CCC 2020
Proceedings of the Creative Construction e-Conference (2020) 006
Edited by: Miroslaw J. Skibniewski & Miklos Hajdu
https://doi.org/10.3311/CCC2020-006

Nested Network for Detecting PPE on Large Construction Sites Based on Frame Segmentation

Mohammad Akbarzadeh[1], Zhenhua Zhu[2] and Amin Hammad[3]

[1] *Concordia University, Montreal, Canada, mohammad.akbarzadeh@concordia.ca*
[2] *University of Wisconsin-Madison, Madison, USA, zzhu286@wisc.edu*
[3] *Concordia University, Montreal, Canada, hammad@ciise.concordia.ca*

Abstract

Safety is a main concern for the construction industry because of the high rate of accidents and casualties on construction sites. Personal Protective Equipment (PPE) is a major part of safety regulations to prevent accidents. However, workers may neglect to wear the required PPE while working, which subsequently increases the potential risk for accidents. Currently, safety managers and inspectors on construction sites are responsible for monitoring safety regulations, which is a time-consuming task. To facilitate safety monitoring, a large number of research studies applied computer vision for detecting PPE on construction sites. Nevertheless, detecting workers and PPE is still a challenge in far-field videos. This research proposes an approach for detecting if anyone on the construction site is wearing the required PPE, even when he or she is far from the surveillance cameras. This method uses a frame segmentation technique and a nested network with two Faster R-CNN models to detect safety noncompliances. The first model detects the human bodies on the construction site, and the second one detects if the detected person is wearing a hardhat and a safety vest. The proposed method is applied to videos from a construction site. The experimental results demonstrate the practicality and robustness of the proposed method to detect PPE in far-field videos. Based on three different test videos, the average precision and recall for the worker detection model were 99.67% and 92.92%, respectively. The PPE detection model had the average precision and recall of 91.25% and 94.77%, respectively.

Keywords: computer-vision, construction safety, far-field surveillance videos, faster R-CNN

1. Introduction

Safety regulations are not always followed on construction sites, which is the main reason for accidents. According to [1], more than 450 workers were killed, and over 63,000 workers were injured on construction sites in Canada in 2017. These accidents cost nearly $19.8B each year. Based on the statistics from the Association of Workers' Compensation Boards of Canada (AWCBC) [2] in 2017, 951 workspace fatalities were recorded in Canada with an increase of 46 from the previous year [3]. One of the most crucial ways of preventing accidents is to use personal protective equipment (PPE). In addition to fatal injuries and casualties, there are other consequences of accidents [4]: time loss of project execution, damaging the reputation of the firm, mental illness of workers, cost of medical care, cost of recruiting and training new workers, compensation cost, cost of repairs and additional supervision, productivity loss, and cost of accident investigation. Safety inspectors are responsible for ensuring that safety regulations are followed by contractors to avoid accidents [5]. Hardhats and safety-vests are the most fundamental PPE. "Employees working in areas where there is a possible danger of head injuries from impact, or from falling or flying objects, or from electrical shock and burns, shall be protected by helmets" [6]. Canadian Centre for Occupational Health and Safety (CCOHS) emphasizes the importance of wearing High Visibility Safety

Apparel (HVSA) for different lighting conditions and working close to moving vehicles [7]. Due to the nature of the construction industry, detecting workers and their PPE from surveillance videos is a challenging task for the following reasons: (1) bad weather conditions, (2) low lighting conditions, (3) low camera resolution, (4) camera height, (5) narrow field-of-view of the camera, and (6) occlusion [8]. Among these challenges, occlusion is the most significant barrier to object detection. To facilitate safety monitoring, a large number of research studies applied computer vision (CV) for detecting PPE on construction sites. Nevertheless, detecting workers and PPE is still a challenge in far-field videos. This paper proposes a novel nested network based on frame segmentation that consists of two Deep Neural Networks (DNN) to facilitate safety monitoring on construction sites. The first model detects the human bodies on the construction site, and the second one detects if the detected person is wearing a hardhat and a safety vest. The proposed method is validated based on the videos collected from a real construction site. The results show the effectiveness and practicality of the proposed method for detecting the workers far from the camera as well as detecting compliance or noncompliance with the PPE regulations.

2. Methodology

In this research, surveillance cameras are installed at a height to reduce the occlusion and cover most of the construction site. Fig. 1 shows the projection of workers on the image plane under these conditions, where the center of projection (COP) is the center of the surveillance camera, and three workers are at different locations. Based on the angle of view of the camera, the worker far from the camera is captured on the upper part of the image plane. The angle of projection for this worker becomes smaller and makes him appear smaller on the image plane because of the perspective view [9]. As will be explained in the case study, the worker in far-field could be about 1/3 the size for workers in near-filed on the image frame. A novel frame segmentation nested network is proposed to overcome the challenge of detecting workers and PPE. The proposed method consists of two nested Faster R-CNN models that are applied sequentially. These models are custom-trained using the transfer learning approach. Fig. 2 shows the overall flow of detection for every frame in the surveillance video.

2.1. Worker detection module

The state-of-the-art Faster R-CNN [10] is a robust object detection algorithm that uses a Region Proposal Network (RPN). Faster-RCNN has an input frame size of 1024×600, which has an aspect ratio of 1:7. High Definition (HD) surveillance cameras used in construction sites have the resolution of 1920×1080. This resolution is larger than the input frame size of the Faster R-CNN model, and the network resizes frames in order to fit the input size in both detection and training stages. As shown in Fig. 1 a worker in the far-field is captured in a small area on the image frame. Additionally, as a result of resizing, the worker becomes even smaller, which makes the detection more challenging and also affects the training performance. The worker detection module has four main steps: frame segmentation, worker detection, detection refinement, and removing duplications. The objectives of the worker detection module are: (1) detecting far-field workers, (2) eliminating the resizing effect, and (3) covering all workers with segments. Fig. 3 shows the components of the worker detection module.

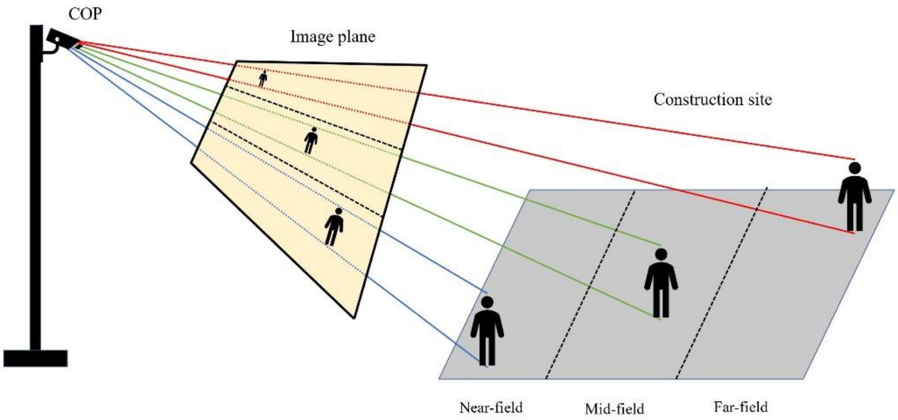

Fig. 1. Projection of workers on the image plane under the real-world condition

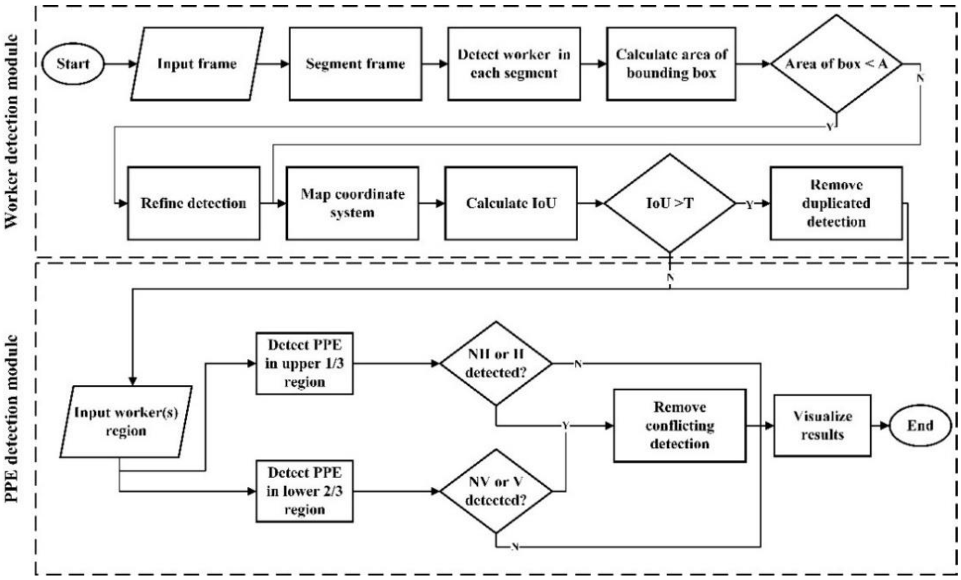

Fig. 2. The overall frame segmentation based nested network detection approach

Three main segments are defined parallel to the horizontal axis of the image frame. I is the near-field strip where workers are captured in the largest size compared to the whole frame, K is where workers are captured in medium-size (mid-field) and J, where workers are captured at the smallest size on the image plane (far-field). Additionally, sub-segments are needed within each main segment to meet the input size and aspect ratio of the network. Workers' width on the image plane is the smaller dimension of a worker and is used to calculate the sub-segments' width. Equation 1 is used in order to get the number of the non-overlapping sub-segments where N is defined as a number of workers fitting into a sub-segment. Different values of N are considered to get the best detection results and compared with the aspect ratio of the Faster R-CNN network. In order to make sure that the detection fully covers workers intersecting with the borders of the main segments/sub-segments, overlapping main segments and sub-segments are defined.

$$\text{Number of sub-segments} = \text{Ceiling}\left(\frac{\text{Frame width}}{\text{N} \times \text{Average width of worker}}\right) \quad (1)$$

$$\text{Sub-segment width} = \frac{\text{Frame width}}{\text{Number of sub-segments}} \quad (2)$$

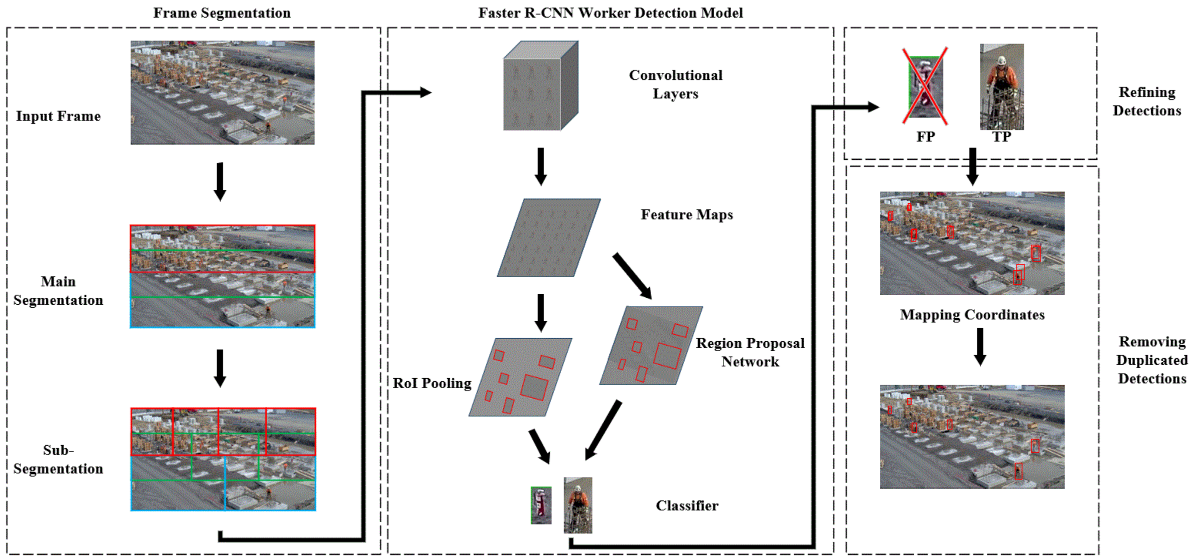

Fig. 3. Components of the worker detection module

In some cases, some objects (e.g., detecting traffic cones) might be detected as a worker. In order to refine the detection results, this study considers the average area of workers in each main segment to remove the false detections from the detection lists. In addition, the detection of workers in overlapping segments results in double counting of workers. The proposed solution is to first map the detection coordinates in the segments to the uniform frame coordinates and then using the IoU cost function to remove duplications. IoU represents the intersection of the ground truth bounding box with the detection bounding box divided by the union area, and is calculated between all the detected objects in the same frame. A threshold (*T*) is defined to remove the duplicates when the IoU is higher than *T*.

2.2. PPE detection module

The next step after detecting workers is to detect whether they are wearing the required PPE (i.e., hardhat and safety-vest). Detecting hardhats and safety-vests is a challenging task, especially under the far-field area with a small angle of view. Therefore, a nested Faster R-CNN model is used, which relies on the detection outputs from the worker detection module. The PPE detection dataset considers the labels H (hardhat), NV (no-safety-vest), NH (no-hardhat), and V (safety-vest). If NH (no-hardhat) or NV (no-vest) are detected, it will be considered as dangerous behaviour. Then, potential regions are defined for each object of interest, where H or NH must be detected in the upper 1/3 region of the detected worker bounding box and V or NV in the lower 2/3 region. In addition, this research considers conflicting detections, where the PPE detection model returns both wearing and not wearing a hardhat or a safety-vest at the same time, based on the confidence level of the detections.

3. Implementation and case study

Surveillance videos from a construction site are used to validate the proposed method. The construction site is a power substation. The construction site is located in Montreal, Canada, where Axis P1425-E surveillance cameras with HD resolution (1920×1080) installed on four poles at about 10 m height. The training for worker detection and PPE detection was done using two primary datasets containing 2200 and 1000 images, respectively. In the worker detection dataset, the main object of interest is human. Since the region of interest for the PPE detection model is the human body, the PPE dataset contains cropped images of persons. The PPE dataset is created by combining the CUHK01 dataset [11] that contains people captured from a high angle of view, as negative examples of workers with no PPE, and the image dataset of workers with PPE from the site. The images in both datasets are annotated using open source software LabelImg [12] using PASCAL (pattern analysis, statistical modelling and computational learning) format [13]. Examples of worker and PPE annotations are shown in Fig. 4.

Fig. 4. Examples of workers and PPE annotations

The three main segments of *I*, *K* and *J*, are defined with an equal size of 1920×540 parallel the horizontal axis of the image frame. However, the size of the main segments does not fit with the input size of the Faster R-CNN network. As explained in Section 2.1, in order to find the optimum number of sub-segments considering the accuracy and detection time, using Equation 1, three values of 5, 15 and 25 are considered for *N*. The average width of workers is measured on the image plane. The average width is 55 pixels in the main segment *I*, 35 pixels in main segment *K*, and 20 pixels in the main segment *J*. The total number, size, and aspect ratio of sub-segments with different *N* are summarized in Table 1.

Table 1. Number of sub-segments with overlaps for detection based on different N

N	Sub-segment information	Main segment I	Main segment K	Main segment J
	No. sub-segments with overlaps	13	21	39
5	W x H (pixels)	274×540	174×540	96×540
	Aspect ratio	0.51	0.32	0.17
	No. sub-segments with overlaps	5	7	13
15	W x H (pixels)	640×540	480×540	274×540
	Aspect ratio	1.19	0.89	0.51
	No. sub-segments with overlaps	3	5	7
25	W x H (pixels)	960×540	640×540	480×540
	Aspect ratio	1.78	1.19	0.89

Three 5-minute validation videos recorded with 30 frames per second are selected from different phases of the project, and detection is performed every one second. Precision, recall, and accuracy are calculated to evaluate detection results. The results are based on assuming the value of 50% for the IoU. Based on the results of the worker detection model, the best detection results are for N equals 25, where the image frame is divided into a total of 15 sub-segments that fit input dimensions and aspect ratio of the Faster R-CNN person detection model. The average precision and recall for the worker detection model are 99.67% and 92.92%, respectively. The proposed method based on frame segmentation improved workers detection on large construction sites compared to the literature [14]. In order to evaluate the PPE detection model, 200 images were gathered, with half of them for workers wearing PPE and the other half for cases where PPE is not used. Precision and recall are calculated to evaluate the PPE detection module, which are summarized in Table 2. The PPE detection module achieved higher precision and recall compared to [15], which similarly detected the PPE within the worker's bounding box.

Table 2. PPE detection results

Classes	Precision (%)	Recall (%)	Accuracy (%)
H	94.06	95.96	90.47
NH	97.93	93.14	91.34
V	80.00	96.97	78.05
NV	93.00	93.00	86.92
Average	91.25	94.77	86.70

4. Summary and conclusions

This paper proposed a nested network for detecting workers and PPE on large construction sites based on frame segmentation techniques. The framework combines two Faster R-CNN models in order to detect workers and PPE. Three main segments of near, mid, and far-fields of view are defined. Sub-segments are defined for each main segment to meet the required input size and aspect ratio of the worker detection model. Detection results are first refined based on comparison with the average area of workers in each main segment; then, the results are mapped from sub-segments to the original frame. Duplicated detections are removed from the detection list of workers. Moreover, the PPE detection module defines potential regions for each type of PPE to be detected. Detection results are compared based on the confidence level of the detection to remove any conflict (e.g., detecting *H* and *NH* at the same time). The final output of the nested network indicates if a worker is wearing a hard and safety vest or not. Based on the case study results, the proposed method improved the detection for far-field workers and PPE.

5. References

[1] Construction workers: 3 or 4 times more accidents - SPI Health and Safety. https://www.spi s.com/en/blog/item/construction-workers-3-or-4-times-more-accidents (accessed May 23, 2019).
[2] Association of Workers' Compensation Boards of Canada / ACATC. http://awcbc.org/ (accessed Apr. 09, 2020).
[3] Canadian Centre for Occupational Health and Safety Government of Canada, Oct. 02, 2019. https://www.ccohs.ca/ (accessed Apr. 09, 2020).
[4] K. Arunkumar and J. Gunasekaran, "Causes and Effects of Accidents on Construction Site," *International Journal of Engineering Science and Computing,* vol. 8, no. 6, p. 9, June 2018.
[5] Employment and social development Canada, "Workplace Safety," aem, Feb. 11, 2009. https://www.canada.ca/en/employment-social-development/services/health-safety/workplace-safety.html (accessed Apr. 26, 2020).
[6] Personal Protective Equipment - Head Protection. https://www.osha.gov/SLTC/etools/logging/manual/logger/head_protection.html (accessed May 12, 2020)
[7] Canadian Centre for Occupational Health and Safety Government of Canada, "High-Visibility Safety Apparel : OSH Answers," Aug. 14, 2019. http://www.ccohs.ca/ (accessed Aug. 14, 2019).
[8] H. Siddiqui, "UWB RTLS for Construction Equipment Localization: Experimental Performance Analysis and Fusion with Video Data," p. 161.
[9] J.-I. Jung, S.-Y. Seo, S.-C. Lee, and Y.-S. Ho, "Enhanced Linear Perspective using Adaptive Intrinsic Camera Parameters," ICESIT2010, p. 7.
[10] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, https://doi.org/10.1109/TPAMI.2016.2577031.
[11] W. Li, R. Zhao, T. Xiao, and X. Wang, "DeepReID: Deep Filter Pairing Neural Network for Person Re-identification," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, Jun. 2014, pp. 152–159, https://doi.org/10.1109/CVPR.2014.27.
[12] Tzutalin, LabelImg. Git code, (2015). https://github.com/tzutalin/labelImg.
[13] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes (VOC) Challenge," Int J Comput Vis, vol. 88, no. 2, pp. 303–338, Jun. 2010, https://doi.org/10.1007/s11263-009-0275-4.
[14] B.E. Mneymneh, M. Abbas, and H. Khoury, "Vision-Based Framework for Intelligent Monitoring of Hardhat Wearing on Construction Sites," *Journal of Computing in Civil Engineering*, vol. 33, no. 2, p. 04018066, Mar. 2019, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000813
[15] N. D. Nath, A. H. Behzadan, and S. G. Paal, "Deep learning for site safety: Real-time detection of personal protective equipment," *Automation in Construction*, vol. 112, p. 103085, Apr. 2020, https://doi.org/10.1016/j.autcon.2020.103085