Deep Learning Detection for Real-time Construction Machine Checking

B. Xiao[a] and S.-C. Kang[a]

[a]Department of Civil & Environmental Engineering, University of Alberta, Canada
E-mail: bxiao2@ualberta.ca, sckang@ualberta.ca

Abstract –

Construction sites require many efforts to be well organized due to the complicated tasks and various construction machines. Recently, computer vision technology has gained success in the construction research field. By deploying single or multiple cameras, we can extract construction information from videos and then help the project manager to understand what happened in real-time. This paper presents a method of checking construction machine status automatically by cameras. We assume that the camera is installed in a high position, which provides clear views for the whole site. This research focuses on extracting the machine information from videos, comparing with construction schedule and feedbacking to project manager for further decision making. In the preliminary stage, a deep learning detector has been employed for detecting active construction machines. Meanwhile, a construction image dataset has been organized for training deep learning models precisely and robustly. This dataset also helps to promote this method to generalized construction scenarios. Comparing the number of active machines of video and the expected number of machines from the schedule, the project manager will get real-time feedback and alerts when missing construction machines. In the future steps, we will develop a method to understand construction activities from videos and highlight important activities automatically. Once the reliable method has been developed, it will benefit the project manager to monitor construction sites from an easier way.

Keywords –

Construction Machine; Computer Vision; Deep Learning

1 Introduction

Construction machines have participated the most of construction activities, and they are essential for any construction processes [1]. Properly using and monitoring of machines contributes to the economy, speed, and quality of projects. It is difficult to track the working machine status, especially in a large construction site. The traditional way of monitoring construction machine is employing a worker to record the information and report to project manager each day. Manually checking of machines is tedious, time-consuming, error-prone, and not safe [2].

Furthermore, the recorded information cannot be feedbacked to the project manager for making a real-time decision. Recently, computer vision technology has succeeded in medical imaging, human-computer interaction and autonomous driving [3]. This technology also shows great potential in construction management. In this paper, we present a method of automatically checking construction machines from the video stream.

By deploying the camera in a high position of construction sites, such as the crane boom and existing high-rise buildings, the camera captures the entire footprint of construction sites. In this research, we have developed a method to extract machine active status from the video stream and comparing the number of active machines with the project schedule. The deep learning detection algorithm YOLOv3[4] has been employed to extract machine category and position from videos. A construction image dataset has been built with the purpose of training deep learning models. Until now, 5,000 images have been collected and manually labeled for construction object detection. This dataset will be kept developing in order to train deep learning detectors and make these methods generalized in more construction scenarios. Then the expected number of active machines can be extracted from project schedules. The comparison results will be summarized in the active chart and feedback to project managers in real-time.

This method benefits project managers and enables them to monitor construction machine status more

directly. If any machine is missing when compared with the planned schedule, the project manager will be notified immediately and they can coordination among construction crews to figure out the problem. The machine active information can be recorded every day for documentation purpose. In the future, we will focus on how to understand construction activates from videos, highlight essential moments for project managers, and help them control the project schedules. A user interface will be designed to display essential information and enable users to search their interested information.

2 Literature Review

In this section, recent studies on object detection methods are reviewed. Then, applications of computer vision technology in construction management are presented. At the end, the evaluation criteria which are widely used has been introduced.

2.1 Development of Object Detection

Object detection refers to detecting instances of semantic objects of certain classes from digital images and videos [5]. The sliding-window paradigm has a successful history in classic object detection, which applies classifier on dense image grid. Viola and Jones [6] have introduced boosted detectors into face recognition. The HOG (Histograms of oriented gradients) [7] provides effective features to pedestrian detection. DPMs [8] are based on part-based deformable models and had achieved the best performance on PASCAL VOC dataset [9] before the introduction of deep learning models.

Recently, the Convolutional Neural Networks (CNNs) methods become the dominant in object detection. The CNNs detector has two categories, which are two-stage detector and one-stage detector. In two-stage detectors, the first stage generates a bunch of candidate regions which may contains expected instances, and the second stage employs classifier to filter all instances into categories. R-CNN [10] has adopted AlexNet [11] and SVM [12] into the second stage and achieved higher performance on VOC dataset. R-CNN has been enhanced in boost precision and speed recently [13], [14].One-stage detectors do not have the first stage of generating candidate regions.

One-stage detectors, such as YOLO and SSD [15], have excellent performance in speed. Recent researches have shown the two-stage detectors perform better in accuracy than one-stage detectors, while one-stage detectors are much fast. Although, recent work shows that two-stage detectors can be faster by reducing the input image resolution and proposal regions [16], one-stage detectors are the better choice for real-time applications. All object detection methods have faced a large class imbalance problem in training, which means these detectors evaluated thousands of candidate regions per image but only a few of them contain objects. The focal loss [17] has been introduced to figure out this problem and allow us training deep learning detectors effectively.

2.2 Applications in Construction

There are many researches and applications that have been developed in construction management with the assisting of computer vision technology. The applications can be categorized as productivity estimation and safety control. For productivity estimation, Weerasinghe and Ruwanpura [18] tracked construction resources with the purpose of reducing waste. Rezazadeh and Brenda [2] have developed an automated method to detect dirt-loading cycles in earth moving tasks. Xiao and Zhu [19] have compared fifteen tracking algorithms in construction scenarios in order to identify the most efficient algorithm. Yang et al. [20] have employed Gaussian background subtraction to detect crane jibs and then make sure the crane operates in good environment.

Construction has become one of the most unsafe industries because of the high risks exist. According to previous study [21], there was more than 2500 annual deaths accompanied in construction sector from 1994 to 2014 in China. Computer vision technology is able to help safety management from comprehensive ways [22]. Han and Lee [23] have developed a system to detect workers and machines. This system protect workers from potential collisions from video streams. Deploying multiple cameras in construction sites can reconstruction 3D clouds of workers, which tracks worker motion precisely [24]

2.3 Evaluation Criteria

Evaluation criteria is important to estimate detectors. Precision and recall are the most common criteria for object detection [25]. Precision refers to the fraction of correct instances among all retrieved instances, while recall means the fraction of correct instances among total ground truth instances. Figure 1. shows the definition of True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN). The precision and recall can be then defined as follows.

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

Mean average precision (mAP) is another metric that measures the performance of CNN detectors [26]. mAP is the average of maximum precisions at different recalls. The average precision can be calculate as the

average of precisions in 11 recall levels, which from 0.0 to 1. The formula has been defined as follow. The mean average precision is the average of AP in each object class. In this study, we will use mAP to evaluate the deep learning detector.

$$AP = \frac{1}{11}\sum_{r\in\{0.0,0.1,\ldots,1\}} AP_r \quad (3)$$

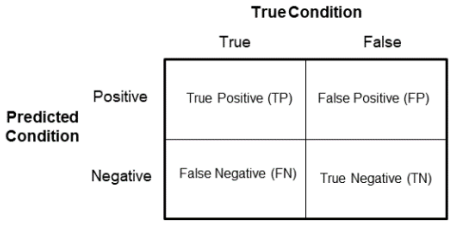

Figure 1. Definition of true positive, false positive, false negative, and true negative

3 Methodology

Figure 2. illustrates the overview of the methodology of machine status checking pipeline. There are four main parts of this method, which are building the dataset, training YOLOv3, detection and visualization, and feedback. In this section, we will introduce each part in details.

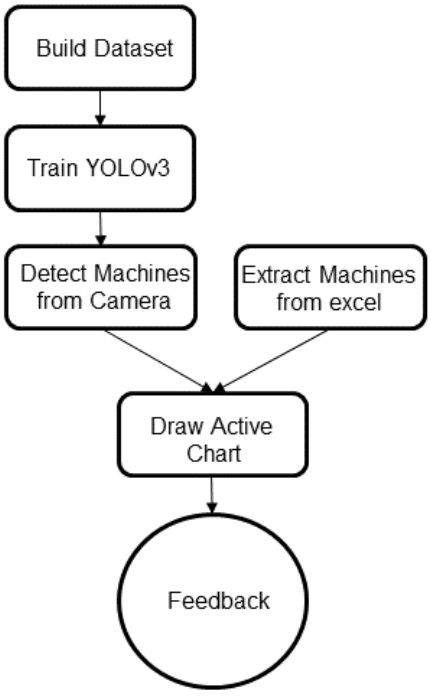

Figure 2. Overview of methodology

3.1 Build Dataset

The biggest challenge for deep learning detection is the limited data. The sufficiently large number of data helps the deep learning model robust and generalized. Since the proposed method will employ the deep learning detector and there not exists construction detection dataset, we have worked on building a new construction image dataset, which contains common construction machines. Over 200 construction videos, which come from YouTube channels, real construction projects, and other datasets, have been collected. 5000 images have been extracted and manually labeled to build this dataset. The entire dataset is divided into training set and testing set, while 80% of images in training set and 20% of images in testing set. In this dataset, there are four types of construction equipment, which are truck, excavator, loader, and backhoe. It shows that 22% of instances in our dataset contain have the size smaller than 5% of the whole image, while 25% of instances have the larger size than the 60% of the entire image. This distribution shows our dataset has both close-up equipment images and bird-view images. A large number of images helps to train deep learning models and generalize into common construction scenarios.

3.2 Training YOLOv3

YOLOv3 is a one-stage state-of-art detector with extremely fast speed. YOLOv3 has shown excellent in COCO dataset [26] with the mAP of 0.553. In this study, the image input size is 416x416 and this algorithm can process 30 images in one second. Compared with some two-stage detectors, the performance of YOLOv3 is slightly low, but the speed is much faster and that is important for real-time applications. The construction detection dataset from the previous step is used for training YOLOv3, which takes 12 hours for the training process. The mAP of YOLOv3 on our testing set is 0.87 from an overall view, where the AP is 0.71 in the truck category, 0.93 in excavator category, 0.91 in loader category, and 0.93 in backhoe category. The validation result shows YOLOv3 fits well in our dataset and could be used in other construction videos.

3.3 Detection and Visualization

This part refers to three steps of methodology in Figure 2, which are detecting machines from the camera, extracting machines from excel and drawing the active chart. In the detection stage, the image stream captured from cameras will be put into YOLOv3 model in real time. The detected images with bounding box will show to users for visualizing the object detection (Figure 3). The detection performance directly effects the overall performance of the entire system. The detected

information, which includes the number of each category of machines, sent to the next step for drawing the active chart.

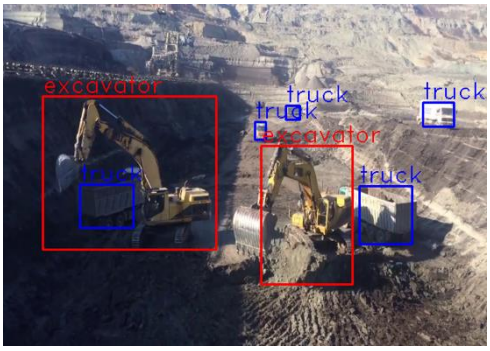

Figure 3. An example of detected images

The excel file of project schedule has been parsed to extract the machine numbers during this time. The information extracted from an excel schedule can be specific to one hour by one hour. The active chart will be drawn to visualize the number of machines active from the camera and construction schedule. Figure 4 shows an example of active chart, which is monitoring excavators and trucks. In the active chart, the jet colormap represents the number of machines, and the horizontal axis represents the time.

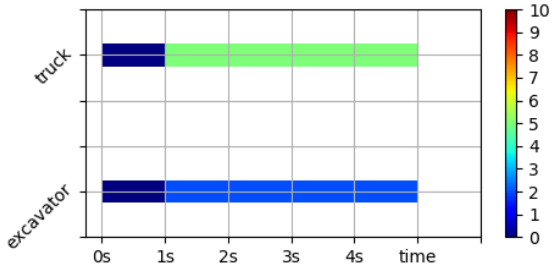

Figure 4. Active chart of describing machine amount

3.4 Feedback

The active chart shows the real-time machine status in construction sites. In each hour or few minutes, the checking system summarizes the active machine chart in this term and saves in local files. The system will notify users with a short sentence, such as "2 trucks and 1 excavator are missed at 14:00pm", when there is a mismatch of detected and scheduled information. The historical information of missing equipment will also be recorded including the timestamp, machine category, missing numbers to local files. This feedback process supports the project manager to document construction activities and active machine status. These files saved with original videos for further purpose. This system expected to help project managers generate construction logs for documentation purpose.

4 Results and Discussion

The construction machine checking system has implemented by Python and C language. All training and detection process was running on an NVIDIA 1080Ti GPU. The testing video has the duration of one hour, which recorded the earthmoving activity. The monitoring object is excavator and trucks. The expected machine number has been manually set up in the excel file, and the number changes every ten minutes. The detection result updated 25 times in each second and the active chart updated by seconds.

To evaluate our system, we use success rate (SR) to describe the performance. Construction sites are slowly moving, and we decided to check the machine numbers in every minute. If the detected machine number is the same as the ground truth from this frame and the feedback information is correct, it will give a positive sign in this minute. Otherwise, it will give a negative sign in this minute. The SR is calculated as Equation 4. The test precision is 95%, which means the machine checking system succeeded 57 times in one-hour test video.

$$SR = \frac{\sum Positive\ Sign}{60} \quad (4)$$

In this study, it found that deep learning detection is robust when well-developed image dataset provided. In computer vision community, there exists some detection dataset such as VOC and COCO. These dataset helps researchers to evaluate new algorithms and applications. It is necessary to build a public image dataset in construction research field, which will benefit the whole community. Since deep learning methods have huge potential in the construction automation field. Detecting all construction objects from images allows us to understand what happened in sites. For project managers, filtering useful information and visualize help them to monitor construction activities and decide in real-time.

Visualization is another concern in construction management. Effective visualization improves the sites communication and help experts understand what happened in construction sites. Since construction sites are always disordered, managers cannot extract useful information directly even with the assistance of cameras, visualized information provided key information to managers to support real-time decision making. Visualization in construction provides an efficient way to training junior workers and engineers.

5 Conclusion

This paper presents a system of checking construction machines automatically from video streams and construction schedule. Cameras deployed in construction sites provide clear views for the whole site and activities. The machine checking system extracts machine information from videos by using deep learning detector YOLOv3. Then it compares detected results and expected number from schedule excel file to feedback to project manager for further decision making. In order to train a general model, a construction image dataset has been built in this study.

In the future steps, the authors will work on developing a method to understand construction activities from videos and highlight important activities automatically. In addition, the authors will keep expanding the construction image dataset.

References

[1] Nunnally, Stephens W. Construction methods and management. Upper Saddle River, NJ: Pearson Prentice Hall, 2007.

[2] Rezazadeh Azar, Ehsan, Sven Dickinson, and Brenda McCabe. "Server-customer interaction tracker: computer vision–based system to estimate dirt-loading cycles." Journal of Construction Engineering and Management 139, no. 7 (2012): 785-794.

[3] Szeliski, Richard. Computer vision: algorithms and applications. Springer Science & Business Media, 2010.

[4] Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." arXiv preprint arXiv:1804.02767 (2018).

[5] Hjelmås, Erik, and Boon Kee Low. "Face detection: A survey." Computer vision and image understanding 83, no. 3 (2001): 236-274.

[6] Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, vol. 1, pp. I-I. IEEE, 2001.

[7] Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 1, pp. 886-893. IEEE, 2005.

[8] Felzenszwalb, Pedro F., Ross B. Girshick, and David McAllester. "Cascade object detection with deformable part models." In Computer vision and pattern recognition (CVPR), 2010 IEEE conference on, pp. 2241-2248. IEEE, 2010.

[9] Everingham, M., Van Gool, L., Williams, C.K., Winn, J. and Zisserman, A., 2010. The pascal visual object classes (voc) challenge. International journal of computer vision, 88(2), pp.303-338.

[10] Girshick, R., Donahue, J., Darrell, T. and Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580-587).

[11] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." In *Advances in neural information processing systems*, pp. 1097-1105. 2012.

[12] Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." *Machine learning* 20, no. 3 (1995): 273-297.

[13] Girshick, Ross. "Fast r-cnn." In Proceedings of the IEEE international conference on computer vision, pp. 1440-1448. 2015.

[14] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Spatial pyramid pooling in deep convolutional networks for visual recognition." In *European conference on computer vision*, pp. 346-361. Springer, Cham, 2014.

[15] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y. and Berg, A.C., 2016, October. Ssd: Single shot multibox detector. In *European conference on computer vision* (pp. 21-37). Springer, Cham.

[16] Huang, Jonathan, Vivek Rathod, Chen Sun, Menglong Zhu, Anoop Korattikara, Alireza Fathi, Ian Fischer et al. "Speed/accuracy trade-offs for modern convolutional object detectors." In *IEEE CVPR*, vol. 4. 2017.

[17] Lin, T.Y., Goyal, P., Girshick, R., He, K. and Dollár, P., 2018. Focal loss for dense object detection. *IEEE transactions on pattern analysis and machine intelligence*.

[18] Weerasinghe, IP Tharindu, and Janaka Y. Ruwanpura. "Automated data acquisition system to assess construction worker performance." In Construction Research Congress 2009: Building a Sustainable Future, pp. 61-70. 2009.

[19] Xiao, Bo, and Zhenhua Zhu. "Two-Dimensional Visual Tracking in Construction Scenarios: A Comparative Study." Journal of Computing in Civil Engineering 32, no. 3 (2018): 04018006.

[20] Yang, Jun, Patricio Vela, Jochen Teizer, and Zhongke Shi. "Vision-based tower crane tracking for understanding construction activity." Journal of Computing in Civil Engineering 28, no. 1 (2012):

103-112.
[21] He, Xueqiu, and Li Song. "Status and future taqsks of coal mining safety in China." Safety Science 50.4 (2012): 894-898.
[22] Guo, Hongling, Yantao Yu, and Martin Skitmore. "Visualization technology-based construction safety management: A review." Automation in Construction 73 (2017): 135-144.
[23] Han, SangUk, and SangHyun Lee. "A vision-based motion capture and recognition framework for behavior-based safety management." Automation in Construction 35 (2013): 131-141.
[24] Brilakis, Ioannis, Man-Woo Park, and Gauri Jog. "Automated vision tracking of project related entities." Advanced Engineering Informatics 25.4 (2011): 713-724.
[25] Davis, Jesse, and Mark Goadrich. "The relationship between Precision-Recall and ROC curves." In Proceedings of the 23rd international conference on Machine learning, pp. 233-240. ACM, 2006.
[26] Lin, Tsung-Yi, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. "Microsoft coco: Common objects in context." In European conference on computer vision, pp. 740-755. Springer, Cham, 2014.