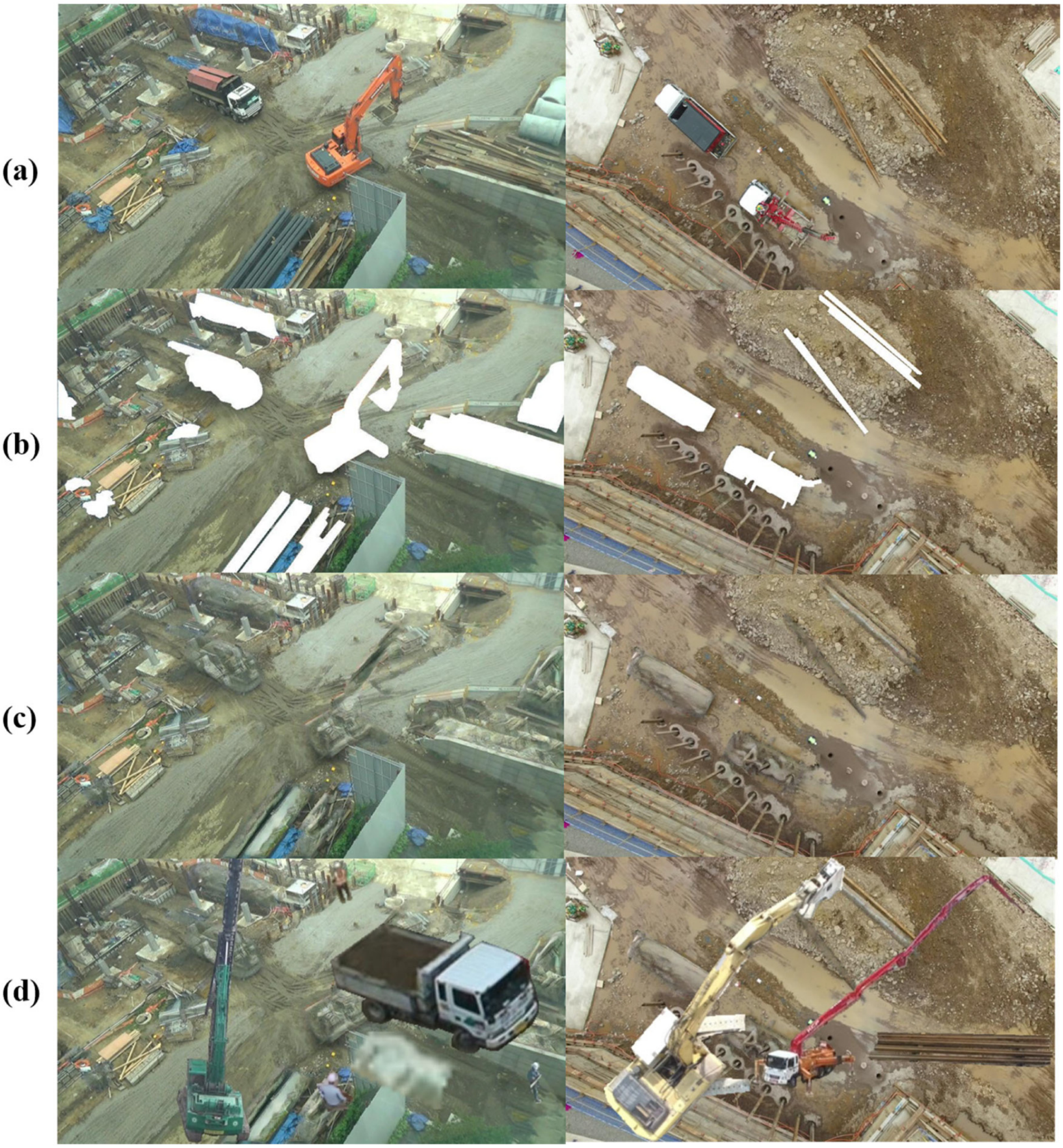


Fig. 2. Comparison among (a) original image; (b) resultant images after the object-removing module; (c) resultant images after the image-inpainting module; and (d) resultant images after the removing-and-inpainting and the cut-and-paste techniques.

module also retains the label information from the original image for the new image.

3.3.3. Scale transformation

The third variable, scale, depends on the flight altitude of UAV when it captures an image. The same object is recognized on a different scale depending on the distance between the camera attached to the UAV and the construction site. To consider the scale changes, the scale transformation module creates new images by applying the affine scale transformation, a linear mapping method. Eq. (3) shows the matrix operation of the affine scale transformation.

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} scale & 0 & 0 \\ 0 & scale & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{3}$$

where x' and y' denote the x- and y-coordinates of a new image, respectively, *scale* denotes a scale factor, and x and y denote the x- and y-coordinates of the original image, respectively. A new image is created by enlarging or reducing the original image by *scale* times. As a new image should have the same resolution as the original image for training a model, zero-value pixels are padded on the void space in the reduced image when *scale* is less than 1. On the contrary, some visual information of the original image is lost when *scale* is larger than 1. The center point of a new image generated using Eq. (3) is not the same as that of the original image. Thus, this module transforms the image using Eq. (4).

$$x' = x \times scale + W \times \frac{(1 - scale)}{2} y' = y \times scale + H \times (1 - scale)/2 \tag{4}$$

where W and H denote the width and height of the original image, respectively. Unlike the intensity variation and image smoothing modules, the scale transformation module converts the position of objects within an image. This module should generate the new label representing the object position of the new image. However, the coordinates of some objects can be out of the resolution of the original image resulting in loss of some visual information, when *scale* is larger than 1. Using images containing objects with lost visual information as

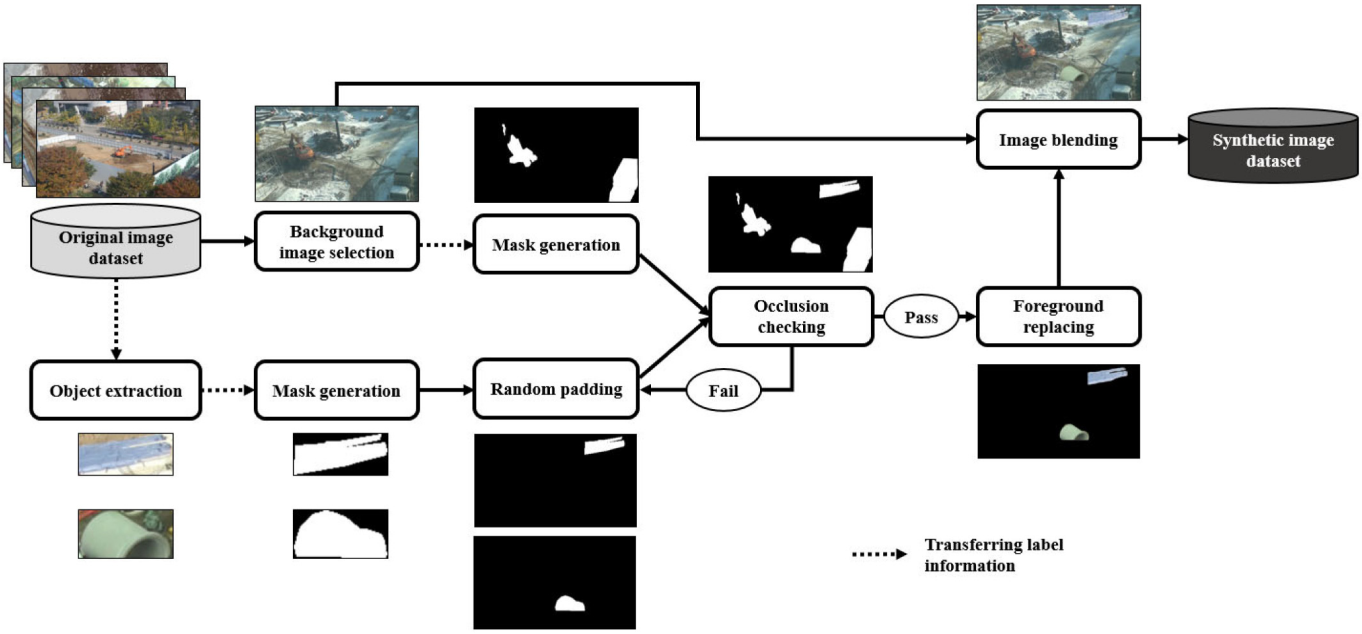


Fig. 3. Procedure of the cut-and-paste technique.

learning data may impair the ability of deep learning models to recognize construction resources. Thus, the scale transformation module deletes the labels of the objects with lost visual information. Fig. 5 shows images generated with a scale factor of 0.8, 1.0, and 1.2. Fig. 5(a) shows the black rectangular frame, which represents zero-value pixels resulting from the reduction in the image size. Fig. 5(c) and (d) illustrate that the visual data corresponding to the outer part of Fig. 5(b) is lost through the scale transformation.

4. Experimental study

4.1. Experimental environment and dataset

The proposed method was implemented using Python, on the Ubuntu 16.04 operating system with a GTX 1080 Ti GPU, and an Intel Core i7-7700 processor. Tensorflow, a deep learning library, was used to train and test a Faster R-CNN. The UAV used to acquire construction site images was a Phantom 3 Professional from the DJI Corporation. The dataset consisted of 656 images collected on six different construction sites (dam, apartment, industrial complex, road, sport complex, and subway station). The images had a resolution of 3840 by 2160; however, they were adjusted to 960 by 540 to reduce the computational burden on the operating system. The images in the dataset were manually annotated using the VGG Image Annotator [57].

4.2. Parameter settings

The removing-and-inpainting and cut-and-paste techniques created 2000 new images based on the original training dataset. Table 1 shows the number of objects per class in the original training dataset, and the probability of being placed in the new image for each class (p_i), calculated using Eq. (2). The original training dataset was unbalanced in terms of the number of objects per class. For example, the number of workers is about 26 times that of concrete trucks. Table 1 also shows the number of objects in the synthetic image dataset, and the ratio of the number of objects in the synthetic image dataset with removing-and-inpainting and cut-and-paste techniques to original training dataset. Higher ratios were obtained for the classes with fewer objects in the original dataset. The gap between the number of workers and concrete trucks was reduced to less than three times. When a class was selected based on the probability (p_i), the size ratios of the different classes were calculated to determine the appropriate size of the object being placed in the new image. Table 2 shows the relative sizes of construction resource classes, assuming the size of the dump truck class as 1.

Table 3 shows the parameters of the three modules of the image-variation technique. As the three modules were applied separately, the image-variation technique using the parameters in Table 3 produced 30 times more images than the original images. As shown in Fig. 1, the image-variation technique was applied to both the original image dataset with 544 images, and the synthetic image dataset with 2000 images. Thus, the combined method of removing-and-inpainting, cut-and-paste, and image-variation increased the size of the dataset from 544 to 76,320 (2544 times 30). As mentioned earlier, Faster R-CNN was used as the construction resource detection model. The learning parameters, learning rate, and weight decay were set to 0.9, 0.003, and 0.0005, respectively. The number of iterations in the training process was set to 300,000.

4.3. Evaluation

To validate the proposed method, Faster R-CNN was trained with four different datasets: the original training dataset, the dataset with cut-and-paste technique, the dataset with removing-and-inpainting and cut-and-paste techniques, and the augmented image dataset with all the techniques. Of the 656 images of the original dataset, 544 images were used as the original training dataset, and the remaining 112 images as the original testing dataset. The four trained networks, corresponding to the four different datasets, were tested on the original testing dataset. The hyper-parameters in Section 4.2 were equally applied to all the experiments. Table 4 shows the comparison of detection performance on construction resources for Faster R-CNNs trained with the four different datasets. The predicted results were measured using two statistical indicators: recall and precision. The criterion for the prediction success was determined by Intersection of union (IoU), which indicates the extent to which the predicted bounding box matches the ground truth. If IoU was greater than 0.5, the prediction result was typically considered to be successful in detecting an object.

As shown in Table 4 and Fig. 6, the Faster R-CNN trained with the dataset using the three techniques achieved the best detection

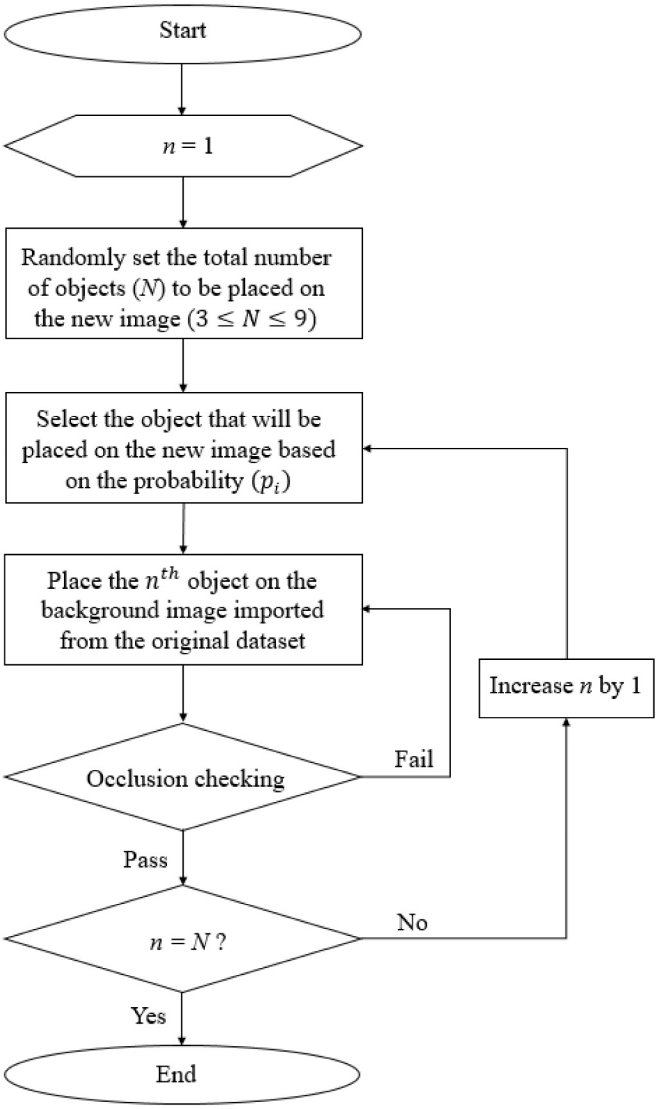

Start
$n = 1$
Randomly set the total number of objects (N) to be placed on the new image ($3 \leq N \leq 9$)
Select the object that will be placed on the new image based on the probability (p_i)
Place the n^{th} object on the background image imported from the original dataset
Increase n by 1
Occlusion checking
Fail
Pass
$n = N$?
No
Yes
End

Fig. 4. Flowchart of the random padding and occlusion checking modules.

performance. The detection result of the Faster R-CNN demonstrates that the three techniques are effective in improving the performance of Faster R-CNN by achieving recall and precision of 60.22% and 57.13%, respectively. Table 4 indicates that the cut-and-paste technique improve precision, while the removing-and-inpainting and the image-variation improves recall, respectively. In object detection tasks, recall is defined as the number of correctly predicted bounding boxes among the number of actual objects, and precision is defined as the number of correctly predicted bounding boxes among the number of predicted bounding boxes. A high precision can be obtained by a more conservative analysis of object pattern; only when a high standard of criteria is satisfied, the network should accept the object as belonging to the particular class. On the contrary, a high recall can be obtained when the network finds as many candidate regions as possible; the network should be trained by using data containing multiple patterns that the object classes can have.

5. Discussion

The cut-and-paste technique uses the objects in the original training dataset without their modification such that the bounding boxes containing the objects have new backgrounds. The new bounding boxes allow the network to learn the core features of the object class effectively because the objects in the bounding boxes are the same as the original objects. The different backgrounds in the bounding boxes become the medium through which the network can better identify the critical features of the object class. According to the result in Table 4, the cut-and-paste technique improves precision by 9.14%. However, the technique reduces recall by 2.26%. When using only cut-and-paste technique, the augmented dataset overlaps with the original dataset in the areas where the technique is not applied. Data bias due to duplication in the image area can reduce the recall because it prevents the network from learning different patterns. The use of removing-and-inpainting technique supplements the shortcomings of the cut-and-paste technique by letting the augmented dataset have new patterns; the technique allows new objects to be placed in the image area that was an object in the original dataset. Table 4 shows that the results of using the removing-and-inpainting technique to the original dataset increase the recall by 7.33% without reducing the precision compared to that using only the cut-and-paste technique. Fig. 6(d) shows an example of the prediction result of the network trained with the dataset using the removing-and-inpainting and the cut-and-paste techniques. Compared to the result shown in Fig. 6(c), the results shown in Fig. 6(d) indicate that the faster R-CNN trained by the datasets with the removing-and-inpainting technique detects more actual objects in the same scene.

The image-variation technique changes the objects in the original training dataset in terms of illumination, blur, and scale. The network trained with the image-variation dataset is likely to find more candidate objects. This technique leads to the improvement of the ability to detect objects with the pattern that does not exist in original datasets. Fig. 6(e) shows an example of the prediction result of the network trained with the dataset using all the techniques of the proposed method. Although there are eight predicted bounding boxes, there exist only five actual objects. It is also worth noting that all of the five objects shown in Fig. 6(a) were successfully detected by the eight bounding boxes. Compared to the result shown in Fig. 6(b), the results shown in Fig. 6(e) and Table 4 demonstrate that the three techniques complement each other, and can achieve the best performance despite the shortcomings of each technique.

Table 5 shows the comparison of detection performance per class with and without the proposed method. Both indicators (recall and precision) were improved by around 20% in most classes. Even crane trucks that were not detected by the network trained with the original dataset could be detected with the proposed method. This is because the visual features of crane trucks, such as the color, shape, texture, patterns, and contexts with surroundings, are quite different between the original training and test dataset. It can be inferred that the proposed method generates some images containing crane trucks with visual features that are closer to those in the test dataset.

Referring to Table 4, the number of training data and the performance of the detection model appear to be proportional. However, the performance of detection models trained with large amounts of training data is not necessarily high. For example, if a construction site image containing a dump truck with similar visual characteristics to an excavator is generated by a data augmentation method, the ability for the detection model to identify the excavator will be reduced. Thus, data augmentation methods need to generate valid data that can improve the performance of the model. Without considering the nature of the construction resources or the conditions at construction sites, data augmentation methods may not achieve performance gain for a detection model.

The purpose of this study is to develop a data augmentation method for improving the model that detects 10 types of construction resources, including construction workers, four kinds of materials, and five types of equipment. To the best of the authors' knowledge, this study is one of the first attempts to detect as many as 10 types of construction resources. This study focuses on improving the generalization capability of models to detect multiple construction resources on various construction sites. The proposed method is expected to be used to detect

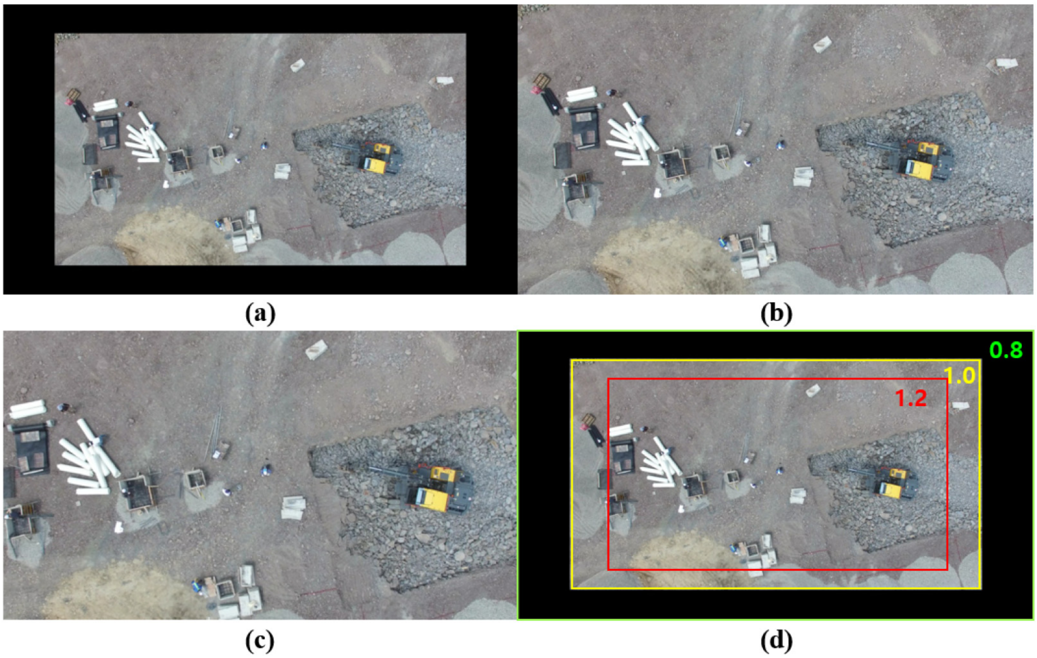


Fig. 5. Examples of the scale transformation: (a) scale 0.8; (b) scale 1.0; (c) scale 1.2; (d) comparison of (a), (b), and (c).

Table 1
The change of number of objects per class using the removing-and-inpainting and the cut-and paste technique.

Class	Number of objects in the original training dataset	Number of objects in the synthetic image dataset	Probability of being placed in the new image (p_i)	Ratio of the synthetic image dataset to the original training dataset
Worker	1226	2749	1.29%	2.24
Tarpaulin	735	1971	2.15%	2.68
Rebar	791	1429	2.00%	1.80
H-beam	435	836	3.63%	1.92
Concrete pipe	626	1737	2.52%	2.77
Drilling	72	586	21.95%	8.14
Crane truck	74	569	21.36%	7.69
Excavator	378	1081	4.18%	2.86
Concrete truck	48	518	32.93%	10.79
Dump truck	198	755	7.98%	3.81
Total	4603	13,168	100.00%	

Table 2
The relative sizes of classes assuming the size of the dump truck class as 1.

Class	Relative Size
Worker	0.17
Tarpaulin	0.86
Rebar	0.89
H-beam	1.04
Concrete pipe	0.65
Drilling	1.88
Crane truck	2.03
Excavator	1.15
Concrete truck	1.07
Dump truck	1.00

Table 3
The parameters of the three modules of the image-variation technique.

Module	Parameters
Intensity variation	(Adding value) −60, −30, 0, 30, 60
Image smoothing	(σ value of Gaussian filter) 0, 1.0
Scale transformation	(Scale factor) 0.8, 1.0, 1.2

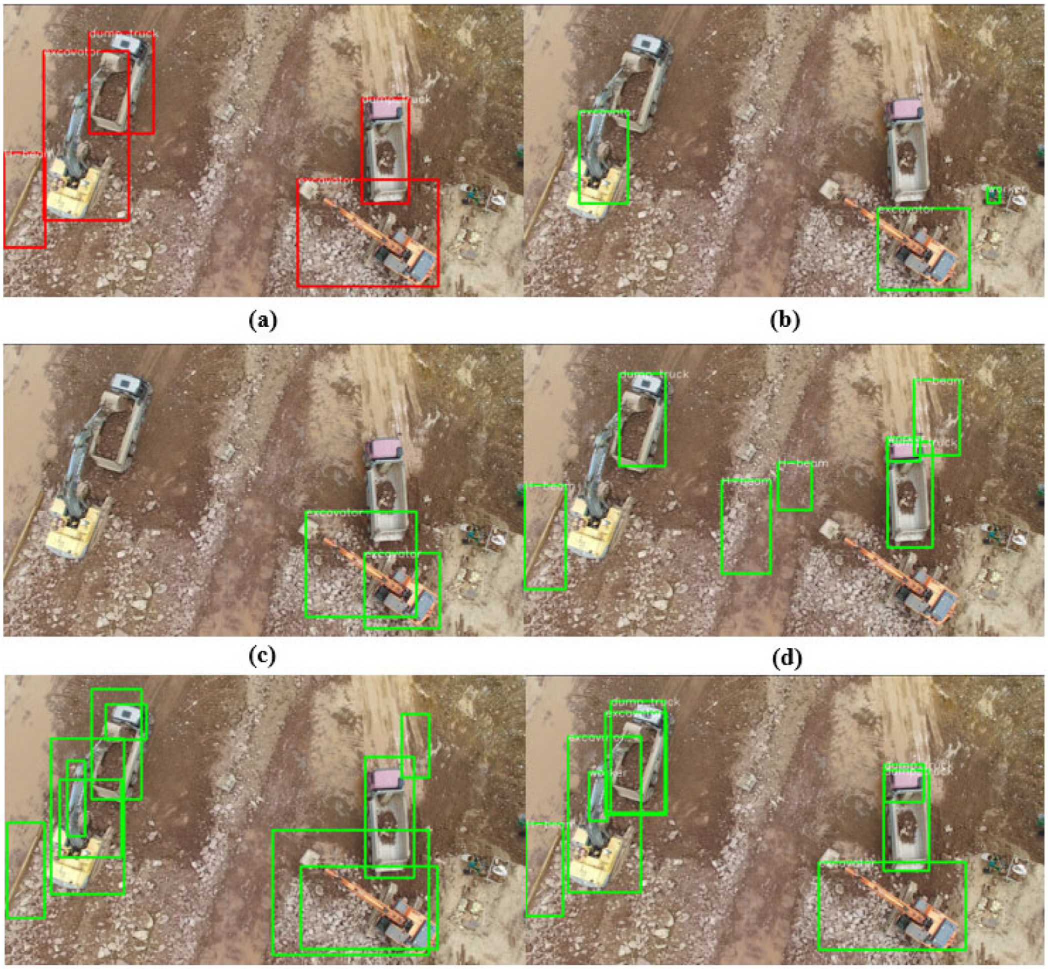
construction resources in a range of construction site images.

6. Limitation and suggestions

The results in Table 5 imply the limitations of the proposed method. Firstly, the cut-and-paste technique places construction resources in empty spaces of images without considering the actual terrain and conditions of construction sites. For example, in the image generated by the technique, construction equipment may not be attached to the ground, or workers can be placed on impassable roads. This technique can create images with object placement that does not exist at actual construction sites, leading to the degradation of detection performance. Thus, the performances of some classes, such as concrete pipe and concrete trucks, decreased in one of the indicators. This result suggests that more images with unrealistically placed construction resources were generated for the classes that yielded decreased performance. Because the object placement in the cut-and-paste technique and the class distribution of the training and test datasets have randomness, gains in detection performance by the proposed method may be different according to object class. Besides, the method creates image data using the variation in illumination, scale, and blur of construction sites but does not generate image data with the variation in color or shape of

Table 4
Detection performance on construction resources by the proposed techniques.

Technique	The number of training images	Recall	Precision	F-measure
Without the proposed techniques	544	35.96%	45.62%	40.22%
Cut-and-paste	2544	33.70%	54.76%	41.72%
Removing-and-inpainting and cut-and-paste	2544	41.03%	55.72%	47.26%
Image-variation	16,320	46.92%	54.27%	50.33%
All techniques	76,320	60.22%	57.13%	58.63%



(a) (b) (c) (d) (e) (f)

Fig. 6. Comparison of the results using the networks trained with different datasets: (a) ground-truth; (b) the original training dataset; (c) the dataset using the cut-and-paste technique; (d) the dataset using the removing-and-inpainting and the cut-and-paste techniques; (e) the dataset using the image-variation technique; and (f) the dataset using all the techniques in the proposed method.

construction resources. If construction site images include construction resources with color or shape, which do not exist in the training dataset, it is difficult for the network trained with the proposed method to detect them from those images. Finally, the number of training images is insufficient, Although the combination can improve the performance of the Faster R-CNN by 24.26% and 11.51% of recall and precision, respectively, the dataset used in this study does not reflect all the patterns of variables at construction sites (e.g., time of day, weather, installation status of the structure, shadows, movement of construction resources, terrain, etc.) The data augmentation method can be improved with more training data acquired at various actual construction sites.

In this context, there are two ways to improve the proposed method. First, Big data is still required for the construction industry. As construction sites have unique characteristics, big data reflecting various situations can contribute to the training of deep learning models. In fact, the proposed method already relied on a large amount of data because the method created over 70,000 images. However, the proposed method, coupled with a larger amount of real construction images, would allow the network to achieve even higher performance. Second, new algorithms are required for object-level data augmentation. The proposed method performs data augmentation using existing objects without creating new objects. If the new algorithms that can simultaneously create construction resource images and their labels are developed, they can overcome the limitations of the proposed method.

7. Conclusion

This paper proposes an image augmentation method to improve the performance of a detection model for construction resources detection in UAV-acquired images. The proposed method consists of three

Table 5
Comparison of detection performance per class with and without the proposed method.

	Without the method		With the method	
	Recall	Precision	Recall	Precision
Worker	24.75%	24.39%	51.49%	41.43%
Tarpaulin	76.58%	62.50%	80.18%	57.79%
Rebar	20.75%	41.51%	50.94%	49.09%
H-beam	7.41%	36.36%	23.15%	59.52%
Concrete pipe	54.79%	80.00%	69.86%	58.62%
Drilling	22.22%	33.33%	44.44%	44.44%
Crane truck	0.00%	0.00%	30.77%	33.33%
Excavator	52.17%	42.86%	80.00%	70.23%
Concrete truck	20.00%	11.11%	10.00%	50.00%
Dump truck	13.85%	90.00%	53.85%	60.34%
Average	35.96%	45.62%	60.22%	57.13%

techniques. First, the removing-and-inpainting technique is proposed to diversify the pattern of object placement using the GAN as the preprocessing of cut-and-paste technique. Second, the cut-and-paste is proposed to solve the class-distribution imbalance of the dataset considering the size and number of construction resource in the images. Third, the image-variation technique is proposed to consider the characteristics of the construction site images taken by UAV using three image transformation techniques. Faster R-CNN, a CNN with promising results in object detection, was used to validate the proposed method. The 656 UAV images were acquired at six different construction sites. Of the 656 images, 544 were used as the original training dataset, and the remaining 112 were used as the experimental dataset to test the network. The network trained with 76,320 training images using the three techniques together presented better performance than the learned network without the method, with recall of 60.22% and precision of 57.13%, respectively. Experimental results show that the cut-and-paste techniques can improve the precision of the network while the removing-and-inpainting and the image-variation technique can improve the recall.

There are two major contributions of this study. First, the authors provide guidelines for developing a data augmentation method consisting of three computer vision technologies, GAN, cut-and-paste, and image transformation techniques, to improve the performance of object detection models for construction resources. The proposed method uses the concept of object removing, GAN-based image inpainting, and the probability and relative size of objects classes to generate fake construction images appropriately. It is difficult to ensure that the proposed method maximizes the performance for construction resource detection. However, to the best of our knowledge, there is no study yet to generate fake construction site images for improving the performance of object recognition models. Thus, the proposed method can be used as a benchmark for future visual data augmentation methods for construction site images. Second, the authors endeavor to create a better understanding of how data augmentation methods affect the precision and recall values. The use of the same objects in different image backgrounds was found to improve the precision value in construction resources detection. On the contrary, the use of variations in intensity, blur, and scale was found to improve the recall value. These findings are expected to serve as guidance to future efforts in construction image augmentation.

The proposed method succeeded in improving the detection performances of construction resources in most classes. However, the detection performances in some classes (e.g., concrete pipe) were degraded by this method. Furthermore, the average detection performance (about 60%) for 10 types of construction resources shows that the proposed method is not reliable enough to be utilized at actual construction sites. This result came from the fact that not only did the proposed method not take into account the conditions and terrain of the construction site, it also did not change the color and shape of the construction resources. To improve the result of the proposed method, algorithms should be developed that can consider the unique characteristics of the construction sites and diversify the visual features of the construction resources.

Acknowledgments

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. Chi, C.H. Caldas, Image-based safety assessment: automated spatial safety risk identification of earthmoving and surface mining activities, J. Constr. Eng. Manag. 138 (3) (2011) 341–351, https://doi.org/10.1061/(ASCE)CO.1943-7862.0000438.

[2] W. Fang, L. Ding, H. Luo, P.E. Love, Falls from heights: a computer vision-based approach for safety harness detection, Autom. Constr. 91 (2018) 53–61, https://doi.org/10.1016/j.autcon.2018.02.018.

[3] W. Fang, L. Ding, B. Zhong, P.E. Love, H. Luo, Automated detection of workers and heavy equipment on construction sites: a convolutional neural network approach, Adv. Eng. Inform. 37 (2018) 139–149, https://doi.org/10.1016/j.aei.2018.05.003.

[4] H. Kim, K. Kim, H. Kim, Vision-based object-centric safety assessment using fuzzy inference: monitoring struck-by accidents with moving objects, J. Comput. Civ. Eng. 30 (4) (2015) 04015075, , https://doi.org/10.1061/(ASCE)CP.1943-5487.0000562.

[5] K. Kim, H. Kim, H. Kim, Image-based construction hazard avoidance system using augmented reality in wearable device, Autom. Constr. 83 (2017) 390–403, https://doi.org/10.1016/j.autcon.2017.06.014.

[6] M.-W. Park, N. Elsafty, Z. Zhu, Hardhat-wearing detection for enhancing on-site safety of construction workers, J. Constr. Eng. Manag. 141 (9) (2015) 04015024, , https://doi.org/10.1061/(ASCE)CO.1943-7862.0000974.

[7] J. Seo, S. Han, S. Lee, H. Kim, Computer vision techniques for construction safety and health monitoring, Adv. Eng. Inform. 29 (2) (2015) 239–251, https://doi.org/10.1016/j.aei.2015.02.001.

[8] P. Tang, G. Chen, Z. Shen, R. Ganapathy, A spatial-context-based approach for automated spatial change analysis of piece-wise linear building elements, Computer-Aided Civil and Infrastructure Engineering 31 (1) (2016) 65–80, https://doi.org/10.1111/mice.12174.

[9] M. Bügler, A. Borrmann, G. Ogunmakin, P.A. Vela, J. Teizer, Fusion of photogrammetry and video analysis for productivity assessment of earthwork processes, Computer-Aided Civil and Infrastructure Engineering 32 (2) (2017) 107–123, https://doi.org/10.1111/mice.12235.

[10] M. Golparvar-Fard, F. Peña-Mora, S. Savarese, Automated progress monitoring using unordered daily construction photographs and IFC-based building information models, J. Comput. Civ. Eng. 29 (1) (2012) 04014025, , https://doi.org/10.1061/(ASCE)CP.1943-5487.0000205.

[11] J. Gong, C.H. Caldas, Computer vision-based video interpretation model for automated productivity analysis of construction operations, J. Comput. Civ. Eng. 24 (3) (2009) 252–263, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000027.

[12] H. Kim, S. Bang, H. Jeong, Y. Ham, H. Kim, Analyzing context and productivity of tunnel earthmoving processes using imaging and simulation, Autom. Constr. 92 (2018) 188–198, https://doi.org/10.1016/j.autcon.2018.04.002.

[13] J. Zou, H. Kim, Using hue, saturation, and value color space for hydraulic excavator idle time analysis, J. Comput. Civ. Eng. 21 (4) (2007) 238–246, https://doi.org/10.1061/(ASCE)0887-3801(2007)21:4(238).

[14] M. Ahmed, C. Haas, R. Haas, Using digital photogrammetry for pipe-works progress tracking, Can. J. Civ. Eng. 39 (9) (2012) 1062–1071, https://doi.org/10.1139/l2012-055.

[15] M. Golparvar-Fard, J. Bohn, J. Teizer, S. Savarese, F. Peña-Mora, Evaluation of image-based modeling and laser scanning accuracy for emerging automated performance monitoring techniques, Autom. Constr. 20 (8) (2011) 1143–1155, https://doi.org/10.1016/j.autcon.2011.04.016.

[16] K. Han, J. Lin, M. Golparvar-Fard, A formalism for utilization of autonomous vision-based systems and integrated project models for construction progress monitoring, Conference on Autonomous and Robotic Construction of Infrastructure, Ames, IA, USA, 2015, pp. 119–131 https://lib.dr.iastate.edu/intrans_reports/141/.

[17] C. Kim, B. Kim, H. Kim, 4D CAD model updating using image processing-based construction progress monitoring, Autom. Constr. 35 (2013) 44–52, https://doi.org/10.1016/j.autcon.2013.03.005.

[18] L. Klein, N. Li, B. Becerik-Gerber, Imaged-based verification of as-built documentation of operational buildings, Autom. Constr. 21 (2012) 161–171 pp. 44-52 https://doi.org/10.1016/j.autcon.2011.05.023.

[19] Y. Turkan, F. Bosché, C. T. Haas, R. Haas, Tracking of secondary and temporary objects in structural concrete work, Constr. Innov. 14 (2) (2014) 145–167, https://doi.org/10.1108/CI-12-2012-0063.

[20] S. Chi, C.H. Caldas, Automated object identification using optical video cameras on construction sites, Computer-Aided Civil and Infrastructure Engineering 26 (5) (2011) 368–380, https://doi.org/10.1111/j.1467-8667.2010.00690.x.

[21] S. Chi, C.H. Caldas, D.Y. Kim, A methodology for object identification and tracking in construction based on spatial modeling and image matching techniques, Computer-Aided Civil and Infrastructure Engineering 24 (3) (2009) 199–211, https://doi.org/10.1111/j.1467-8667.2008.00580.x.

[22] X. Yang, H. Li, T. Huang, X. Zhai, F. Wang, C. Wang, Computer-aided optimization of surveillance cameras placement on construction sites, Computer-Aided Civil and Infrastructure Engineering 33 (12) (2018) 1110–1126, https://doi.org/10.1111/mice.12385.

[23] M. Golparvar-Fard, A. Heydarian, J.C. Niebles, Vision-based action recognition of earthmoving equipment using spatio-temporal features and support vector machine classifiers, Adv. Eng. Inform. 27 (4) (2013) 652–663, https://doi.org/10.1016/j.aei.2013.09.001.

[24] J. Kim, S. Chi, Adaptive detector and tracker on construction sites using functional integration and online learning, J. Comput. Civ. Eng. 31 (5) (2017) 04017026, , https://doi.org/10.1061/(ASCE)CP.1943-5487.0000677.

[25] 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, Int. J. Comput. Vis. 115 (3) (2015) 211–252, https://doi.org/10.1007/s11263-015-0816-y.

[26] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, Microsoft COCO: Common objects in context, Proceedings of the 2014 European Conference on Computer Vision (ECCV), Zurich, CH, 2014, pp. 740–755, , https://doi.org/10.1007/978-3-319-10602-1_48.

[27] A. Kuznetsova, H. Rom, N. Alldrin, J. Uijlings, I. Krasin, J. Pont-Tuset, S. Kamali, S. Popov, M. Malloci, T. Duerig, The Open Images Dataset V4: Unified image classification, object detection, and visual relationship detection at scale, arXiv preprint arXiv:1811.00982, 2018.

[28] A. Zhang, K.C. Wang, B. Li, E. Yang, X. Dai, Y. Peng, Y. Fei, Y. Liu, J.Q. Li, C. Chen, Automated pixel-level pavement crack detection on 3D asphalt surfaces using a deep-learning network, Computer-Aided Civil and Infrastructure Engineering 32 (10) (2017) 805–819, https://doi.org/10.1111/mice.12297.

[29] S. Bang, S. Park, H. Kim, H. Kim, Encoder–decoder network for pixel-level road crack detection in black-box images, Computer-Aided Civil and Infrastructure Engineering 34 (8) (2019) 1–15, https://doi.org/10.1111/mice.12440.

[30] Y.J. Cha, W. Choi, G. Suh, S. Mahmoudkhani, O. Büyüköztürk, Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types, Computer-Aided Civil and Infrastructure Engineering 33 (9) (2018) 731–747, https://doi.org/10.1111/mice.12334.

[31] Y. Ham, K.K. Han, J.J. Lin, M. Golparvar-Fard, Visual monitoring of civil infrastructure systems via camera-equipped unmanned aerial vehicles (UAVs): a review of related works, Visualization in Engineering 4 (1) (2016) 1–8, https://doi.org/10.1186/s40327-015-0029-z.

[32] L. Wang, F. Chen, H. Yin, Detecting and tracking vehicles in traffic by unmanned aerial vehicles, Autom. Constr. 72 (3) (2016) 294–308, https://doi.org/10.1016/j.autcon.2016.05.008.

[33] 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, Autom. Constr. 84 (2017) 70–80, https://doi.org/10.1016/j.autcon.2017.08.031.

[34] D. Kim, M. Liu, S. Lee, V.R. Kamat, Remote proximity monitoring between mobile construction resources using camera-mounted UAVs, Autom. Constr. 99 (2019) 168–182, https://doi.org/10.1016/j.autcon.2018.12.014.

[35] S. Bang, H. Kim, H. Kim, Vision-based 2D map generation for monitoring construction sites using UAV videos, 34th International Symposium on Automation and Robotics in Construction (ISARC 2017), Taipei, Taiwan, pp. 830–833, doi:10.22260/ISARC2017/0116.

[36] N. Elasal, D.M. Swart, N. Miller, Frame augmentation for imbalanced object detection datasets, Journal of Computational Vision and Imaging Systems 4 (1) (2018) 3–3 https://openjournals.uwaterloo.ca/index.php/vsl/article/view/341.

[37] I. Brilakis, M.-W. Park, G. Jog, Automated vision tracking of project related entities, Adv. Eng. Inform. 25 (4) (2011) 713–724, https://doi.org/10.1016/j.aei.2011.01.003.

[38] A. Dimitrov, M. Golparvar-Fard, Vision-based material recognition for automated monitoring of construction progress and generating building information modeling from unordered site image collections, Adv. Eng. Inform. 28 (1) (2014) 37–49, https://doi.org/10.1016/j.aei.2013.11.002.

[39] H. Son, C. Kim, N. Hwang, C. Kim, Y. Kang, Classification of major construction materials in construction environments using ensemble classifiers, Adv. Eng. Inform. 28 (1) (2014) 1–10, https://doi.org/10.1016/j.aei.2013.10.001.

[40] K.K. Han, M. Golparvar-Fard, Appearance-based material classification for monitoring of operation-level construction progress using 4D BIM and site photologs, Autom. Constr. 53 (2015) 44–57, https://doi.org/10.1016/j.autcon.2015.02.007.

[41] H. Hamledari, B. McCabe, S. Davari, Automated computer vision-based detection of components of under-construction indoor partitions, Autom. Constr. 74 (2017) 78–94, https://doi.org/10.1016/j.autcon.2016.11.009.

[42] H. Kim, K. Kim, H. Kim, Data-driven scene parsing method for recognizing construction site objects in the whole image, Autom. Constr. 71 (2016) 271–282, https://doi.org/10.1016/j.autcon.2016.08.018.

[43] J. Yang, Enhancing action recognition of construction workers using data-driven scene parsing, J. Civ. Eng. Manag. 24 (7) (2018) 568–580, https://doi.org/10.3846/jcem.2018.6133.

[44] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Advances in Neural Information Processing Systems, 2012, pp. 1097–1105 http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networ.

[45] H. Kim, H. Kim, Y.W. Hong, H. Byun, Detecting construction equipment using a region-based fully convolutional network and transfer learning, J. Comput. Civ. Eng. 32 (2) (2017) 04017082, , https://doi.org/10.1061/(ASCE)CP.1943-5487.0000731.

[46] H. Son, H. Choi, H. Seong, C. Kim, Detection of construction workers under varying poses and changing background in image sequences via very deep residual networks, Autom. Constr. 99 (2019) 27–38, https://doi.org/10.1016/j.autcon.2018.11.033.

[47] J. Kim, S. Chi, Action recognition of earthmoving excavators based on sequential pattern analysis of visual features and operation cycles, Autom. Constr. 104 (2019) 255–264, https://doi.org/10.1016/j.autcon.2019.03.025.

[48] Y. Annadani, C. Jawahar, Augment and adapt: A simple approach to image tampering detection, 24th International Conference on Pattern Recognition (ICPR 2018), Beijing, China, IEEE, pp. 2983–2988, https://doi.org/10.1109/ICPR.2018.8545614.

[49] D. Dwibedi, I. Misra, M. Hebert, Cut, paste and learn: Surprisingly easy synthesis for instance detection, The IEEE international conference on computer vision (ICCV 2017), Venice, Italy, arXiv preprint https://arxiv.org/abs/1708.01642.

[50] H. Inoue, Data augmentation by pairing samples for images classification, arXiv preprint arXiv:1801.02929, 2018.

[51] S.C. Wong, A. Gatt, V. Stamatescu, M.D. McDonnell, Understanding data augmentation for classification: when to warp? arXiv preprint arXiv:1609.08764, 2016.

[52] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, Adv. Neural Inf. Proces. Syst. (2014) 2672–2680 arXiv preprint arXiv:1406.2661 (2014).

[53] L. Bi, J. Kim, A. Kumar, D. Feng, M. Fulham, Synthesis of positron emission tomography (PET) images via multi-channel generative adversarial networks (GANs), arXiv preprint arXiv:1707.09747, 2017.

[54] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, H. Greenspan, GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification, Neurocomputing 321 (2018) 321–331, https://doi.org/10.1016/j.neucom.2018.09.013.

[55] T. Karras, S. Laine, T. Aila, A style-based generator architecture for generative adversarial networks, arXiv Preprint arXiv:1812.04948, 2018.

[56] B. Zhou, A. Khosla, A. Lapedriza, A. Torralba, A. Oliva, Places: an image database for deep scene understanding, arXiv Preprint arXiv:1610.02055, 2016.

[57] A. Gupta Dutta, A. Zissermann, VGG image annotator (VIA), arXiv Preprint arXiv:1904.10699, 2019.