Enhancing Deep Neural Network-based Trajectory Prediction: Fine-tuning and Inherent Movement-driven Post-processing

Daeho KIM[1], Houtan JEBELLI[2], SangHyun LEE[3], and Vineet R. KAMAT[4]

[1] Ph.D. Candidate, Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI 48109; PH (734) 764-8495; FAX (734) 764-4292; email: daeho@umich.edu

[2] Assistant Professor, Department of Architectural Engineering, The Pennsylvania State University, University Park, PA 16802; PH (814) 863-6442; email: hkj5117@psu.edu

[3] Professor, Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI 48109; PH (734) 764-9420; FAX (734) 764-4292; email: shdpm@umich.edu

[4] Professor, Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI 48109; PH (734) 764-4325; FAX (734) 764-4292; email: vkamat@umich.edu

ABSTRACT

As a proactive means of preventing struck-by accidents in construction, many studies have presented proximity monitoring applications using wireless sensors (e.g., RFID, UWB, and GPS) or computer vision methods. Most prior research has emphasized proximity detection rather than prediction. However, prediction can be more effective and important for contact-driven accident prevention, particularly given that the sooner workers (e.g., equipment operators and workers on foot) are informed of their proximity to each other, the more likely they are to avoid the impending collision. In earlier studies, the authors presented a trajectory prediction method leveraging a deep neural network to examine the feasibility of proximity prediction in real-world applications. In this study, we enhance the existing trajectory prediction accuracy. Specifically, we improve the trajectory prediction model by tuning its pre-trained weight parameters with construction data. Moreover, inherent movement - driven post-processing algorithm is developed to refine the trajectory prediction of a target in accordance with its inherent movement patterns such as the final position, predominant direction, and average velocity. In a test on real-site operations data, the proposed approach demonstrates the improvement in accuracy: for 5.28 seconds' prediction, it achieves 0.39 meter average displacement error, improved by 51.43% as compared with the previous one (0.84 meters). The improved trajectory prediction method can support to predict potential contact-driven hazards in advance, which can allow for prompt feedback (e.g., visible, acoustic, and vibration alarms) to equipment operators and workers on foot. The proactive intervention can lead the workers to take prompt evasive action, thereby reducing the chance of an impending collision.

INTRODUCTION

At the construction site where every circumstance dynamically evolves, contact-driven accidents often happen in various forms, resulting in a significant number of construction fatalities. According to *The Center for Construction Research and Training, the U.S.*, from 2011 to 2015, a total of 1,079 construction workers died due to the forcible contact or impact by a mobile vehicle or a piece of equipment [# of struck-by = 804 (CPWR 2017a) and # of caught-in/between = 275 (CPWR 2017b)]. The figure accounted for 24% of overall contact-driven fatalities in the U.S. and was unmatched by other U.S. industries (CPWR 2017a; CPWR 2017b). Notably, such fatalities continue to rise annually; it reached to 367 in 2017, accounting for 37% of the overall construction fatalities during that year (BLS 2019).

A major research area for this issue has been attuned to automating onsite proximity monitoring. And, various technologies—such as radio frequency identification (Teizer et al. 2010; Marks et al. 2012), magnetic field (Teizer et al. 2015), global positioning system (Ruff 2001), bluetooth low energy (Park et al. 2016), or computer vision (Kim et al. 2019a; Kim et al. 2017; Kim et al. 2016)—have been considered in many previous studies to this end. They proved the great potential of such technologies, demonstrating a high accuracy of proximity monitoring in onsite applications.

Most previous studies to date have been dominated by proximity detection or monitoring. In many cases, however, prediction can be more effective and important for contact-driven accident prevention. This is because the sooner workers (e.g., equipment operators and workers on foot) are informed of their proximity to each other, the more likely they are to avoid the impending collision. Nevertheless, few studies have attempted to address it.

To address the proximity prediction, our earlier study developed a trajectory prediction model for mobile construction resources (Kim et al. 2019b). We applied an established deep neural network (DNN) for trajectory prediction, called Social GAN (Gupta et al. 2018), and particularly tuned its hyper-parameters and network architecture so that we can make a longer prediction (5.24 seconds) than the original one (3.96 seconds).

Based on this prior work, this round of study focuses on enhancing the model's prediction accuracy. To this end, we conduct fine-tuning of the pre-trained model with construction data and develop a post-processing algorithm that can automatically refine the trajectory prediction of each target. In addition, a test on construction operations data follows so as to demonstrate the effect of the fine-tuning and post-processing on prediction accuracy improvement.

PRIOR WORK: TRAJECTORY PREDICTION USING A DNN

Trajectory prediction studies have been led by data-driven learning approaches. This is because the movement of an entity is so diverse and uncertain that it is more viable to predict it based on what can be observed from given data, rather than by a standardized algorithm. Traditionally, it often adopted hand-crafted feature-based methods (Helbing et al. 1995; Antonini et al. 2006; Yamaguchi. Et al. 2011) or statistical learning methods (Tay and Laugier 2008; Trautman et al. 2015). However, nowadays, many studies for trajectory prediction are motivated to use a DNN since it is more suitable for the task that requires learning complex and intractable probability distribution.

In recent years, a number of DNN architectures for trajectory prediction have been proposed: to name a few, Social LSTM (Alahi et al.2016), Crowd Interaction DNN (Xu et al. 2018), Interaction Aware DNN (Pfeiffer et al. 2018), and Social GAN (Gupta et al. 2018). Enjoying the support of an increased dataset, stronger computing power, and advanced learning algorithms, such DNNs continue to improve the performance of trajectory prediction.

Among such DNNs, Social GAN (Gupta et al. 2018) shows several distinctive features. It enables to learn social behavior (e.g., collision avoidance) as well as an entity's moving pattern by integrating a social pooling layer and an LSTM encoder-decoder. In addition, by realizing generative adversarial network (GAN), it enhances intractable probabilistic computation and accordingly behavioral inference of which other DNNs are incapable. The original work (Gupta et al. 2018) used Social GAN for 2.6/4.0 seconds prediction and showed its superior performance over other DNNs.

In this regard, we applied Social GAN (Gupta et al. 2018) to address the trajectory prediction of mobile construction resources and thereby pro-active detection of contact-driven hazards (e.g., struck-by or caught-in/between). Figure 1 illustrates how the trajectory prediction can address the pro-active hazard detection. Moreover, our earlier study tuned its architecture and hyper-parameters (e.g., prediction and observation lengths) so that it can make longer, yet valid prediction. Note that longer prediction and thereby earlier notice are needed to provide a worker in a danger with enough time for evasive action.

To this end, we modified two major hyper-parameters—observation and prediction lengths—and trained it with two benchmark dataset of pedestrian trajectories [ETH (Pellegrini et al. 2010) and UCY (Leal-Taixe et al. 2014)]. As a result, we developed several models to have different observation length (from 2.64 seconds to 6.60 seconds at every 0.66 seconds), but be capable of predicting 5.28 seconds. Overall, all the models demonstrated a promising accuracy in a test; they achieved the displacement error of 0.88 meters on average.

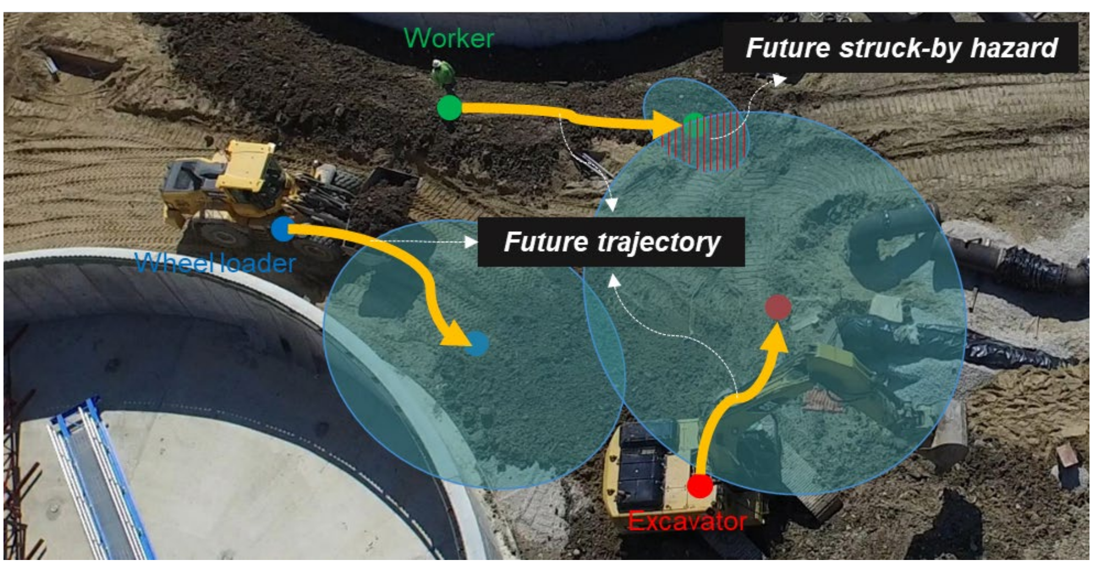

Figure 1. Trajectory Prediction and Pro-active Hazard Detection

RESEARCH OBJECTIVE: ENHANCING PREDICTION ACCURACY

While the earlier study focused on the architecture modification of Social GAN (Gupta et al. 2018) to achieve longer prediction length, this study put major emphasis on enhancing the accuracy of trajectory prediction. To this end, we conduct fine-tuning of the prior trajectory prediction model and further adds a post-processing algorithm. In addition, a new test on real construction operations data follows so as to validate the effect of fine-tuning and post-processing on prediction accuracy improvement.

IMPROVEMENT #1: FINE-TUNING WITH CONSTRUCTION DATA

In the prior study, we already pre-trained the empty architecture of Social GAN (Gupta et al. 2018) with the two benchmark dataset, taking into account several different sets of hyper-parameters such as observation and prediction length (Kim et al. 2019b). At this round, we further fine-tune the pre-trained model with additional construction data, thereby better fitting it to construction settings.

We first collected and annotated a real construction video that captures human-equipment interaction for the purposes of testing as well as fine-tuning. Specifically, several UAV-captured construction site videos were collected, of which 916 sequential frames were used for fine-tuning, and 398 frames for test. Each trajectory (i.e., a set of x-y coordinates) of targets—a worker, a wheel loader, and an excavator—were manually annotated over the whole frames and a complete inspection followed to ensure the annotation validity.

We fine-tuned the weights of the pre-trained model by continuing training with the integrated dataset of ETH (Pellegrini et al. 2010), UCY (Leal-Taixe et al. 2014), and the construction data. We considered the below details in this task:

- Number of epoch: to avoid under-fitting, the number of epoch was increased to 400 from the prior default value of 200.
- Batch size: considering the limit of hardware support (e.g., memory capacity), it the batch size was set to 16.
- Parallel computing: to accelerate the tuning process, we used a graphical processing unit (GPU, NVIDIA TITAN V) as well as a central processing unit. The tuning was continued for four days in our setting until reaching to the number of epoch.
- Architecture-related hyper-parameters: the Social GAN consists of three components—generator, pooling module, and discriminator—and there are many architecture-related hyper-parameters, which need to be considered for successful training. Table 1 summarizes the value of hyper-parameters that we actually applied in the tuning process.

Table 1. Examples of Architecture-related Hyper-Parameters

Component	Hyper-parameter	Description	Value
Generator	encoder_h_dim_g	Dimensions of the hidden layer in the encoder	32
	decoder_h_dim_g	Dimensions of the hidden layer in the decoder	32
	noise_dim	Dimensions of the noise added to the input of the decoder	8
	noise_type	Type of noise to be added	Gaussian
	clipping_threshold	Threshold at which the gradients is clipped	2
	g_learning_rate	Learning rate for generator	0.0001
Pooling	bottleneck_dim	Output dimension of pooled vector	8
	neighborhood_size	neighborhood size for social pooling	2
	gird_size	the size of grid to determine neighborhood	8
Discriminator	encoder_h_dim_d	Dimensions of the hidden layer in the encoder	48
	d_learning_rate	Learning rate for discriminator	0.001
	clipping_threshold	Threshold at which the gradient is clipped	0

IMPROVEMENT #2:
INHERENT MOVEMENT-DRIVEN POST-PROCESSING

The fine-tuned model is designed to output a sixteen time-steps future trajectory [16x2, x-y coordinates of 16 time-steps (equivalent to 160 frames or 5.24 seconds)] for an observation input [12x2, x-y coordinates of 12 time-steps (equivalent to 120 frames or 4.00 seconds)]. We devised a post-processing algorithm that extracts an individual entity's inherent movement attributes from the observation input and reuse it to refine the prediction output.

An entity (e.g., construction worker) tends to keep its own inherent movement attributes such as predominant direction and average velocity, and more so if it does not sense a situational change (e.g., vehicle coming closer) and thereby does not have a sudden motivation. The developed algorithm utilizes such constancy underlying each individual entity. Specifically, it considers the three inherent movement attributes: the final position, predominant direction, and average velocity. The details are as below:

- Inspection of starting position (Figure 2): this algorithm first examines the validity of the predicted trajectory's starting position. Specifically, it updates the starting position to the extrapolated coordinates of the last two observed positions if the starting position of prediction is far from the final position of observation over a certain threshold. We used the distance between the last two observed positions as the threshold value.

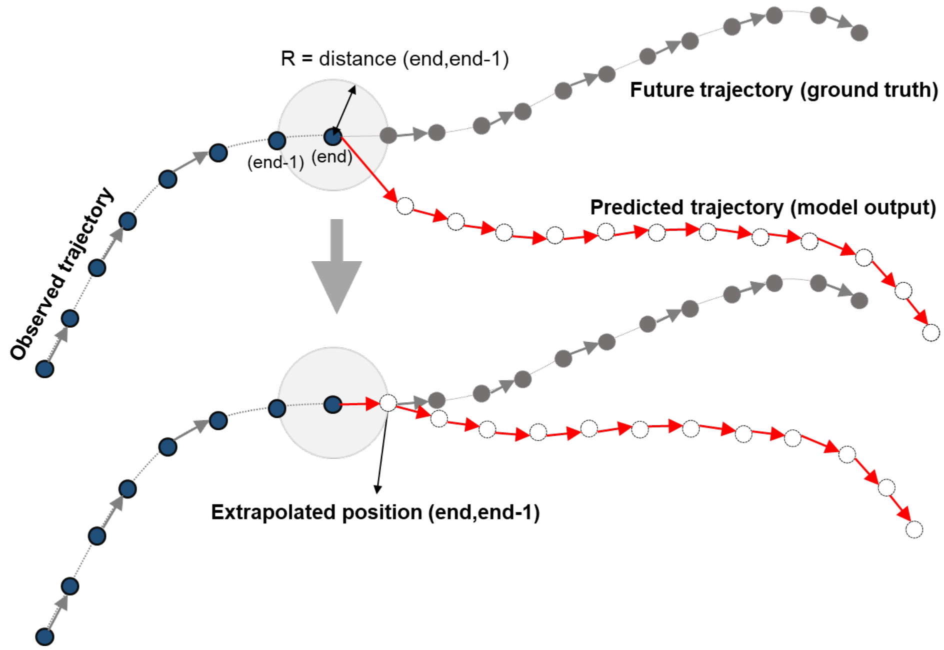

Figure 2. Post-processing: Inspection of Starting Position

- Correction of predominant direction (Figure 3): second, the algorithm updates predicted trajectory's predominant direction. It refines the direction to the observed one if the direction vector of the predicted trajectory is warped to that of observed trajectory over a certain threshold (8 degrees).

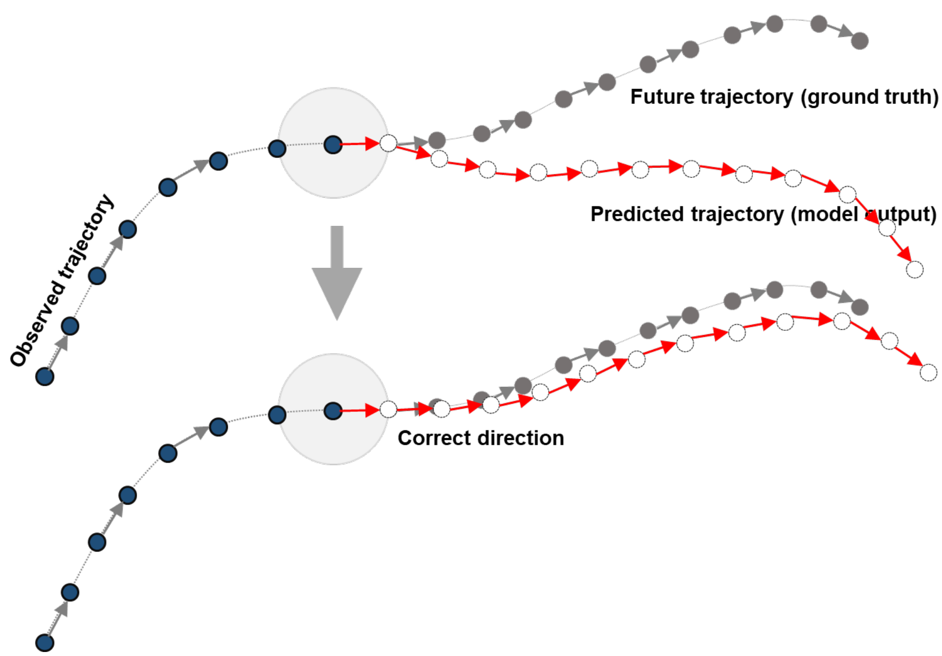

Figure 3. Post-processing: Correction of Direction

- Correction of average velocity (Figure 4): lastly, the algorithm corrects the predicted trajectory's average velocity. It adjusts the predicted trajectory's average velocity to the original one and accordingly refines trajectory prediction if it exceeds the original average velocity more than 20%.

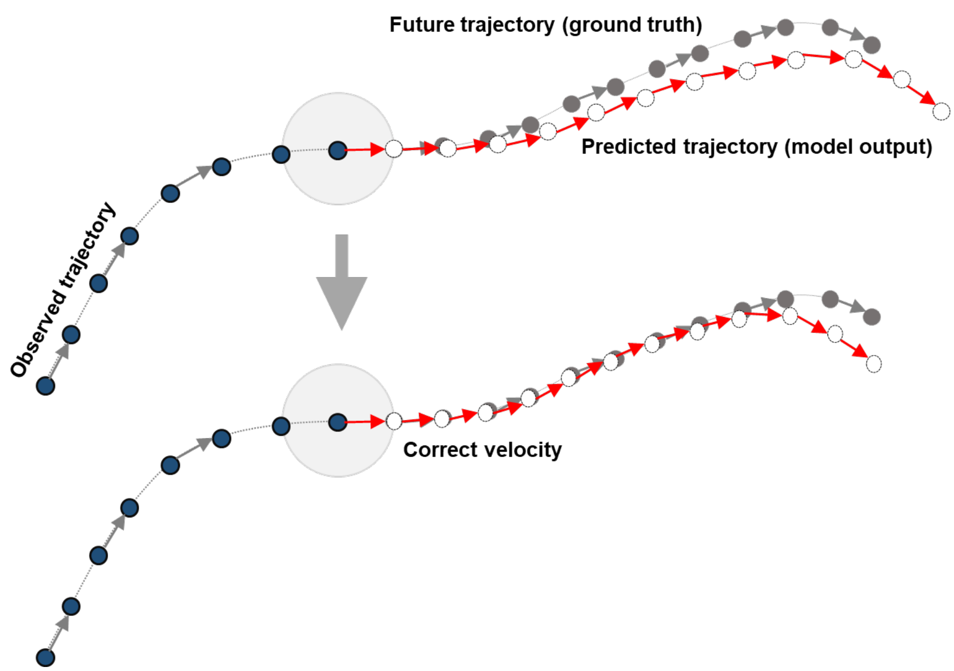

Figure 4. Post-processing: Correction of Average Velocity

TEST RESULT AND DISCUSSION

To demonstrate the effect of fine-tuning and inherent movement -driven post-processing on the improvement of prediction accuracy, a test on real construction data was conducted. As evaluation metrics, we applied average displacement error (ADE) and final displacement error (FDE) that were used in our earlier study (Kim et al. 2019b).

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

Table 2. Test Result

Metric	Pre-trained	Fine-tuned (w/o post-processing)	Fine-tuned (w/ post-processing)
Average displacement error	0.84 m	0.45 m	0.38 m
Final displacement error	1.42 m	0.87 m	0.75 m

Table 2 summarizes the ADE and FDE of the original pre-trained model and fine-tuned models without and with the post-processing. Overall, all three cases showed a promising accuracy in this test: the ADEs and FDEs for all cases were less than one meter and 1.5 meters, respectively.

It turned out that the fine-tuning was highly effective in improving trajectory prediction accuracy. Compared to the original pre-trained model, it reduced ADE and

FDE by 46% and 38%, reaching down to 0.4498 m and 0.8761 m, respectively. Notably, the fine-tuned network also showed a good performance on the original benchmark test dataset (i.e., ETH): the ADE and FDE were 0.40 m and 0.58 m, respectively.

The post-processing algorithm also improved the trajectory prediction accuracy: both the ADE and FDE was further reduced by 13% and consequently achieved 0.3894 m ADE and 0.7571 m FDE. The fine-tuned Social GAN itself showed promising performance in this test; however, the limitation of learning with a limited amount of data was certain. The developed post-processing algorithm showed it can help to offset the limitation by reusing the inherent movement attributes underlying each individual's movement history.

CONCLUSION

To enhance the performance of trajectory prediction, this study conducted fine-tuning of a pre-trained trajectory prediction model and developed a post-processing algorithm that can automatically refine the trajectory prediction of each target. As a result, we achieved 0.39 meters ADE in a 5.28 seconds prediction task, improved by 51% as compared with the previous one (0.84 meters).

During construction operations, contact-driven hazards can easily arise in various scenarios. In the unpredictable situations, the presented method will enable the advanced detection of a potential hazard, thereby making timely intervention possible. The proactive intervention will save construction workers from potentially fatal hazards and ultimately help to promote safer work environments in construction projects.

ACKNOWLEDGMENT

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 especially thank Weston Tanner and John McGlennon from WALSH Construction Co. for their considerate assistance in collecting onsite data.

REFERENCES

Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., and Savarse, S. (2016). "Social LSTM: Human trajectory prediction in crowded spaces." *The IEEE Conference on Computer Vision and Pattern Recognition (CPVR)*, 961-971, 2016.

Antonini, G., Bierlarire, M., and Weber, M. (2006). "Discrete choice models of pedestrian walking behavior." *Transportation Research Part B: Methodological*, 40(8):667-687, 2006.

BLS. (2019), "Census of fatal occupational injuries-Current and revised data." *Bureau of Labor Statistics*, 2019, <https://www.bls.gov/iif/oshcfoi1.htm>

CPWR. (2017a), "Struck-by injuries and prevention in the construction industry." *The Center for Construction Research and Training*, 2017.

CPWR. (2017b), "Caught-in/between injuries and prevention in the construction industry." *The Center for Construction Research and Training*, 2017.

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

Helbing, D and Molnar, P. (1995). "Social force model for pedestrian dynamics." *Physical Review E*, 51(4282), 1995.

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

Kim, D., Liu, M., Lee, S., and Kamat, V.R. (2019b). "Trajectory prediction of mobile construction resources toward pro-active struck-by hazard detection." *36th International Symposium on Automation and Robotics in Construction*, Alberta, Canada, 2019.

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

Kim, H.J., Kim, K.N., and Kim, H.K. (2016). "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

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

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

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

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

Pfeiffer, M., Paolo, G., Sommer, H., Nieto, J., Siegwart, R., and Cadena, C. (2018). "A data-driven model for interaction-aware pedesrtrain motion prediction in object cluttered environments." *The IEEE Conference on Robotics and Automation (ICRA)*, 21-25, 2018.

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

Tay, M.K.C. and Laugier, C. (2008). "Modeling smooth paths using Gaussian processes." *Field and Service Robotics*, 381-390, 2008.

Teizer, J. (2015). "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

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

Trautman, P., Ma, J., Murray, R.M., Krause, A. (2015). “Robot navigation in dense human crowds: Statistical models and experimental studies of human-robot cooperation.” *The International Journal of Robotics Research*, 34(3):335-356, 2015.

Xu, Y., Piao, Z., and Gao, S. (2018). “Encoding crowd interaction with deep neural network for pedestrian trajectory prediction.” *The IEEE Conference on Computer Vision and Pattern recognition (CPVR)*, 5275-5284, 2018.

Yamaguchi, K., Berg, A.C., Ortiz, L.E., and Berg, T.L. (2011). “Who are you with and where are you going?” *The IEEE Computer Vision and Pattern Recognition (CPVR)*, 3488-3496, 2015.