

CLASSIFICATION OF CONSTRUCTION TRADE AND QUANTIFICATION OF WORK EFFICIENCY USING POSTURE RECOGNITIONS AND DEEP NEURAL NETWORKS

Ling CHEN¹

1. Ph.D., Postdoctoral Fellow, Department of Building and Real Estate, Faculty of Construction and Environment, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong.
(email: ling.a.chen@connect.polyu.hk)

Kaixin Johnny LIN²

2. Research Assistant, Department of Building and Real Estate, Faculty of Construction and Environment, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
(email: johnny.lin@polyu.edu.hk)

Ming-Fung Francis SIU^{3*}

3*. Assistant Professor, Corresponding author, Department of Building and Real Estate, Faculty of Construction and Environment, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong (email: francis.siu@polyu.edu.hk)

Yu-Hong WANG⁴

4. Associate Professor, Department of Civil and Environmental Engineering, Faculty of Construction and Environment, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong (email: ceyhwang@polyu.edu.hk)

Ping-Chuen Albert CHAN⁵

5. Head of Department, Chair Professor, Department of Building and Real Estate, Faculty of Construction and Environment, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong (email: bsachan@polyu.edu.hk)

Chi-Fai Daniel LAU⁶

6. Director, Able Engineering Company Limited, 155 Waterloo Road, Kowloon Tong, Kowloon, Hong Kong (email: daniellau@ableeng.com.hk)

ABSTRACT

In construction, the planners always estimate the work productivity and work efficiency of limited workers on site for delivering the projects on time. Time study is traditionally used for analysing the worker's performance to derive the productivity and efficiency. However, this manual method is too tedious and time-consuming. In Hong Kong, video cameras are always installed on construction sites to record site operations and workers' workflows. Site videos are always archived and provided solid evidence in case of construction site accidents. The videos captured tons of data which can be of better used for benchmarking worker productivity and work efficiency. Although the research endeavors advanced object recognition technology to automate the productivity analysis using modern computers, its applications of productivity analysis are highly limited. In this research, we proposed a novel approach for video-based productivity analysis by combining the use of posture recognition and deep neural networks. First, a prediction model using deep neural network and transfer learning is calibrated. Next, a knowledge base of posture information of specialty trades is developed by capturing actual postures of known specialty trades. Then, driven by the knowledge base, the workers of unknown trades are smartly classified by the computers using posture recognitions (instead of face recognition) and deep neural networks such that the work efficiency of particular trades can be determined. To demonstrate the application of the proposed method, a practical case study of a housing project in Hong Kong is used. The scope of the study is limited to one trade of steel bar bender and fixer. The computational effectiveness of using the new approach is reported. In conclusion, this new approach outperforms the traditional one in terms of result reliability and time efficiency.

INTRODUCTION

In construction, the project planners are struggled to improve the productivity of limited workforce on site so as to deliver the construction projects on time. In consideration of the work productivity and work efficiency of individual workers, worker allocation strategies can be made in the hope of improving the productivity of a project. For example, a worker, who may be suffered from low efficiency after working for few hours, should swap the activities with another one with high efficiency in the same trade such that the project progress will not be delayed.

The productivity of a worker can be measured using activity-oriented model (Thomas et al., 1990) as per Equation (1). In general, the “output” is defined as the work content being delivered such as the number of tons of steel being fixed, and the “input” is defined as the worker-hours that can be calculated by multiplying the number of deployed workers and their consumed working hours. For example, the productivity benchmarks are available in commercial market such as RS Means® (Murphy, 2019).

$$\text{Work productivity} = \frac{\text{Output (installed quantity)}}{\text{Worker-hour}} \quad (1)$$

The production rate is dependent on workers’ work efficiency (Thomas et al., 1990). The worker efficiency is recognised as the percentage of the work duration that a worker contributed to effective work and contributory work (rather than non-effective work). The production rate can be discounted by the work efficiency. For example, the work efficiency of oil and sand industry was benchmarked as 30% (CII, 2010, COAA, 2014). However, these yardsticks were generally provided at industry or project level. There is no existing technique for benchmarking the work efficiency at activity level for a particular project.

In Hong Kong, video cameras are always installed on construction sites to record site operations and workers’ workflows. Site videos are always archived and provided solid evidence in case of construction site accidents (Hong Kong Government, 1955). The videos captured tons of data which can be of better used for assessing worker productivity and work efficiency. Time study and work sampling techniques are always used for analysing these videos to derive worker productivity and work efficiency. However, this manual method is too tedious and time-consuming (Yates, 2014).

In recent years, the research endeavors advanced information technology to automate the productivity analysis using modern computers. For instance, the research proposed to use advanced face recognition technology to track the activities of workers. When applying this technology to analyse our collected site videos, we found that the results of face recognition are not reliable because the workers often cover their mouths, noses, and eyes to prevent physical hurts on site such that their essential facial features cannot be recognised by computers.

Postures, on the other hand, are a basic representation of workers’ activities at different time points. An activity can be represented by a collection of postures. Posture detection has been used in construction applications such as risk alert (Yu et al., 2012). Postures can be recognised using either non-vision-based or vision-based methods. For instance, wearable sensors can be used to track the movement and detect the postures of a worker. However, the workers may feel physically and mentally uncomfortable for wearing such tracking sensors. Image processing methods, as an alternative, are used to track the activities of workers.

In this research study, we first calibrated deep neural networks and transfer learning to classify the postures and works of the workers. Then, we developed the knowledge base for storing the information of predefined worker posture and work classification (effective, contributory, and non-effective) based on the site videos. Based on the results of posture detection, the work efficiency for measuring the productivity is defined in accordance with the detected postures of construction workers.

LITERATURE REVIEW

Research studies of benchmarking work efficiency on construction site were conducted for improving the effectiveness of utilizing the limited resources (Thomas et al., 1990). Most of the studies characterized the analytical relationships between work productivity and critical factors. For instance, Sonmez and Rowings James (1998) used a neural network to quantitatively evaluate the effects of multiple factors on concrete pouring productivity using a linear model. Tam et al. (2002) used a neural network with non-linear activation functions to model the non-linear relationship between the productivity of excavators and critical factors. Moradi et al. (2017) used hybrid system dynamics approach and discrete event simulation to identify the continuous context and discrete operational factors such as environmental condition, site condition, crew size, work quantity, material inventory, scheduled overtime, and construction technologies, and their effects on worker productivity. However, the data used to generate the analytical models is project-specific which may not be applicable to future projects (Halligan David et al., 1994). Moreover, this method is not designated for quantifying the work productivity of particular trades or individual workers.

To determine the work efficiency of workers on site, field sampling method such as work sampling technique was commonly used (Oglesby et al., 1989). Work sampling technique measures the effort of a worker contributed to a site activity. In general, the site work associated with a worker can be classified into “effective”, “contributory”, and “ineffective”. The worker effort can be measured by calculating the ratio of “effective work plus 0.25 times contributory work” over “the total number of observation” (Yates, 2016). When sampling the work, the manager, who is familiar with the job nature of a trade, track the workers’ work performance using manual observation. Undoubtedly, the site workers may be pressurized while the work classification by managers may be subjective and time-consuming.

Due to the advancement of computing power, activity analyses based on computer vision become feasible in recent years. The researchers focused on detecting, recognizing, and tracking any moving entity such as workers and equipment on site (Park et al., 2011). For example, Gong and Caldas (2011) integrated a number of vision-based object detection and tracking methods to automatically interpret construction actions information, such as work processes, cycle time, and delays. Park and Brilakis (2012) used HOG (Histogram of Oriented Gradient) features and k-NN (k-Nearest Neighbors) to detect construction workers in video frames. Kim et al. (2018) used transfer learning to detect construction equipment in images. Konstantinou and Brilakis (2018) used vision-based method to track construction workers from different views and perspectives for reducing the influence of site occlusion.

Nevertheless, Gong et al. (2011) realized that work productivity is associated with the working pose and gesture of site workers. They used background subtraction to extract human pose at each frame from near-field videos, and trained a neural network for classifying worker performance into three classes of effective, ineffective, and contributory work. However, the postures cannot be easily recognized because the background subtraction can only track moving objects based on the site videos and a posture can be associated with the workers of other trades.

NEW METHODOLOGY

This section is structured into three sub-sections: (i) establishment of prediction model using deep neural network and transfer learning to detect postures of site workers based on the collected site videos; (ii) development of knowledge base by defining the inputs and outputs of trades and postures; and (iii) classification of specialty trades and quantifications of work efficiency for productivity measurement.

Establishment of prediction model using deep neural network and transfer learning

A deep neural network (DNN) commonly consists of an input layer, multiple hidden layers, and an output layer.

- Input layer: $\chi = \{d_k\}$ denotes the input layer. The input layer loads and stores the data from a resized image frame.
- Hidden layer: $H_i = \{h_{i,j}\}$ denotes the hidden layer. $\{h_{i,j}\}$ is the j^{th} node or unit in H_i (i^{th} hidden layer). Convolutional layers, named as convolutional neural network (CNN), are often used as hidden layer for image processing.
- Output layer: $H_\infty = \{e_l\}$ denotes the output layer. The layer stores the data of posture classification and worker localization. Typically, the location of a worker in the images is defined by the coordinates of the corners of a rectangle bounding box.

When establishing a CNN model, several considerations should be noted. Firstly, the input image should be resized to a proper scale for efficient computation. Bottleneck features (Yu and Seltzer, 2011) have been used to compute features in order to improve the efficiency of deep learning. Secondly, filter shapes in the convolutional layers should be properly designed such that the right level of granularity can be determined so as to compute feature maps with the proper scale for a given dataset. Thirdly, the output layer should be designed by including the class labels (i.e., posture type) and bounding boxes of detected objects for posture recognition.

The proposed framework of deep neural networks for posture detection based on input images is shown in Figure 1. The input is a two-dimensional image. The depth of the input image is three when the image is an RGB image. The middle layers are CNNs (convolutional neural networks) which act as the image feature extractors. The output layer contains the information about the bounding boxes and posture labels of onsite workers. Other information, such as posture mask (i.e., silhouette) and trade label, can be added in future application.

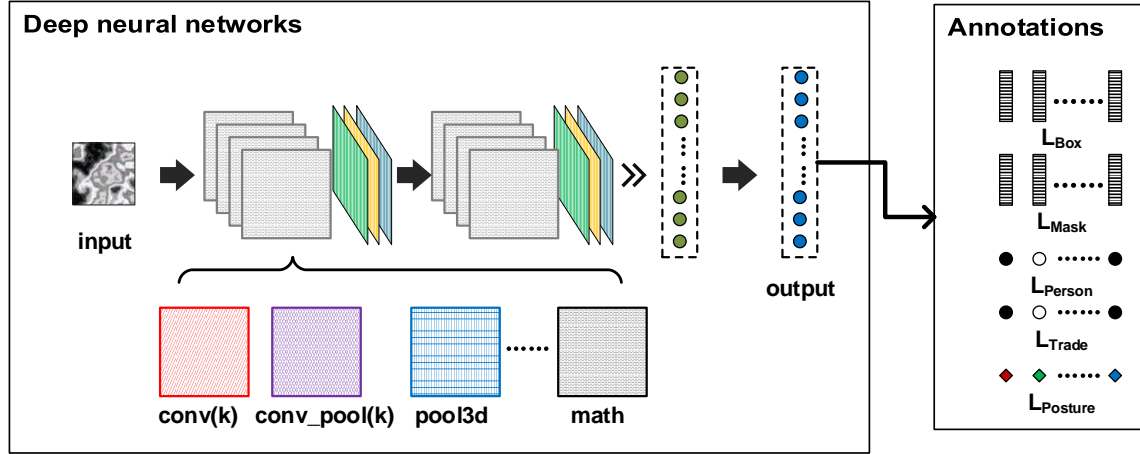


Figure 1: Proposed framework of posture detection using deep neural networks

There are existing techniques for improving the efficiency of network training for deep learning. For example, fully convolutional networks can be used to reduce the training time (Kim et al., 2018), dropout method (Srivastava et al., 2014) can be used in a fully-connected dense layer to reduce the training time and to prevent the data overfitting, transfer learning can be used to pertain certain parameters for calibrating the hidden convolutional layers such that training time can be reduced (Kolar et al., 2018, Kim et al., 2018).

In this research study, Faster-RCNN model (Ren et al., 2015) with ResNet structure (He et al., 2016) was selected to detect workers' postures (Huang et al., 2017). Transfer learning using pre-trained parameters given by MS COCO (Lin et al., 2014) was used to calibrate the convolutional networks since the size of our collected training sets was small.



To measure the precision of bounding box, Intersection over Union (IoU) is used. The overlapped area of a detected bounding box and its ground-truth bounding box is denoted as $A_D \cap A_{GT}$. The united area of a detected bounding box and its ground-truth bounding box is denoted as $A_D \cup A_{GT}$. The IoU is determined as the ratio of $A_D \cap A_{GT}$ to $A_D \cup A_{GT}$. When IoU exceeds a predefined value, i.e., 0.5, the detected object is assumed to be correct. As such, average precision (AP) is used to measure the accuracy of posture detection with respect to classes (postures).

Development of knowledge base by defining the inputs and outputs for trade and posture classifications

There are three steps to establish a knowledge base for storing training dataset of posture detection using site videos as shown in Table 1. Firstly, the videos of the targeted specialty trade are manually selected and trimmed. Secondly, the keyframes are extracted from selected videos using a uniform temporal sampling method (Chen and Wang, 2017). Finally, the bounding boxes and posture labels of workers are manually labeled from the keyframes.

Table 1. Knowledge base for storing the training dataset of posture detection

No.	Image region sample	Ground truth information "posture"<bounding box>
-----	---------------------	---

1		Crouch < 0.198958 0.825926 0.048958 0.094444 > Bend < 0.307292 0.533333 0.064583 0.131481 > Crouch < 0.562500 0.566204 0.042708 0.106481 >
2		Walk < 0.317448 0.283796 0.064062 0.210185 > Bend < 0.429948 0.498611 0.066146 0.187963 > Crouch < 0.268229 0.767593 0.051042 0.088889 > Crouch < 0.704427 0.490741 0.042188 0.129630 >

The locations of the site workers are detected from each frame based on bounding boxes of postures. Then, their postures can be classified in accordance with postures. For example, the postures of a bar bender and fixer when fixing the steel bars for a reinforced concrete wall can be classified as “stand”, “walk”, “bend”, “climb” and “crouch”. The primary postures (bend, crouch, stand, and walk) and the corresponding projection data are shown in Figure 2.

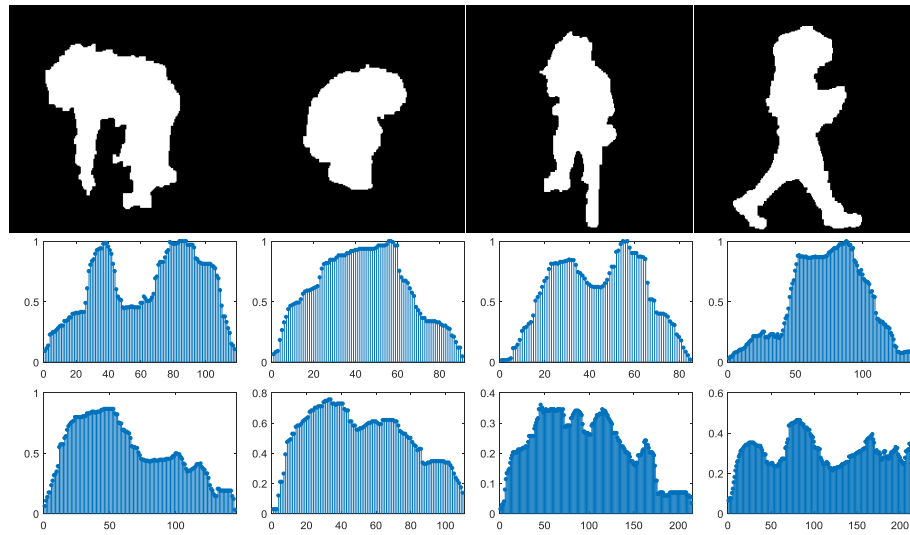


Figure 2: Four postures (bend, crouch, stand, walk) for fixing steel bars

Notably, deep learning method is used to detect onsite workers and recognise their postures simultaneously. Using a pre-established training dataset, a posture detection deep neural network model was trained for detecting the postures of steel bar benders and fixers. After that, the work productivity and efficiency can be analysed.

Classification of specialty trades of site workers and work efficiency for productivity measurement

On site, the workers of different trades may work in the same working zone. There is a need to classify the workers into different trades. The working trade can be classified using posture detection method, by recognizing the key poses of a worker of particular trade and analysing temporal distributions of the workers.

For instance, the classification rules of the steel workers were proposed as follows:

- The key poses of the steel fixing are observed, e.g., the workers stand and crouch with their

- hands reaching the steel bars.
- The repetition duration and frequency of the key poses exceed certain thresholds, e.g., one-third of the working time.

Since the work associated with a worker can be classified as “effective”, “contributory”, and “ineffective”, Equation (2) is proposed to quantify the effective work of bar bender and fixer in consideration of the time and the weights of different postures. The variables β_{absent} , β_{stand} , β_{walk} , β_{crouch} , β_{bend} , and β_{climb} are the regression coefficients which represent the weights (i.e., effectiveness) of postures when delivering bar bending and fixing activity on site.

$$\text{Effective work}_{\text{Steel fixing}} = \beta_{\text{absent}}t_{\text{absent}} + \beta_{\text{stand}}t_{\text{stand}} + \beta_{\text{walk}}t_{\text{walk}} + \beta_{\text{crouch}}t_{\text{crouch}} + \beta_{\text{bend}}t_{\text{bend}} + \beta_{\text{climb}}t_{\text{climb}} \quad (2)$$

Equation (3) shows the effectiveness of the postures for fixing the steel bars according to the experience of the site planners. b_1 and b_2 are the predefined lower and upper bounds of weight. Monte Carlo simulation is used to compute the work efficiency.

$$0 < \beta_{\text{walk}} < \beta_{\text{bend}} \equiv \beta_{\text{climb}} < \beta_{\text{crouch}} \equiv \beta_{\text{stand}} \leq 1 \\ b_1 \leq \beta_{\text{absent}} \leq b_2 \quad (3)$$

PRACTICAL CASE STUDY

To illustrate the method application of the proposed method, experiments were conducted for classifying the trade and postures associated with bar benders and fixers. A 20-minute video of steel fixing for a concrete wall was used.

Classifications of trades and postures

One hundred image frames with manually labeled ground truth were used to calibrate a pre-trained Faster-RCNN model with COCO dataset, and were used to evaluate the precision of posture detection. Figure 3 shows the results of posture detection in sample image frames.



Figure 3: Posture recognition results from twelve image frames

Table 2 shows the AP values of posture detection. The low AP values for posture detection of “bend”, “walk”, and “climb” are mainly caused by the limited number of instances used in training and testing dataset. Notably, the evaluation results can be potentially improved if the training and testing set is large (i.e., more than 100) and instances in different classes are more evenly distributed.

Table 2: Precision of posture detection in steel fixing

Precision metric	Classification					Bounding box localization
	Bend	Crouch	Stand	Walk	Climb	
Instances in ground truth in training dataset	27	115	116	28	1	287
Instances in ground truth in testing dataset	15	45	193	4	2	259
AP@0.5IoU	0.30	0.82	0.98	0	0	0.43

Classification of worker productivity and work efficiency

Four steelworkers were detected in this video. The postures of these four workers were detected in each second. The results of posture detection in 1,200 seconds (20 minutes) is shown in Figure 4.

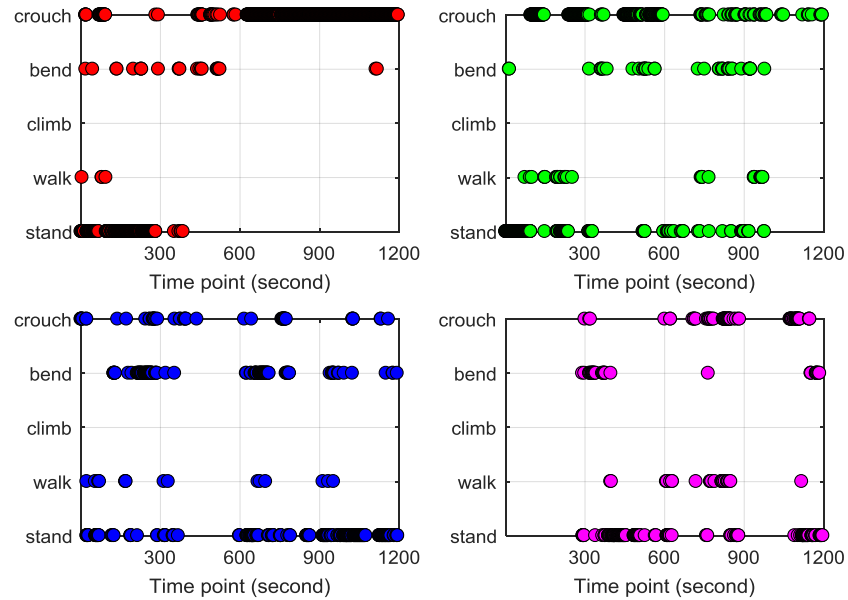


Figure 4: Detected postures of 4 steelworkers in a 20-minute video

Figure 5 shows the efficiency of workers in every five minutes to serve as an alternative for five-minute work sampling.

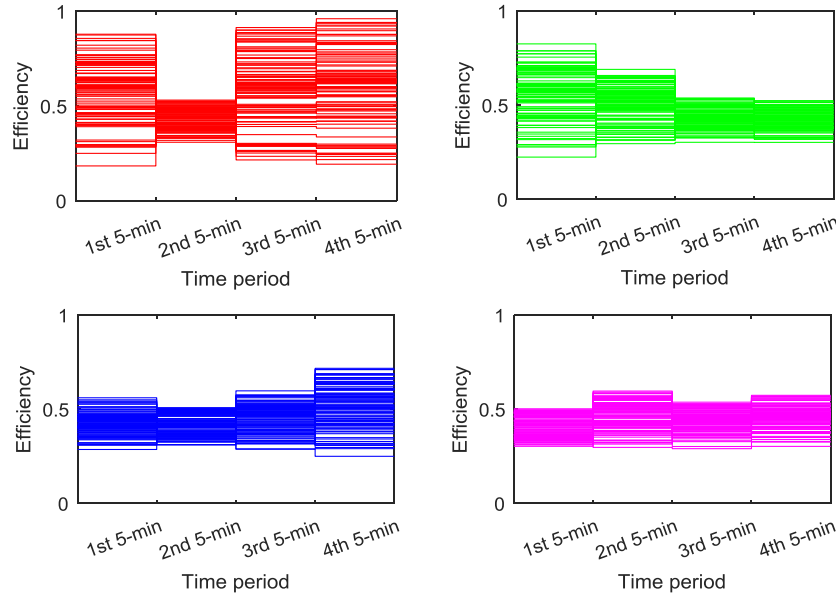


Figure 5: Work efficiency of four steelworkers in a 20-minute video

The average efficiency is also computed using Monte Carlo simulation for 100 iterations. Figure 6 shows the simulated results. The results show that the first worker has the highest working efficiency, and the fourth worker has the lowest working efficiency.

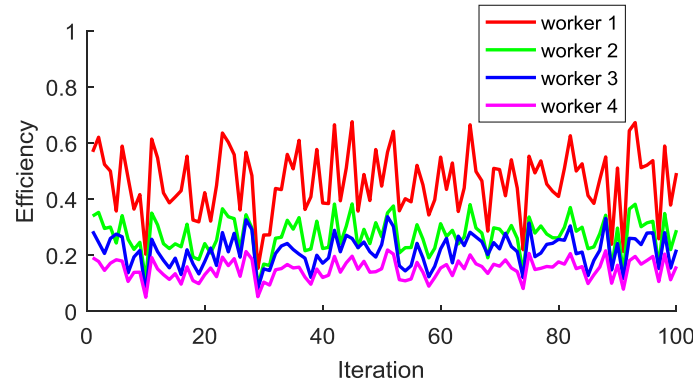


Figure 6. Average efficiency of four steelworkers in a 20 minutes video using Monte Carlo simulation

To validate the experiment results, site planners with more than 20 years of site experience were consulted. He confirmed that, when fixing steel bars, the “bend” and “crouch” postures can be harmful to human health, and excessive “walk” posture can be caused by the improper placement of the materials on site. The proposed method can potentially improve the work productivity and efficiency in real time as per the following measures:

- When excessive “walk” is detected, the managers should assure whether the materials are stored in a proper place for delivering the activities.
- When “climb” is detected, the managers should pay extra attention to such workers who may be in danger.
- When excessive “bend” or “crouch” is detected, the managers should swap his work with other workers in the same trade to balance the workload and prevent physical hurt.

CONCLUSIONS

In construction practice, manual methods of benchmarking work productivity and worker efficiency are not commonly used by planners because it is too time-consuming and tedious. As such, this research study proposed an automatic method for determining work productivity by establishing a knowledge base of worker postures and work classes based on site videos using modern computing algorithms coupled with computer power.

The experiments show that deep learning method has potential for detecting the workers' postures accurately, providing that the training data is sufficient. Based on the transfer learning algorithm, a deep neural network is trained with a relatively small training dataset. 89% of accuracy was achieved. Practical case study was used to demonstrate the classification the trade and work of steel bar bender and fixers. The work trade and postures of the four steel workers in the video frames were successfully detected. This proved that the proposed method can be used as an alternative of work sampling to compute the worker efficiency automatically driven by the knowledge base of workers' postures and work classes.

In future, the proposed method can be used for benchmarking the worker productivity and work efficiency by correlating the factors such as weather. Real-time resource allocation and optimization on the basis of individual workers can be possible. For example, based on the detection results, the planners should place the site materials in proper place such that excessive "walk" can be avoided and should swap the workload among the individual workers in the same trade such that excessive "bend" and "crouch" can be avoided.

ACKNOWLEDGMENT

The authors would like to thank The Hong Kong Polytechnic University (Project title: Data-driven resource planning for building works in Hong Kong; PI: Corresponding author), and Able Engineering Company Limited (Project title: Development of smart resource-scheduling system for improving resource productivity performance; PI: Corresponding author) for financially supporting this research work. The authors would also like to thank Lawrence Tse and Mark Mak from Able Engineering Company Limited for the assistance during data collection.

REFERENCES

- Chen, L. & Wang, Y. (2017). "Automatic key frame extraction in continuous videos from construction monitoring by using color, texture, and gradient features". *Automation in Construction*, 81, 355-368.
- CII. (2010). "Guide to Activity Analysis". The University of Texas at Austin.
- COAA. (2014). "COAA workface planning rules". Edmonton, AB, Canada. 2014
- Gong, J. & Caldas, C. H. (2011). "An object recognition, tracking, and contextual reasoning-based video interpretation method for rapid productivity analysis of construction operations". *Automation in Construction*, 20(8), 1211-1226.
- Gong, J., Caldas, C. H. & Gordon, C. (2011). "Learning and classifying actions of construction workers and equipment using Bag-of-Video-Feature-Words and Bayesian network models". *Advanced Engineering Informatics*, 25(4), 771-782.
- Halligan David, W., Demsetz Laura, A., Brown James, D. & Pace Clark, B. (1994). "Action - Response Model and Loss of Productivity in Construction". *Journal of Construction Engineering and Management*, 120(1), 47-64.
- He, K., Zhang, X., Ren, S. & Sun, J. (2016). "Deep Residual Learning for Image Recognition". *The*

- IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Hong Kong Government (1995). "Buildings Ordinance". Laws of Hong Kong. Hong Kong. 1955
- Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Alireza Fathi, I. F., Wojna, Z., Song, Y., Guadarrama, S. & Murphy, K. (2017). "Speed/accuracy trade-offs for modern convolutional object detectors". *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Kim, H., Kim, H., Hong, Y. W. & Byun, H. (2018). "Detecting Construction Equipment Using a Region-Based Fully Convolutional Network and Transfer Learning". *Journal of Computing in Civil Engineering, ASCE*, 32(2), 04017082.
- Kolar, Z., Chen, H. & Luo, X. (2018). "Transfer learning and deep convolutional neural networks for safety guardrail detection in 2D images". *Automation in Construction*, 89, 58-70.
- Konstantinou, E. & Brilakis, I. (2018). "Matching Construction Workers across Views for Automated 3D Vision Tracking On-Site". *Journal of Construction Engineering and Management*, 144(7), 04018061.
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Doll R, P. & Zitnick, C. L. "Microsoft COCO: Common Objects in Context". 2014 Cham. Springer International Publishing, 740-755.
- Moradi, S., Nasirzadeh, F. & Golkhoo, F. (2017). "Modeling Labor Productivity in Construction Projects using Hybrid SD-DES Approach". *Scientia Iranica*, 24(6), 2752-2761.
- Murphy, J. D. (2019). *RS Means Labor Rates for the Construction Industry 2019*, RS Means Company.
- Oglesby, C. H., Parker, H. W. & Howell, G. A. (1989). *Productivity improvement in construction*, McGraw-Hill College.
- Park, M.-W. & Brilakis, I. (2012). "Construction worker detection in video frames for initializing vision trackers". *Automation in Construction*, 28, 15-25.
- Park, M.-W., Makhmalbaf, A. & Brilakis, I. (2011). "Comparative study of vision tracking methods for tracking of construction site resources". *Automation in Construction*, 20(7), 905-915.
- Ren, S., He, K., Girshick, R. & Sun, J. (2015). "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks". *Advances in Neural Information Processing Systems* 28. Curran Associates, Inc.
- Sonmez, R. & Rowings James, E. (1998). "Construction Labor Productivity Modeling with Neural Networks". *Journal of Construction Engineering and Management*, 124(6), 498-504.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. (2014). "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". *Journal of Machine Learning Research*, 15, 1929-1958.
- Tam, C. M., Tong, T. K. L. & Tse, S. L. (2002). "Artificial neural networks model for predicting excavator productivity". *Engineering, Construction and Architectural Management*, 9(5-6), 446-452.
- Thomas, H. R., Maloney William, F., Horner, R. M. W., Smith Gary, R., Handa Vir, K. & Sanders Steve, R. (1990). "Modeling Construction Labor Productivity". *Journal of Construction Engineering and Management*, 116(4), 705-726.
- Yates, J. (2014). "Productivity Improvement for Construction and Engineering: Implementing Programs that Save Money and Time".
- Yates, J. K. (2016). "Productivity Improvement Data Analysis Techniques". *Productivity Improvement for Construction and Engineering*.
- Yu, D. & Seltzer, M. L. (2011). "Improved bottleneck features using pretrained deep neural networks". *12th Annual Conference of the International Speech Communication Association*. Florence, Italy.
- Yu, M., Rhuma, A., Naqvi, S. M., Wang, L. & Chambers, J. (2012). "A Posture Recognition-Based Fall Detection System for Monitoring an Elderly Person in a Smart Home Environment".

Trans. Info. Tech. Biomed., 16(6), 1274-1286.