Automation in Construction 89 (2018) 58–70

Contents lists available at ScienceDirect

Automation in Construction

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

Transfer learning and deep convolutional neural networks for safety guardrail detection in 2D images

Zdenek Kolar, Hainan Chen, Xiaowei Luo*

Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong

A R T I C L E I N F O

Keywords:
Computer vision
Construction safety
Guardrail detection
Convolutional neural networks
Transfer learning
VGG-16

A B S T R A C T

Safety has been a concern for the construction industry for decades. Unsafe conditions and behaviors are considered as the major causes of construction accidents. The current safety inspection of conditions and behaviors heavily rely on human efforts which are limited onsite. To improve the safety performance of the industry, a more efficient approach to identify the unsafe conditions on site is required to supplement the current manual inspection practice. A promising way to supplement the current manual safety inspection is automated and intelligent monitoring/inspection through information and sensing technologies, including localization techniques, environment monitoring, image processing and etc. To assess the potential benefits of contemporary technologies for onsite safety inspection, the authors focused on real-time guardrail detection, as unprotected edges are the ones cause for workers falling from heights.

In this paper, the authors developed a safety guardrail detection model based on convolutional neural network (CNN). An augmented data set is generated with the addition of background image to guardrail 3D models and used as training set. Transfer learning is utilized and the Visual Geometry Group architecture with 16 layers (VGG-16) model is adopted to construct the basic features extraction for the neural network. In the CNN implementation, 4000 augmented images were used to train the proposed model, while another 2000 images collected from real construction jobsites and 2000 images from Google were used to validate the proposed model. The proposed CNN-based guardrail detection model obtained a high accuracy of 96.5%. In addition, this study indicates that the synthetic images generated by augment technology can be used to create a large training dataset, and CNN-based image detection algorithm is a promising approach in construction jobsite safety monitoring.

1. Introduction

As reported by Occupational Safety and Health Administration (OSHA), approximately 40% of all fatalities at construction sites are caused by falls from heights, followed by struck by objects, electrocution, and caught-in/between [1]. More than 4500 US workers (3.4 per 100,000 full-time workers, more than 13 deaths every day) died on the job in 2015 [1]. According to OSHA's statistics, “Fall protection, construction” is at top of the list of the most frequently violated OSHA standards [2].Reducing accidents of falling from height can significantly improve the safety performance in the construction industry. To reduce those accidents, both proactive and passive approaches have been proposed for implementation on construction jobsite.

Construction safety planning and training are considered as a proactive approach to improve worker's safety awareness and reduce the falling risk [3–5]. Although proper planning can reduce the falling risks on site, it cannot eliminate them. Due to worker's absent-mind or overconfidence, workers are still exposed to high risk of falling. The passive falling prevention measures, such as guardrail, warning lines, fall arrest systems [6,7], act as the onsite approach for reducing falling risk. Guardrail systems are used on many work surfaces (rooftops, scaffolds, platforms and etc.) to prevent the workings from falling onto a lower level. There have been accidents, in which the required guardrail systems or part of the system were missing [8]. In current practice, the guardrail system inspection belongs to the responsibility of safety officers on site and they have to manually check whether the guardrail system is in-place and complete. However, not all the missing safety guardrail situation can be found in time because the limited-number safety officers with other workloads cannot cover all the areas on the jobsite at every moment. The advancement and pervasiveness of information technology enable the inspection work to be done automatically in real time [9–12]. In a real-time safety monitoring system, once the guardrail system is missing where it is required, an alert can be sent to the responsible person immediately.

* Corresponding author.
E-mail address: xiaowluo@cityu.edu.hk (X. Luo).

https://doi.org/10.1016/j.autcon.2018.01.003
Received 12 May 2017; Received in revised form 2 January 2018; Accepted 9 January 2018

To accomplish automatic checking of whether the guardrail system is set up appropriately, the first step would be to detect the existence of guardrail system correctly. Traditionally, the feature based computer vision techniques are used to process the videos/images for object detection and classification [13,14]. For each individual task, a precise feature model should be established first, and this model development heavily relies on complex statistic, mathematic and image processing theories. In addition, those traditional image processing techniques have strict constraints on the inputs of the model, including sizes, pixel-perfect precision, and channels of the images [15]. Therefore, the flexibility of those technologies is greatly limited. Different from the conventional image processing techniques, the deep neural network is an end-to-end process, instead of the step-by-step one. The feature extraction is accomplished through neural network model training. Since the neural network is a black box algorithm, the features can be extracted without any prior knowledge. Using CNN with different convolutional core and convolutional layers of different numbers, the features of one image can be extracted into the different precision levels and multiple dimension space. Therefore, once the CNN have enough cores and layers, the whole feature space of one image can be obtained through model training without external interaction. The shortcoming of the deep neural network is that training a deep neural network model is time-computing and computationally intensive, which means longer time and more expensive hardware are required. As the deep neural network is not sensitive to the features, the models trained for some special purposes can be used or partially integrated into the new model with the help of transfer learning. In this way, the time and hardware cost would be significantly reduced, making the deep neural network more practical. In this paper, the authors proposed a CNN-based guardrail detection model with the integration of the core part of the VGG-16 model and multi-layers perception (MLP) network, and validated it using image data collected from jobsite and internet. The remaining of the paper is organized as follows. Section 2 summarizes the related research in computer vision and construction sensing domain. Section 3 describes the CNN and transfer learning techniques used in the study. Section 3.1 explains the validation dataset and results. Results are presented and explained in Section 4. Finally, conclusions are drawn in Section 5.

2. Literature review

In this section, the authors first review the commonly used algorithms for computer vision in the construction-related area. To address the limitation of the conventional computer vision methods used in construction, the techniques in computer vision-based object detection in the computer science discipline were reviewed. Based on the comparison of available techniques in computer vision based object detection, CNN is considered as a promising approach to address the mentioned limitations. Therefore, CNN is introduced in the 3rd subsection.

2.1. Application of computer vision in construction

Computer vision techniques have been used in construction-related research for object detection. The studies are focused on detection of construction workers, site machinery and progress tracking in construction [16, 17, 18, 19, 20]. Teizer [21] describes the status quo and challenges in computer vision in construction. Of the computer vision techniques used in construction, the histogram of oriented gradients (HOG) is one of the widely used techniques. Park and et al. [16] use HOG and the histogram of HSV (Hue, Saturation, Value) colors as an input for k-nearest neighbors (KNN) classifier. HOG, Histogram of Optical Flow (HOF), Motion Boundary Histogram (MBH) were used for action recognition of construction workers [22]. Besides HOG, Haar Cascade [23] is another popular technique used in construction. Du et al. [24] use an approach based on Haar Cascade to detect the hard hats of workers on construction. Kim et al. [25] used a combination of KNN and Scale Invariant Feature Transform (SIFT) algorithms to parse a complete image from a construction site. It is quite straightforward that the extracted features are used to code one image, and then to conduct classification and clustering for the images labeling. However, for traditional computer vision technologies, the features are extracted by the predefined and special-purpose optimized models. Those models can only be manually developed when high-dimensional features are required. Therefore, when multi-features models are simultaneously considered, the conventional computer vision methods (HOG, HOF, Motion Boundary Histogram (MBH), etc.) lose their advantages or even may fail to perform the designed tasks.

2.2. Techniques in computer vision-based object detection

The computer vision techniques in construction-related research derive from computer vision disciplines. In computer vision, there have been two main approaches for object detection and classification. The first one aims to determine the correct class of a given sample based on a set of features (often handcrafted) that are specific to the given class. An example of algorithms to meet this need is SIFT, proposed by Lowe [26]. This approach can be extended with visual bag-of-words, which aims to determine the class of a sample based on the frequencies of individual features in the image [27]. In the traditional computer vision process, the object detection and classification is still subjected to the feature extraction. Without an appropriate feature model, it is difficult to implement accurate object detection and classification.

The second approach is designed to automatically extract a large number of image features and then use these features for image classification. Examples include CNNs, HOGs, etc. The advantage of those algorithms is that they have the self-learning ability from a given dataset. Since the deep neural network is an end-to-end process, it is not necessary to conduct feature extraction in advance. Along with the advancement in computer hardware and software, the deep neural network is considered as the most powerful technology to process the image and to solve the computer-vision related problems.

With established objective detection system, visual sensors can play a useful and important role in various management work. Veres et al. [28] proposed a robust approach for workflow classification in industrial environments with employing computer vision processing. Zhu et al. [29] investigated the workforce and equipment detection and tracking on construction jobsite with video processing technology. Park et al. [30] employed computer vision based objects detection to detect and track construction workers position on construction site for safety and productivity monitoring. Along with the rapid development of the image processing technology, computer vision based objects detection is considered as a very valuable approach for construction management.

2.3. Convolutional neural networks

Images, seen as matrices, contain a great number of values. A small image, used for example for digit classification, may consist of 28 × 28 pixels in grayscale mode. That is 784 features, each may take value from the interval between 0 and 255. The size of modern-era images may, however, be far greater than that. Due to the number of features and loss of the spatial relationship between pixels, artificial neural networks (ANNs) are not the ideal solution for image classification [31].

In order to overcome the disadvantages of standard ANNs for image classification, LeCun et al. [31] developed a technique for digit recognition for the US Post, based on CNNs. For training and validation, the authors used the Modified National Institute of Standards and Technology (MNIST) dataset [32].

The limitation of CNNs at that time was, due to the limited computational power available, it was possible to process only small images (the resolution of the MNIST dataset is only 28 × 28 pixels). Thanks to