ICCEPM 2022
The 9th International Conference on Construction Engineering and Project Management
Jun. 20-23, 2022, Las Vegas, NV, USA

Enhancing Occlusion Robustness for Vision-based Construction Worker Detection Using Data Augmentation

Yoojun Kim[1], Hyunjun Kim[2], Sunghan Sim[3], Youngjib Ham[4]

[1] *Department of Construction Science, Texas A&M University, 3137 TAMU, College Station, TX 77843, USA,* E-mail address: yoojun@tamu.edu
[2] *Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA,* E-mail address: hyunjun@illinois.edu
[3] *School of Civil, Architectural Engineering and Landscape Architecture, Sungkyunkwan University, Suwon 16419, Republic of Korea,* E-mail address: ssim@skku.edu
[4] *Department of Construction Science, Texas A&M University, 3137 TAMU, College Station, TX 77843, USA,* E-mail address: yham@tamu.edu

Abstract: Occlusion is one of the most challenging problems for computer vision-based construction monitoring. Due to the intrinsic dynamics of construction scenes, vision-based technologies inevitably suffer from occlusions. Previous researchers have proposed the occlusion handling methods by leveraging the prior information from the sequential images. However, these methods cannot be employed for construction object detection in non-sequential images. As an alternative occlusion handling method, this study proposes a data augmentation-based framework that can enhance the detection performance under occlusions. The proposed approach is specially designed for rebar occlusions, the distinctive type of occlusions frequently happen during construction worker detection. In the proposed method, the artificial rebars are synthetically generated to emulate possible rebar occlusions in construction sites. In this regard, the proposed method enables the model to train a variety of occluded images, thereby improving the detection performance without requiring sequential information. The effectiveness of the proposed method is validated by showing that the proposed method outperforms the baseline model without augmentation. The outcomes demonstrate the great potential of the data augmentation techniques for occlusion handling that can be readily applied to typical object detectors without changing their model architecture.

Key words: computer vision, AI, site monitoring, occlusion, data augmentation

1. INTRODUCTION

Recent advances in computer vision have contributed to automated monitoring of construction workers, such as worker detection and tracking [1–3], personal protective equipment compliance [4], and worker activity recognition [5]. Despite its great potential, the overall performance varies in practice due to the intrinsic dynamics of construction scenes (e.g., illumination variations [4], varying sizes [6], occlusions [1–3], etc.). Among them, occlusion is one of the most challenging problems that significantly undermines performance. Prior works on occlusion handling in construction leverage the prior information from the sequential images, such as worker's location and image features [1–3]. However, the prevailing methods cannot be employed in non-sequential images (e.g., a single image) and require additional computational costs, thereby increasing processing time. Furthermore, the prior methods are systemically vulnerable under continuous occlusions in which the prior information is also influenced by occlusions.

Data augmentation is a technique to artificially transform the dataset, which has the potential to address the occlusion problem. Because the occlusion problem is also critical to the classification of natural images, a wide variety of augmentation techniques have been developed in the field of computer science. The data augmentation methods generate different types of occlusion during training, such as a rectangle [7] and uniformly distributed square regions [8]. Although data augmentation in deep learning has shown improved classification and detection performance, its application to typical occlusions on the construction site has not been fully addressed.

This study presents a data augmentation strategy to construct a robust object detector, enhancing the improved detection of construction workers. Here, the proposed approach is specially designed to handle one of the most common occlusions caused by the steel bars for concrete reinforcement (hereafter termed rebar occlusion) installed on the construction site. Rebar occlusions are a unique occlusion type in construction scenes in which most workers behind such rebars are partially hidden unlike non-porous occlusions (e.g., concrete walls and columns) as the steel bars for concrete reinforcement are projected onto the image as grid patterns. As a proof of concept to investigate the effectiveness of data augmentation techniques for occlusion handling in construction scenes, the proposed approach in this study is particularly optimized to identify the construction works under rebar occlusion.

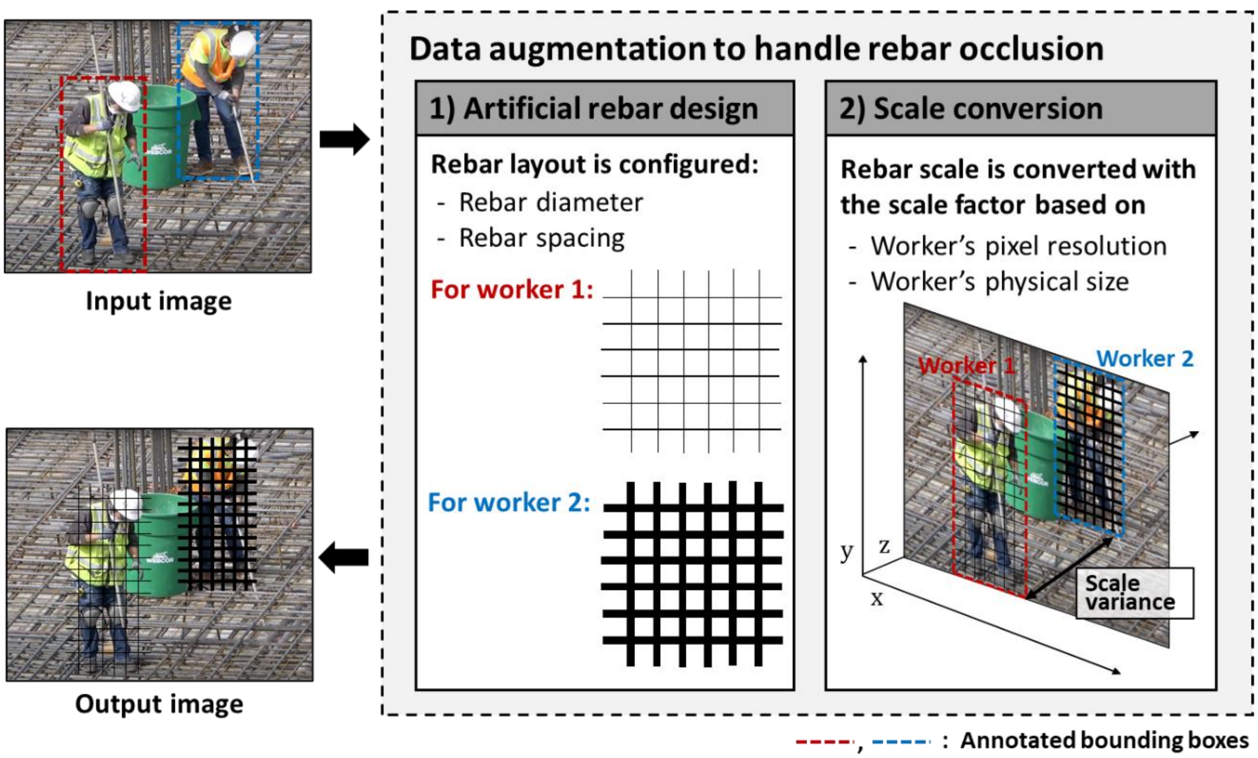

Figure 1. Overview of data augmentation technique to handle rebar occlusion

2. METHODOLOGY

This study proposes an occlusion handling method using data augmentation, constructing a robust object detector for the construction workers under the rebar occlusion. To effectively augment a dataset for the occlusion problem, the following two points are carefully addressed in this study: (1) the possible occlusions caused by the steel bars for concrete reinforcement are artificially designed and (2) its scale effect due to different working distances from the camera is additionally considered for the generation of the realistic dataset. Note that as the proposed method provides a data augmentation technique, it can be applied to typical object detectors without changing the architecture. Fig. 1 represents the overview of the proposed method to handle the rebar occlusion. Following the American Concrete Institute (ACI) 318-19 [9], artificial rebar layout is formed based on the design standards of steel bars for concrete reinforcement. To generate the rebar occlusion for each instance in the image, physical size of the worker is utilized to resize the artificial rebar layout. Thus, the realistic rebar occlusions can be generated for the training stage, building a robust object detector for the construction workers.

2.1. Artificial rebar design

The first step of the proposed method is to design artificial rebar layout based on the steel bars for concrete reinforcement. In general, the common rebars covering the view of the construction workers from a surveillance camera are the vertically installed rebars for the concrete wall that is the major structure in the construction industry [10]. Thus, the artificial rebar layout in this study is followed by the design standards of the concrete wall. Here, rebar diameter and spacing are the predefined parameters in the proposed method, as shown in Fig. 2. To consider the practical rebar sizes widely used in the construction projects [11], seven types of the diameter (i.e., 10, 12, 16, 20, 25, 32, and 40 mm) are employed for the proposed method. In addition, rebar spacing, a constant distance between rebars, is utilized as another parameter to generate the realistic rebar layout. According to ACI 318-19 [9], the minimum rebar spacing for the concrete wall can be determined as the greatest of 38.1 mm, 1.5 rebar diameter, and (4/3) maximum size of coarse aggregate. Because the standard for the size of coarse aggregate in the concrete wall is undefined, the remaining two criteria can be used for the proposed method to calculate the minimum rebar spacing. However, the sole use of the minimum rebar spacing generates a huge amount of rebar occlusions, potentially decreasing the performance of the trained network. In this regard, rebar spacing level, a variable to control a distance between the rebars, is additionally adopted to mitigate the severe rebar occlusion and diversify the level of rebar occlusion. The final rebar spacing s can be calculated as follows:

$$s = s_{\min} \times s_{level} \tag{1}$$

where smin and slevel are the minimum rebar spacing and rebar spacing level, respectively. The increase of the spacing level results in relatively small rebar occlusions in the image. In addition, if the rebar occlusion is conducted in the proposed data augmentation, a random value of the spacing level is selected between 1 and the predefined maximum value. Once rebar diameter and rebar spacing are determined, the artificial rebar layout is designed starting from the top-left corner.

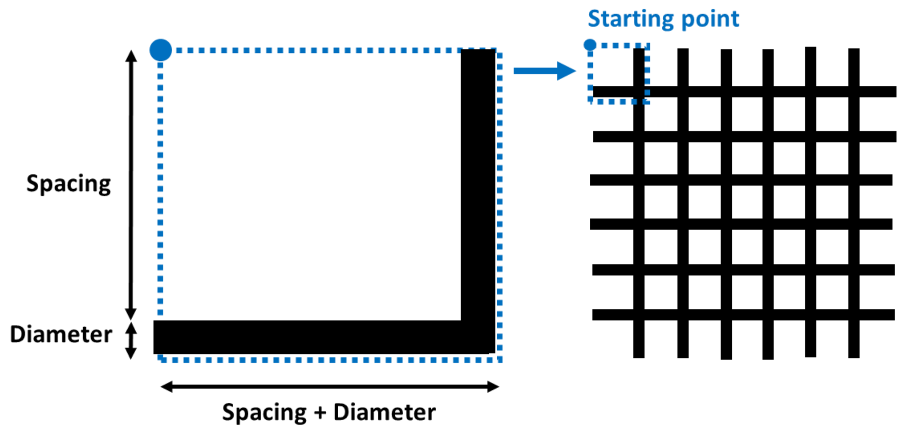

Figure 2. Overview of artificial rebar layout

2.2. Application of artificial rebars to images

To reflect the realistic rebar occlusion, the designed artificial rebar layout is scaled with respect to each size of the construction worker in the image. In general, a reference object with known dimensions on the construction site (e.g., construction worker [12] and column foundation [13]) is leveraged for the scale conversion. Likewise, the proposed method employs the construction worker's width as a reference object for the scale conversion in which its size is predefined as 425 mm by referring to the average width of Chinese male [14]. To apply the artificial rebar layout to each construction worker in the image, two assumptions are presented. First, the worker and the corresponding artificial rebar are in the same location. As surveillance cameras from a remote distance are typically used to monitor a wide range of the construction site in practice, a distance between the construction worker and the rebar is smaller than that from the camera to the rebar. For example, the distance between the worker and the rebar, denoted as d_2 in Fig. 3(a), is negligibly small compared to the distance from the camera to the rebar, defined by d_1. In this regard, it is assumed that the construction worker and the artificial rebar are projected onto the camera based on the same distance. The second assumption is that the width of the bounding box that includes an instance represents the construction worker's shoulder width, regardless of the worker's pose, as shown in Fig. 3(b). Given these assumptions, a scale factor for each worker in the image can be calculated as the bounding box width divided by the worker's shoulder width of 425 mm.

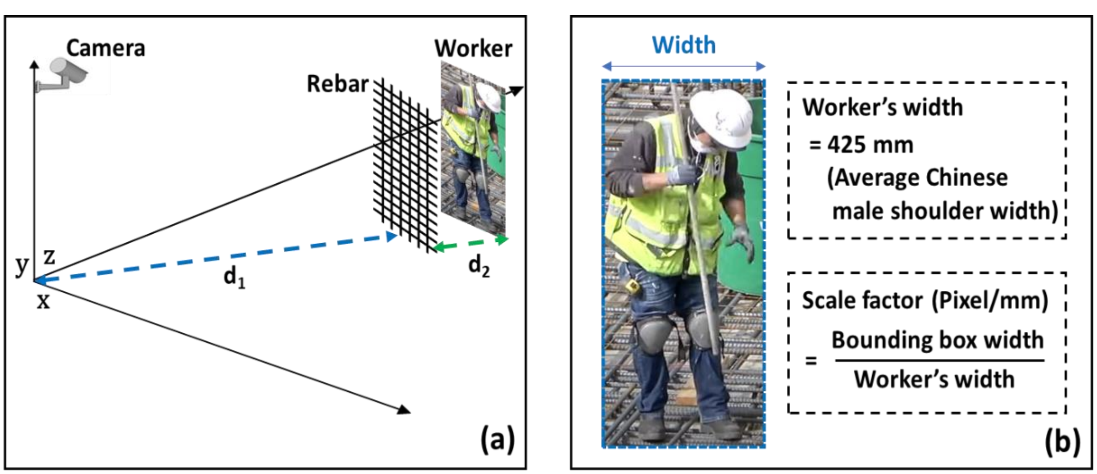

Figure 3. Assumptions for the scale conversion

3. EXPERIMENTAL VALIDATION

3.1. Data preparation

A set of images taken from the Moving Objects in Construction Sites (MOCS) dataset [15] are used to investigate the performance of the proposed method. MOCS is the largest public dataset, including 43,000 images collected from 174 construction sites in China and Pakistan from 2005 to 2019. A total of 222,861 instances for 13 categories (e.g., construction worker, tower crane, roller, bulldozer, and pump truck) were manually annotated by the bounding boxes to be used for training the object detector. For training dataset in this study, the MOCS training dataset images including 77,770 instances of construction worker are entirely used. For testing dataset, the images including rebar occlusion are manually seleceted from MOCS testing dataset. As a result, a total of 4,175 instances of construction worker in 701 images is collected. To compare the effectiveness on rebar occlusions, the collected images are further classified into 'test set A (rebar occlsuion)' and 'test set B (no rebar occlusion)'. Table 1 represnts the summary of the two test sets. The test set A is designed to validate the detection performance with regard to rebar occlusion. To this end, the construction workers who are not under rebar occlusion in the test set A are erased. On the other hand, the purpose of the test set B is to test the detection performance under no rebar occlusion. Thus, the construction workers under rebar occlusions are erased contrary to the test set A. Fig 4 shows the example images in test set A (2nd image) and test set B (3rd image).

Table 1. Summary of testing dataset

	Test set A (rebar occlusion)	Test set B (no rebar occlusion)
Number of workers	1,716	2,459
Number of images	701	527

Figure 4. Example images in raw dataset (1st), test set A (2nd), and test set B (3rd)

3.2. Experimental setup

Faster R-CNN [16] with a ResNet-50 [17]-FPN [18] is used as a benchmark object detector for three methods. In addition, the default model parameters in Detectron2 from Facebook AI Research [19] are used except for the training epoch that is doubled to compensate for the additional occlusions in the same manner as [15]. For the model parameters, the learning rate starts with 0.005, and further decrease to 0.0005 after 80 epochs and to 0.00005 after 100 epochs until 110 epochs. During the experiments, the models were initialized with the pre-trained weight of ImageNet [20] and then fine-tuned on the training dataset in Dectectron2. For the hyperparameters of the proposed method, the probability (i.e., frequency of occlusion per worker) is set as 0.5.

3.3. Experimental results

The quantitative results in detecting the construction workers are compared to validate the proposed data augmentation method, as shown in Table 2. The object detectors with three different augmentation techniques are used to compare their identification accuracies with a baseline model in which data augmentation is not used during training. Here, average precision (AP) in PASCAL VOC [21], an area under a precision-recall curve, is selected as a measure to evaluate the performance.

Table 2. Comparison of the performance with and without the proposed data augmentation

Model	Test set A (rebar occlusion)	Test set B (no rebar occlusion)
Without augmentation	52.196	77.795
With augmentation	59.777 (+7.581)	81.459 (+3.664)

Table 2 shows the evaluation results with the test set A and test set B. The results show that the proposed approach outperforms the baseline by 7.581% on the test set A (rebar occlusion). Compared to the baseline approach (i.e., without augmenataion), the proposed method also improves the worker detection performances by 3.664% on the test set B (no rebar occlusion). This indicates that the proposed approach improves the robustness of the models against not only occlusions, but also to other image features (e.g., illumination, clutter, etc.). This is because the approach generating artificial occlusion during training enables the models to train relatively less distinctive image features, thus improving the perception ability in the same manner of the other occlusion-based data augmentation techniques. Note that this performance improvement is obtained by only augmenting the input data without changing the model architecture.

4. CONCLUSION AND FUTURE WORK

We proposed a data augmentation-based mechanism that enhances the model's detection performance for the construction workers under rebar occlusions. The rebar occlusion is artificially generated by considering the standards of rebar size and spacing, as well as the effect of its scale factor caused by different distances from the camera. As the performance of the proposed approach is successfully validated from the images on the construction site, continuous monitoring of the workers can be performed to ensure the safety issue. This study provided a data augmentation technique customized to handle rebar occlusions that can be easily applied in any object detector. The effectiveness of the proposed method was systematically validated for the cases of rebar occlusion and no rebar occlusion. The experimental results demonstrated the great potential of the proposed data augmentation approach. However, there are some limitations of this study that need to be addressed by future work. For instance, the proposed method predetermined the pixel value of occlusion as 0 (i.e., black) and the patterns of artificial rebar layout that horizontal and vertical rebars are uniformly distributed on the entire construction worker. Also, the parameters of the proposed methods such as rebar diameters and rebar spacing are presumed based on the widely used seven diameters and ACI 318-19. In this regard, future research could center on optimizing the artificial rebar design with different patterns (e.g., reinforced concrete columns) or addressing the characteristics of certain construction jobsite (e.g., structural drawings of rebars) for the better efficiency of the proposed method.

REFERENCES

[1] M. Park, I. Brilakis, “Continuous localization of construction workers via integration of detection and tracking”, Automation in Construction, vol. 72, pp. 129–142, 2016.

[2] H. Son, C. Kim, “Integrated worker detection and tracking for the safe operation of construction machinery”, Automation in Construction, vol. 126, pp. 10367, 2021.

[3] O. Angah, A. Chen, “Tracking multiple construction workers through deep learning and the gradient based method with re-matching based on multi-object tracking accuracy”, Automation in Construction, vol. 119, pp. 103308, 2020.

[4] N. Nath, A. Behzadan, S. Paal, “Deep learning for site safety: Real-time detection of personal protective equipment”, Automation in Construction, vol. 112, pp. 103085, 2020.

[5] X. Luo, H. Li, X. Yang, Y. Yu, D. Cao, “Capturing and Understanding Workers’ Activities in Far-Field Surveillance Videos with Deep Action Recognition and Bayesian Nonparametric Learning”, Computer-Aided Civil and Infrastructure Engineering, vol. 34, pp. 333–351, 2019.

[6] 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”, Automation in Construction, vol. 99, pp. 27–38, 2019.

[7] Z. Zhong, L. Zheng, G. Kang, S. Li, Y. Yang, “Random Erasing Data Augmentation”, Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 13001–13008, 2020.

[8] P. Chen, S. Liu, H. Zhao, J. Jia, “GridMask Data Augmentation”, 2020.

[9] American Concrete Institute (ACI), “ACI 318-19: Building code requirements for structural concrete”, 2019.

[10] A. Jarkas, P. Eng, C. Eng, “Analysis and Measurement of Buildability Factors Influencing Rebar Installation Labor Productivity of In Situ Reinforced Concrete Walls”, Journal of Architectural Engineering, vol. 18, pp. 52–60, 2012.

[11] M. Kim, J. Thedja, H. Chi, D. Lee, “Automated rebar diameter classification using point cloud data based machine learning”, Automation in Construction, vol. 122, pp. 103476, 2021.

[12] K. Kim, H. Kim, H. Kim, “Image-based construction hazard avoidance system using augmented reality in wearable device”, Automation in Construction, vol. 83, pp. 390–403, 2017.

[13] D. Kim, M. Liu, S. Lee, V. Kamat, “Remote proximity monitoring between mobile construction resources using camera-mounted UAVs”, Automation in Construction, vol. 99, pp. 168–182, 2019.

[14] Y. Daozhong, “Report on the investigation and research of the physique of Chinese staff members and workers”, FUJIAN SPORTS SCIENCE AND TECHNOLOGY, 1997.

[15] A. Xuehui, Z. Li, L. Zuguang, W. Chengzhi, L. Pengfei, L. Zhiwei, “Dataset and benchmark for detecting moving objects in construction sites”, Automation in Construction, vol. 122, pp. 103482, 2021.

[16] S. Ren, K. He, R. Girshick, J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", Advances in Neural Information Processing Systems, vol. 28, 2015.

[17] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, 2016.

[18] T. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, S. Belongie, "Feature Pyramid Networks for Object Detection", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2117-2125, 2017.

[19] Y. Wu, A. Kirillov, F. Massa, W. Lo, R. Girshick, Detectron2, https://github.com/facebookresearch/detectron2/, 2019.

[20] J. Deng, W. Dong, R. Socher, L. Li, K. Li, L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 248-255, 2009.

[21] M. Everingham, L. Van Gool, C. Williams, J. Winn, A. Zisserman, "The Pascal Visual Object Classes (VOC) Challenge", International Journal of Computer Vision, vol. 88, no. 2, pp. 303-338, 2009.