Trajectory Prediction of Mobile Construction Resources Toward Pro-active Struck-by Hazard Detection

D. Kim[a], M. Liu[a], S. Lee[a], and V. R. Kamat[a]

[a]Department of Civil and Environmental Engineering, University of Michigan, MI, US.
E-mail: daeho@umich.edu, meiyin@umich.edu, shdpm@umich.edu, vkamat@umich.edu

Abstract –
In construction, unanticipated struck-by hazards often arise, which have resulted in a significant number of construction fatalities. To address this problem, many studies have attempted to automate proximity monitoring and struck-by hazard detection using various technologies, such as wireless sensors and computer vision methods. While this technology focuses on understanding what is happening as hazards arise, it is not equipped to detect future hazards. In impending situations, detecting current hazards may not provide enough time for workers to take evasive actions. To address this challenge this study develops a trajectory prediction model for mobile construction resources. Specifically, this study conducts hyper-parameter tuning of a deep neural network, called Social Generative Adversarial Network to develop a prediction model capable of predicting more than five seconds. Further, a test on a real construction operations data follows to validate developed models' trajectory prediction accuracy. As a result, a developed model could achieve promising accuracy: the average displacement error and the final displacement error were 0.78 and 1.27 meters, respectively. The trajectory prediction allows for detecting future hazards, which will support pro-active intervention in hazardous situations. It will ultimately contribute to promoting a safer working environment for construction workers.

Keywords –
Struck-by hazard; Pro-active intervention; Trajectory prediction; Deep neural network; Hyper-parameter tuning

1 Introduction

In construction, mainly due to unstructured and limited workspaces, unanticipated struck-by hazards involving mobile vehicle or equipment often arise, contributing to the significant number of construction fatalities [1]. According to *The Center for Construction Research and Training, United States*, from 2011 to 2015, total 925 struck-by fatalities were reported from construction [2]. The figure accounted for 24% of overall struck-by fatalities in the U.S. and was unmatched by other U.S. industries [2]. Notably, the number of struck-by fatalities rose 34% from 2010 (N=121) to 2015 (N=162) [2].

A critical element of construction safety management is "a proactive, ongoing process to recognize hazards that are present or that could have been anticipated" [3]. However, such continuous monitoring has not been viable in practice as manual observation and inspection is notoriously time-consuming, labor-intensive, and costly [4].

A major research area for this issue is attuned to automating object localization, proximity monitoring, and accordingly struck-by hazard detection. Prior research leveraged various technologies—such as wireless sensors [5-9] and computer vision methods [1, 10-11]—and made a great progress on automation of struck-by hazard detection. It is expected that the successful deployment of such technologies will allow for prompt feedback to involved workers, thereby reducing the chance of an impending collision [1, 5, 10].

However, there remains a critical challenge that has not been tackled yet: how to recognize not only current hazards but also the ones that will be present in the near future for pro-active intervention. All prior works using wireless sensors and computer vision are limited to understand what is happening. That is, these technologies are only capable of detecting current hazards because they depend on current locations of entities of interest. In many cases, however, letting a worker know "now you are in a danger" may not provide enough time for him/her to take a proper evasive action. Therefore, predicting what will happen (i.e., knowing future position of entities and detecting future hazards) is critical in the prevention of potential accidents.

As a preliminary study to address this challenge, this research examines the potential of trajectory prediction for mobile construction resources. To this end, this study develops a trajectory prediction model through hyper-parameter tuning of a deep neural network (DNN), called Social Generative Adversarial Network (GAN) [12], and conducts test on a real construction operations data to evaluate the developed model's prediction accuracy.

2 Previous Works

Considerable research efforts have been made to automate the struck-by hazard detection in construction. Some studies applied wireless sensors—such as Radio Frequency Identification (RFID) [5-6], Magnetic Field (MF) [7], Global Positioning System (GPS) [8], and Bluetooth Low Energy (BLE) [9]—to instantly detect hazardous proximity between entities of interest. On the other hand, other studies applied deep neural networks (DNNs)-based object detection framework—such as Faster R-CNN [13], R-FCN [14], and YOLO-V3 [1]—for continuous object localization and proximity monitoring.

The previous works have made large strides in automating struck-by hazard detection. However, trajectory prediction and accordingly future hazard detection have not been tackled yet. All prior works using wireless sensors and computer vision are limited to detect current hazards. To provide enough time for workers to take a prompt evasive action, to predict what will happen, namely recognizing future position of entities and detecting future hazards, is required.

One possible solution that can address the above issue is to enable trajectory prediction, which stands for a task to predict a target's future trajectories (a set of future positions) by observing the target's moving pattern. Recently, the trajectory prediction has made a great progress with the advancement of DNNs, such as Long-Short Term Memory (LSTM) [15], Gated Recurrent Unit (GRU) [16], and social pooling layers [17], and Generative Adversarial Network (GAN) [12]. Alahi et al. 2016 [17] first presented social pooling layers-embedded LSTM architecture (called Social LSTM), which showed remarkable progress in trajectory prediction. This work demonstrated the Social LSTM can learn not only each entity's moving pattern, but also social behaviour of human in crowded settings (e.g., collision avoidance). The interconnected use of individual and social features in turn showed a great performance in trajectory prediction: predicted trajectories by the Social LSTM only had 0.72 meter displacement error on average, compared to the ground truth trajectories.

Encouraged by this progress, Gupta et al. 2018 [12] more improved the Social LSTM [17] by using GAN. This work developed unique generator and discriminator by integrating LSTM encoder-decoder and social pooling layers (Figure 1, please refer to Gupta et al. 2018 [12] for detailed information). Consequently, the strict supervision by the discriminator successfully improved the model's prediction performance: the displacement error on average was 0.58 meters.

Despite the promise, applying the trajectory prediction DNN (i.e., Social GAN) to our problem involves another challenge: how to modify the original network so that it can predict longer time-steps. Note that the published Social GAN model has 2.64 s prediction length. To provide a worker in a danger with enough time for evasive action, longer prediction and accordingly more early notice are needed.

3 Research Objective and Framework

With this background, this study conducts hyper-parameter tuning of the Social GAN [12] to develop a trajectory prediction model for mobile construction resources. In essence, the longer prediction is needed for more pro-active hazard detection. This study sets five seconds as the target to predict with the assumption that it would be enough for workers to take prompt evasive actions. Further, tests on real construction operations data are conducted so as to demonstrate the developed model's potential in real-world applications. This study follows the below framework to achieve these aims (Figure 1).

- Data collection: for the purpose of hyper-parameter tuning, ETH [18] and UCY [19] dataset widely used for trajectory prediction are used. In addition, a real construction operations data that captures interactions between construction resources is collected for the test purpose.
- Hyper-parameter tuning: training the Social GAN [12] with multiple hyper-parameter scenarios is conducted to develop trajectory prediction models capable of predicting more than 5 seconds.
- Test on a real construction operations data: the trained models then are tested on a real construction operations data for evaluation.

4 Data Collection

The more extensive data is used for training, the higher performance of a model can be reached. For the hyper-parameter tuning purpose, this study thus benchmarked two sets of human trajectory data, ETH [18] and UCY [19], which are the most widely used dataset in trajectory prediction studies [12,17]. In total, the two datasets captures four different crowded scenes and contains 1,536 human trajectories. The trajectories reflect various human-human interactions, including (i) crossing each other; (ii) collision avoidance, (iii) group forming; and (iv) dispersing [17] (A in Figure 1).

In addition, this study collected a real construction operations data for the purpose of test. UAV captured construction site videos were collected. Of these, the total of 916 sequential frames were sampled that captures interactions between a worker, an excavator, and a wheel loader (B in Figure 1). Each trajectory (i.e., a set of x-y coordinates) of the three entities were manually annotated over the whole frames and a complete inspection ensured the validity of the annotations.

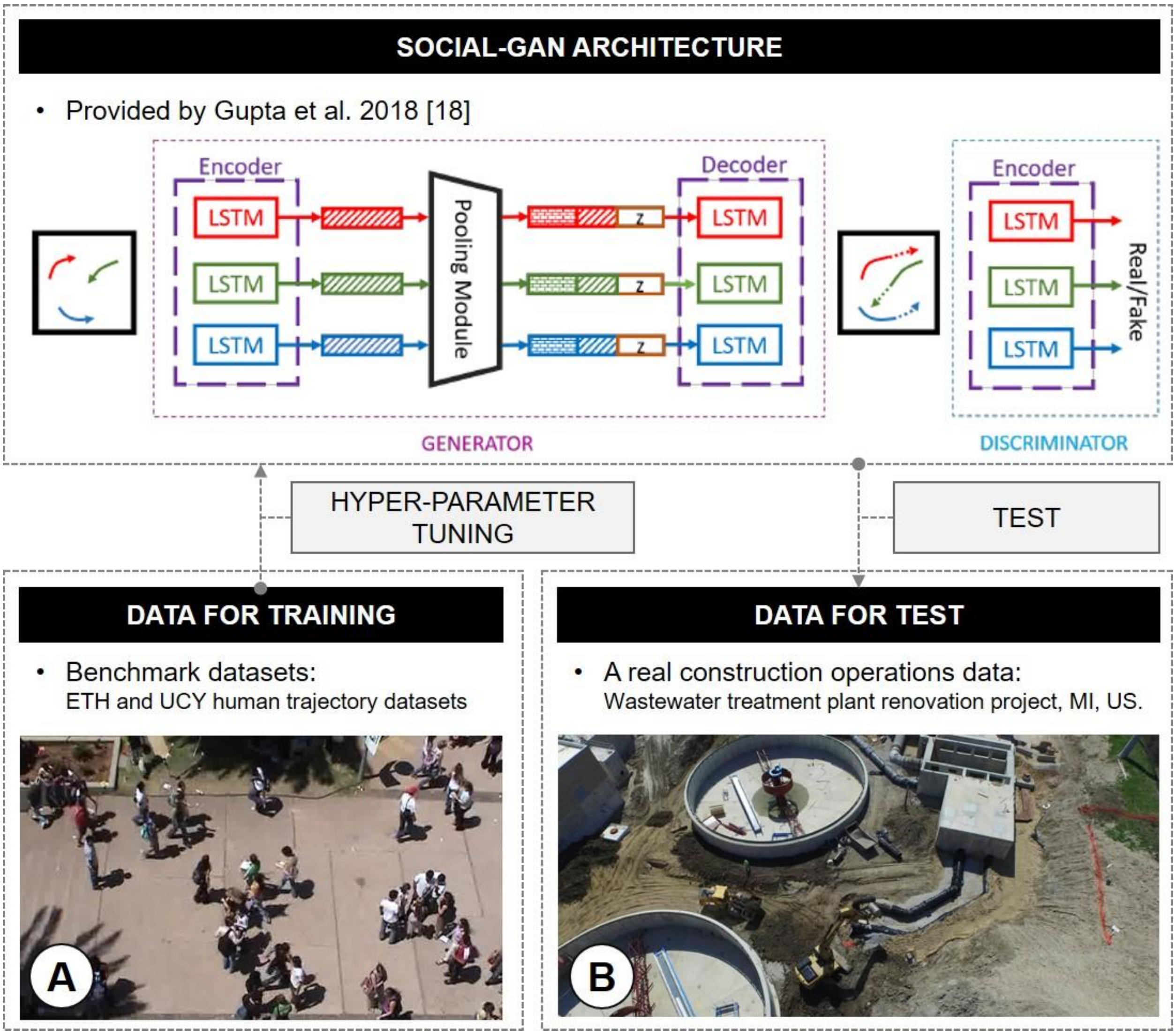

Figure 1. Research framework

5 Hyper-Parameter Tuning

To develop a long-term trajectory prediction model for construction mobile resources, this study conducted hyper-parameter tuning of the Social GAN [12].

The Social GAN has around 40 hyper-parameters that might need to be considered for successful training: batch size, number of iteration, number of epoch, model dimensions, observation length, and prediction length, to name a few. A small change in each hyper-parameter might be able to affect training and a trained model's final performance; however, examining all possible combinations is not viable as training a model with a graphical processing unit (GPU, e.g., Tesla K40c) in general takes more than five days. Hence, this study selected two important hyper-parameters, prediction and observation length, as tuning targets with the following reasons:

- Prediction length: the prediction length is the most important hyper-parameter that literally determines how many time-steps the model will predict. To achieve the prediction model capable of predicting more than five seconds, this study changed the prediction length from the default value (8 time-steps, 2.64 s) to 16 time-steps (5.28 s).
- Observation length: the observation length was selected as the second important hyper-parameter that needs to be tuned. The major input consumed for inferring a set of future trajectory in the Social GAN is a set of observed trajectory. The length of

observation is therefore bound to have a significant impact on a trained model's prediction performance. This work considered seven different observation lengths: from 8 time-steps (2.64 s) to 20 time-steps (6.6 s) with 2 time-steps interval (0.66 s).

Considering the above two hyper-parameters, total seven different tuning scenarios were established and applied in training (Table 1).

Table 1. Hyper-parameter tuning scenarios

Scenarios	Hyper parameter			
	Observation length		Prediction length	
	Time-steps	Seconds	Time-steps	Seconds
#1	8	2.64	16	5.28
#2	10	3.30	16	5.28
#3	12	3.96	16	5.28
#4	14	4.62	16	5.28
#5	16	5.28	16	5.28
#6	18	5.94	16	5.28
#7	20	6.60	16	5.28

6 Test Result and Discussion

To evaluate the trajectory prediction accuracy of the trained models on real construction settings, a test on a real construction operations data is conducted. As evaluation metrics, this study applied average displacement error (ADE) and final displacement error (FDE) that are commonly used metrics in trajectory prediction studies [12,17].

- ADE: average *L2* distance (i.e., mean square error) between ground truth and prediction over all predicted time-steps [12,17].
- FDE: distance between the predicted final destination and the ground truth destination at the end of the prediction period [12,17].

Table 2 summarizes the ADE and FDE of each trained model on the test dataset. Overall, all seven of the trained models showed promising accuracy in this test: the ADEs and FDEs for all the models were less than one meter (avg. ADE=0.88 meters) and 1.6 meters (avg. FDE=1.51 meters), respectively. Given a set of observation, predicting position of far time-step is naturally more challenging than close one. Accordingly, it was shown that the FDEs are 0.6 meters higher than the ADEs on average.

In this test, it was revealed that longer observation length does not necessarily guarantee higher accuracy. Longer observation means that the trajectory of less relevant time-steps are more consumed in the prediction. For example, in the scenario #7, not that all 20 time-steps observation are closely relevant to the future time-steps positions. The first several time-steps observation may have less relevancy to the future trajectory than the last several ones, which can be noises and have a negative impact on the model's prediction performance. In actual, the ADEs and FDEs slightly increased as observation length increased (Table 2).

It turned out that the best model for 16 time-steps prediction (5.28 s) is the one with 8 time-steps (2.64 s) observation (Table 2). This model consumes the shortest observation in the prediction, which however has the highest relevancy to the future trajectory. Consequently, this model outperformed the others and showed the most promising result: ADE=0.78 meters and FDE=1.27 meters (Table 2). Figure 2 illustrates the best model's prediction performance. Note that in this figure, green, blue, red lines stand for predicted trajectory of the worker, wheel loader, and excavator, respectively. And white circles stand for their ground truth position. As shown in Figure 2, the ground truth position of each entity well follows the predicted trajectory in process of time. This fact visually verifies validity of the predicted trajectories.

The developed model also demonstrated that it can continuously update the trajectory prediction at every 0.33 s without significant time-lag. With the use of a GPU (i.e., Tesla K40c), the model predicts three sets of trajectories for 5.28 s (16 time-steps) within 0.12 s. Then, at the next time step, 0.21 s later after completing the previous prediction, it predicts new sets of trajectories with the latest observation. That is, the model can continuously provide trajectory prediction for 5.16 s (5.28 s – 0.12 s) at every 0.33 s.

This test shows the great potential for the developed model in predicting construction mobile resources' trajectories. However, there is still room for further improvement. This study only focuses on hyper-parameter tuning, not considering fine-tuning with augmented construction data. Once an extensive dataset for construction resources' trajectory is available, this work will have another chance that can likely improve the prediction performance.

Table 2. Test result

Scenarios	Observation length	ADE (meters)	FDE (meters)
#1	80	0.78	1.27
#2	100	0.87	1.48
#3	120	0.84	1.42
#4	140	0.89	1.53
#5	160	0.89	1.54
#6	180	0.98	1.80
#7	200	0.91	1.57

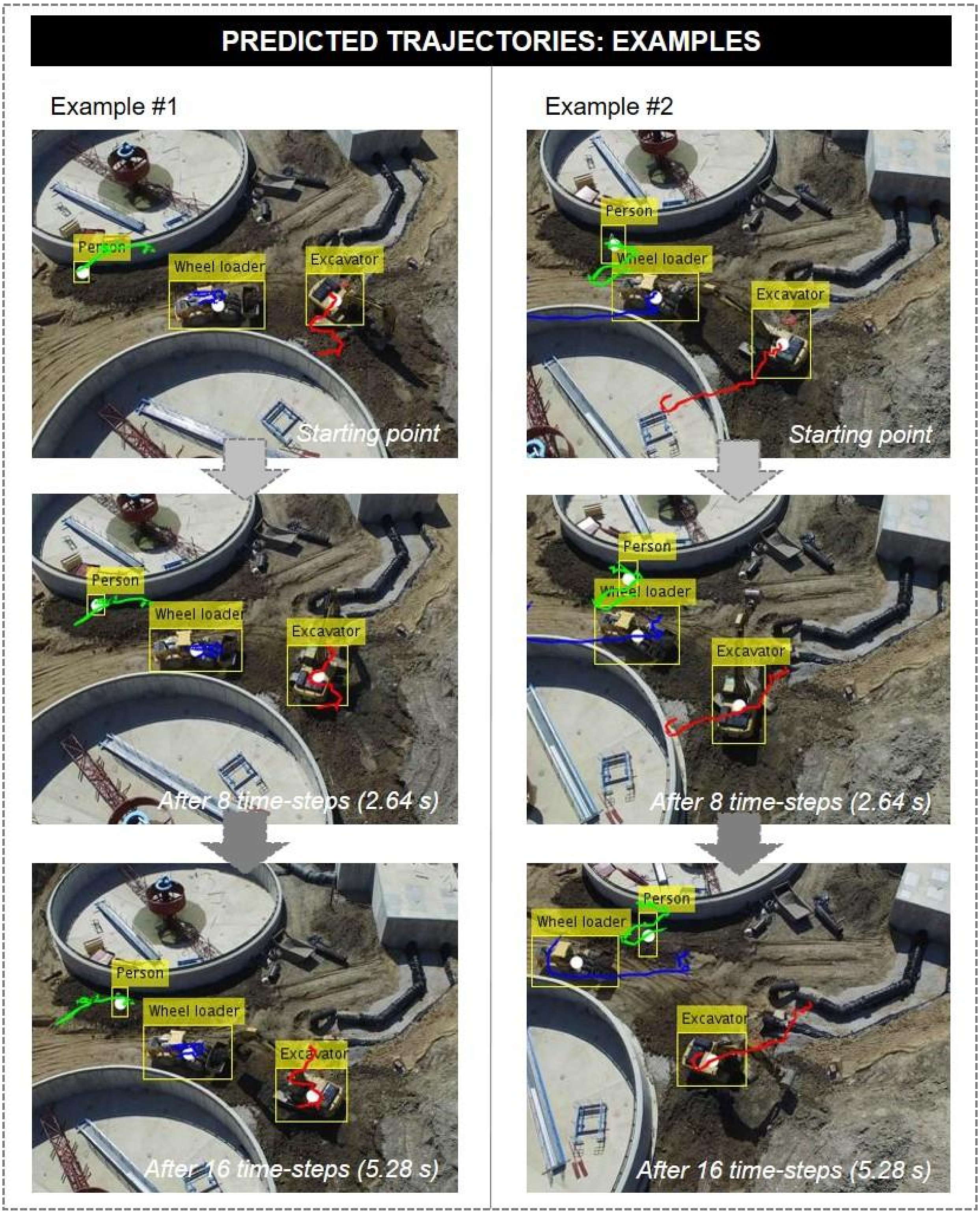

Figure 2. Predicted trajectory vs. ground truth positions

7 Conclusion

To support the pro-active struck-by hazard detection in construction, this study developed a trajectory prediction model for construction mobile resources. Specifically, this study conducted hyper-parameter tuning of a deep neural network, called Social GAN and developed a prediction model capable of predicting target's trajectory more than five seconds.

As a result, the best model (i.e., scenario #1) could achieve promising prediction accuracy: the ADE of 0.78 meters and the FDE of 1.27 meters. However, there still remain critical opportunity to improve the prediction accuracy—such as the fine-tuning with augmented construction training dataset.

With such refinement, an updated model would likely result in a more robust and accurate prediction on real construction operations data. The proposed trajectory prediction allows for detecting hazards that will be present in the near future, which will support pro-active intervention in hazardous situations. It will ultimately contribute to promoting a safer working environment for construction workers.

8 Acknowledgement

The work presented in this paper was supported financially by a National Science Foundation Award (No. IIS-1734266, 'Scene Understanding and Predictive Monitoring for Safe Human-Robot Collaboration in Unstructured and Dynamic Construction Environment'). Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation. Lastly, the authors wish to specially thank Weston Tanner and John McGlennon from WALSH Construction Co. for their considerate assistance in collecting onsite data.

References

[1] Kim, D., Liu, M., Lee, S., and Kamat, V.R. Remote proximity monitoring between mobile construction resources using camera-mounted UAVs. *Automation in Construction*, 99(2019):168–182, 2019.

[2] CPWR, Struck-by injuries and prevention in the construction industry. *The Center for Construction Research and Training*, 2017.

[3] OSHA, Recommended Practices for Safety & Health Programs in Construction. *Occupational Safety and Health Administration*, 2017.

[4] Bahn, S. Workplace hazard identification and management: The case of an underground mining operation. *Safety Science*, 57(2013):129-137, 2013.

[5] Teizer, J., Allread, B.S., Fullerton, C.E., Hinze, J. Autonomous pro-active real-time construction worker and equipment operator proximity safety and alert system. Automation in Construction, 19(2010):630-640, 2010.

[6] Marks, E. and Teizer, J. Proximity sensing and warning technology for heavy construction equipment operation. *Construction Research Congress 2012*, West Lafayette, IN, USA, 2012

[7] Teizer, J. Wearable, wireless identification sensing platform: Self-monitoring alert and reporting technology for hazard avoidance and training (smarthat). *Electronic Journal of Information Technology in Construction*, 20:295-312, 2015

[8] Ruff, T.M. Monitoring blind spots: A major concern for haul trucks. *Engineering and Mining Journal*, 202(12):17-26, 2001.

[9] Park, J.W., Marks, E., Cho, Y.K., and Suryanto, W. Performance test of wireless technologies for personnel and equipment proximity sensing in work zones. *Journal of Construction Engineering and Management*, 142(1): 04015049, 2016

[10] Kim, D.H., Yin, K., Liu, M., Lee, S.H., and Kamat, V.R. Feasibility of a drone-based on-site proximity detection in an outdoor construction site. *IWCCE 2017*, Seattle, WA, USA, 2017

[11] Kim, H.J., Kim, K.N., and Kim, H.K. Vision-based object-centric safety assessment using fuzzy inference: Monitoring struck-by accidents with moving objects. *Journal of Computing in Civil Engineering*, 30: 04015075, 2016

[12] Gupta, A., Johnson, J., Fei-Fei, L., Savarese, S. and Alahi, A. Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[13] Fang, Q., Li, H., Luo, X., Ding, L., Luo, H., Rose, T.M., and An, W. Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. *Automation in Construction*, 85(2018):1-9, 2018.

[14] Kim, H.J., Bang, S.D., Jeong, H.Y., Ham, Y.J., and Kim, H.K. Analyzing context and productivity of tunnel earthmoving process using imaging and simulation. *Automation in Construction*, 92(2018):188-198, 2018.

[15] Gers, F.A., Schmidhuber, J. and Cummins, F. Learning to forget: Continual prediction with LSTM. 1999.

[16] Chung, J., Gulcehre, C., Cho, K. and Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.

[17] Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L. and Savarese, S. Social lstm: Human trajectory prediction in crowded spaces.

Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 961-971, 2016.

[18] Pellegrini, S., Ess, A., and Van Gool, A. Improving data association by joint modeling of pedestrian trajectories and groupings. *Computer Vision–ECCV 2010*, 452–465, 2010.

[19] Leal-Taix´e, L., Fenzi, M., Kuznetsova, A., Rosenhahn, B., and Savarese, S. Learning an image-based motion context for multiple people tracking. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.