Fake it till you make it: training deep neural networks for worker detection using synthetic data

Tohidifar, A., Saffari S. F., Kim D.
University of Toronto, Canada
civdaeho.kim@utoronto.ca

Abstract. The construction industry's productivity and safety have long been a source of concern, while the broad use of deep neural network (DNN)-based visual AI has transformed other industries. Automation and digitalization powered by DNN provide intriguing answers; yet the lack of high-quality, diversified data prevents the construction sector from leveraging the benefits. This paper presents a novel computational framework that enables synthetic data generation for DNN training to overcome the time-consuming manual data collection and avoid data privacy problems. The suggested framework uses graphics engines to create a virtual duplicate of the construction site that generates non-real yet realistic visuals. The proposed framework randomizes crucial scene elements such as worker pose, clothes, camera viewpoint, and lighting conditions to enhance the variety of the synthetic dataset. The findings of this study present promising potential of synthetic data in DNN training.

1. Introduction

Accounting for 13 percent of global GDP, construction, an industry that provides infrastructure essential to daily life, is one of the largest industries worldwide (Oxford Economics, 2021). Despite its vital importance, the construction industry is facing several challenges. It is renowned as the least automated (Neythalath et al., 2021), experiencing significant safety (Zhou et al., 2015) and productivity problems (Dixit et al., 2019). In recent years, Deep Neural Network (DNN)-based visual Artificial Intelligence (AI), under the umbrella term of Industry 4.0, made its way into different industries and improved the productivity and safety of the work environments (Chien et al., 2020). In spite of DNNs' promises in other industries, significant challenges hinder the construction industry from reaping the benefits (Rao et al., 2022). On the one hand, the construction sites' dynamic, complex, and unstructured nature demand a higher level of AI than other industries (Daeho Kim et al., 2020; Laufer Alexander et al., 2008). On the other hand, the unavailability of data and privacy concerns around the collected data (Akinosho et al., 2020) raised significant hindrances to the widespread adoption of DNN-based visual AI in the construction industry. Even though many recent studies focused on applying DNNs, the development of field-applicable models appears to be over-ambitious (Bang et al., 2019). Insufficient high-quality and diversified data in construction studies resulted in incomplete and overfitted models, ultimately limiting the accuracy and scalability of DNN solutions (Grosse et al., 2021).

To address the data shortage of DNN solutions, a significant portion of previous studies were focused on the manual collection and labelling of data. However, manual labelling is laborious, time-consuming, and expensive (Assadzadeh et al., 2022). While the exact costs of labelling construction data remain unknown, the scale of the problem can be conveyed by looking at the crowdsourcing data labelling services. Semantic segmentation provided by Google Cloud would incur 0.87 USD per class for a single image (Dahun Kim et al., 2020). Besides the prohibitively time-consuming nature of manual data collection, the collected data is susceptible to human error and biases (Assadzadeh et al., 2022). Incorrect annotations would decrease the accuracy of experiments (Xiao and Kang, 2021). Since training high-quality DNNs require millions of labelled data, notable investments from both time and expenses are involved in the manual collection, labelling, and ensuring the quality of the data.

https://doi.org/10.7146/aul.455.c229

An essential element of DNNs' prosperity in other industries is the availability of publicly shared datasets and access to a benchmarking system. Comparative benchmarking, however, in the construction industry is limited due to the lack of willingness of stakeholders to share project datasets (Hwang et al., 2018). The stratified analysis showed that practitioners are primarily unsatisfied with the level of data sharing among stakeholders in the construction industry (Ayodele and Kahimo-shakantu, 2021). Furthermore, recording personal visual data has brought privacy concerns (Akinosho et al., 2020). Ethical issues regarding privacy are of concern in many countries, which resulted in the passing of strict laws to protect privacy (Fabbri et al., 2021). In light of the recent General Data Protection Regulation (GDPR), which applies to all companies holding EU citizen data, care must be taken with data sources (Koops, 2014). Ethical issues regarding privacy are also critical in the US. As an instance of privacy concerns, datasets for re-identification modules of DukeMTMC were taken offline recently (Fabbri et al., 2021; Ristani et al., 2016).

One possible solution to the data limitation issue is to leverage the available computational power to augment or synthesize data. The research community has already recognized the potential of synthetic data created by powerful graphical engines to compensate for the data shortage (Amato et al., 2019; Bousmalis et al., 2017) and create non-real, but real-looking comparative benchmarking (Fabbri et al., 2021). However, promising as it may seem, the potential of computer-generated data for training DNNs in the construction domain remained uninspected. There is a lack of understanding of DNNs' behaviour when trained with synthetic data, specifically in the construction domain. No study has been conducted regarding the feasibility of training DNNs only using synthetic data to understand construction workers. We do not know how the realism of synthetic images impacts the DNNs performance. In addition, there is no consensus about the optimal size of the synthetic dataset to achieve state-of-the-art performance. Would privacy be an issue? Should synthetic images include occluded and cluttered scenes? These are the questions that remained unanswered. As a preliminary step toward extensive experiments to answer the above questions, this study aims to develop a computational method of synthetic data generation for human workers in construction scenes and to visually verify the performance of synthetic data-trained DNN on real construction images.

Section 2 reviews the related works to the objective of the study and highlights the current state of the art and their limitations. Section 3 elaborates on the proposed method for synthetic data generation and explains the adapted DNN model. Training details, real-world validation dataset, and the employed evaluation metrics, along with achieved results are presented in Section 4. Lastly, the research contributions and limitations are covered in Section 5.

2. Related works

To tackle the data limitation and privacy concerns, two mainstream research can be seen: (i) data augmentation and (ii) synthetic data generation.

2.1 Data Augmentation

Data augmentation is an approach that utilizes the available dataset to augment or increase the dataset size by introducing linear transformations, interpolations and distortion, and probabilistic approaches (Delgado and Oyedele, 2021). Linear transformation approaches can be applied when the features of the datasets are not affected by alterations. Since images are transformation invariant, linear transformation techniques such as cropping, flipping, and

altering the colour space are easy and efficient methods to be implemented (Shorten and Khoshgoftaar, 2019). Interpolations and distortion methods introduce non-linear distortions and randomness to augmented data. This method has shown to be effective in shallow and deep neural networks (Delgado and Oyedele, 2021). Probabilistic methods, on the other hand, generate new data considering the distribution of the variables in the training set instances in a probabilistic manner (Delgado and Oyedele, 2021). To properly handle the probabilistic characteristic of a new dataset, this method is usually combined with deep learning models with probabilistic characteristics. Although many studies advocate the efficacy of data augmentation, this method do not augment new content to the dataset and is limited in the stimulation of the model's parameters during training (Luo et al., 2020, 2018). Moreover, it might discard important information that is essential as the labels. The label-preserving augmentation is of more concern when multiple images are being combined. Therefore, the scope of where and when these transformations can be applied is relatively limited (Shorten and Khoshgoftaar, 2019).

2.2 Synthetic Data

Another approach to increase the data size and quality is to generate synthetic data. In this approach, the computational powers are leveraged to simulate new data from analogous mediums. One of the most renowned pieces of work in this domain is the flying chairs study (Dosovitskiy et al., 2015). The proposed method of this study uses 3 dimensional (3D) models of chairs in a virtual world and generates synthetic images. Generated synthetic images are proven to be effective in improving DNN training. Since labels are extracted automatically in this approach, data generation is a seamless effort (Neuhausen et al., 2020); thus, an unlimited number of synthetic data is accessible as seen in the SYNTHIA dataset (Ros et al., 2016).

Despite the wide application of synthetic data in other disciplines, construction-related research expressed interest lately. With the rapid increase in the prevalence of building information modelling (BIM) (Alvanchi et al., 2021), research is underway to generate synthetic data from BIM. Acharya used the indoor images derived from BIM as training data for a deep learning network that estimates the pose of a camera (Acharya et al., 2019). Ma et al. (Ma et al., 2020) used synthetic point clouds obtained from BIM as the training data for point cloud segmentation networks. Hong et al. (Hong et al., 2021) employed BIM to construct a synthetic dataset that contains the annotation information of infrastructure elements. BIM-powered synthetic data generation is, inevitably, limited in scope to the building components. Human workers and equipment, as one of the most essential entities in productivity and safety of the construction site, are neglected in this approach.

To incorporate human workers and equipment in the synthetic data generation process, recent studies used 3D virtual models of construction sites. Soltani et al. (Soltani et al., 2018) applied 3D modelling tools to generate automatically labelled synthetic training images. They used virtual images of construction resources extracted from various views for training by using a 3D excavator model. Kim and Kim (Kim and Kim, 2018) reconstructed a 3D model of a concrete mixer truck using the multi-view stereo algorithm to generate synthetic data. Following the same approach, Mahmood et al. (Mahmood et al., 2022) developed a synthetic dataset for training DNN models that estimated the 3D pose of an excavator. In a very recent study by Neuhausen et al (Neuhausen et al., 2020), synthetically generated images are leveraged to train DNNs with a focus on human worker detection and tracking model. A comparative study of the developed model on both synthetic and real-world data identified the synthetic images as a viable solution for vision-based DNN training. Although synthetic data are applied widely out of the construction industry, there exists a promising potential to be used in

construction-related activities. In an attempt to investigate the capabilities of synthetic data on DNN training, this study uses the human workers images at virtual construction site to train a worker detection model.

3. Training a deep neural network using synthetic construction images

Following the promising results of prior studies in the computer vision domain (Fabbri et al., 2021), this study adopts a graphical engine (Blender (2022)) to simulate a virtual construction site. The virtual construction site is then used to render diversified images from various scenarios of construction activities. Figure 1 illustrates different parts of the proposed framework.

The first part of the framework is to build and prepare a construction worker avatar. In this part, a human worker conducts multiple construction activities. The worker's body motion is collected using the motion capture suit. This capturing method records the movement of all the body joints and anonymizes the worker that has done the activity. Meanwhile, a 3-dimensional (3D) avatar of the human worker is modelled. Both the motion information and the worker avatar are input to the graphical engines. In the second part, the 3D models of the buildings under construction are extracted from both points clouds or BIM models.

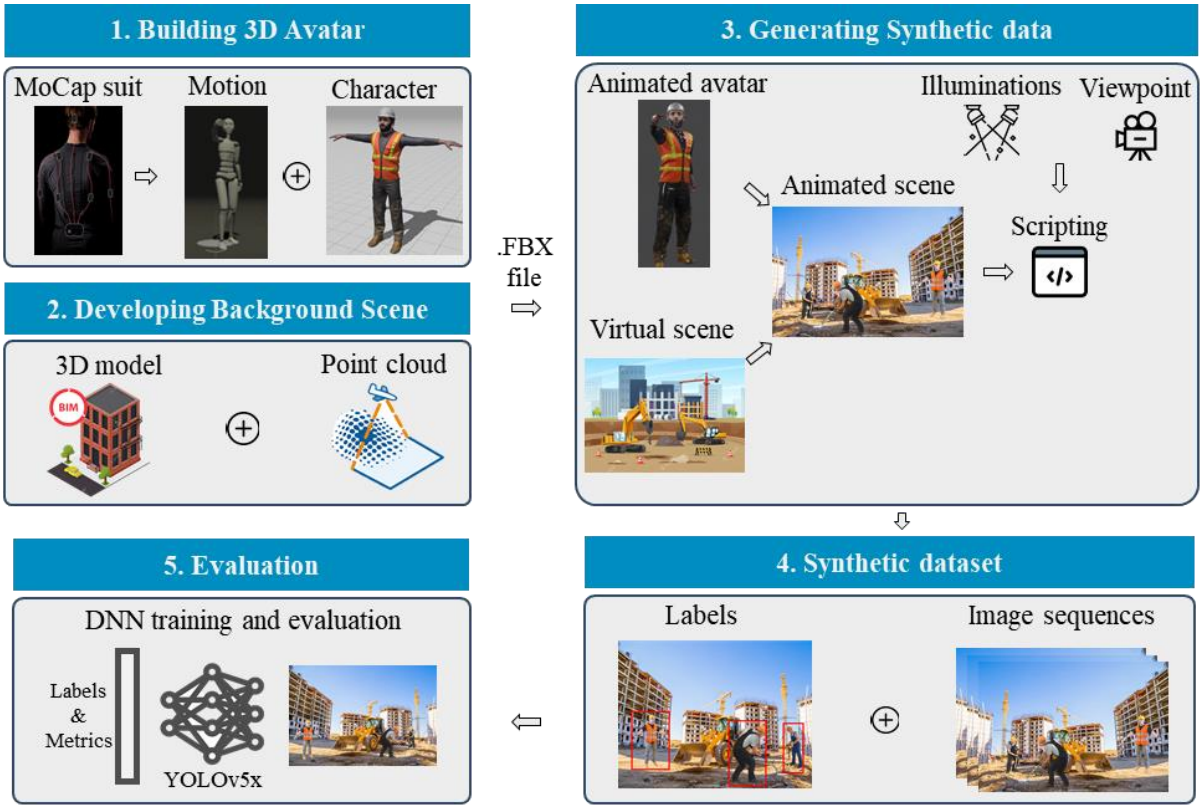

Figure 1. Synthetic data generation framework

In the third part, the collected information of the body joints' movements is augmented into a 3D worker avatar, to create an animated worker in the virtual environment. By placing multiple characters of animated avatars conducting various construction tasks on the developed digital construction replica, an animated construction scene is generated. By scripting into the graphical engine, we automatically randomized the camera location, its viewpoint in the digital world, and lighting conditions. By randomizing the scene, multiple screenplays are generated in the virtual world. Rendering each of these screenplays produces sequences of images. Since

all the manipulations take place on a graphical engine, extraction of entities' exact information is a seamless procedure. By using the scripting capability, the required ground truth labels are extracted from the virtual world as the last step of the third part. In the fourth part, rendered image sequences along with ground truth labels are organized as training and validation datasets. The fifth part of the framework uses the synthetic dataset to train DNN models for worker detection.

3.1 Dataset Generation

The synthetic data generation process of this study is comprised of three different processes: namely, screenplay settings, renderings, and label generations. This section will elaborate more on each process.

Screenplays

The first process is the screenplay design, which provides the basis for synthetic image generation. The screenplay contains the arrangement of everything that is seen on the scene, including locations, and avatars' actions. To design screenplays, we manually placed animated avatars in different locations of the scene and randomized the location and viewpoint of the camera. Three different construction-related activities of walking, digging, and coordinating are assigned to the workers. Two unique worker avatars are also used in our screenplays. To obtain diversified viewpoints in the synthetic dataset, we randomized the location of the camera within a range of 35 meters from a target worker. The target worker is a specific avatar selected to be tracked by the camera in the scene. It is critical to have the target avatar to ensure the existence of at least a single worker in every rendered image.

Rendering

After setting up screenplays, we simulated the virtual construction scene and rendered various viewpoints from the scene. By randomizing the camera location in the scene and rendering the animation of avatars, we generated 41 screenplays. Each of the screenplays yields image sequences of the animated construction site, from different viewpoints. For the case of this study, we rendered 14 seconds of screenplays at a rate of 10 frames per second, which resulted in 140 image sequences for each screenplay. To obtain as diverse images as possible, we randomized the sun direction and daytime of the recordings. By rendering all of the screenplays, we were able to generate 5,740 images with a click of a single button within two days.

Label generation

Since graphical engines are used in this framework, every rendered image comes with precise information about the scene. Blender's data structure enables the extraction of 3D annotations of visible or occluded body parts, 2D and 3D bounding boxes, instance segmentation, and depth maps (see Figure 2). While this study sufficed to the exploitation of the 2D bounding boxes for worker detection in the scene, there exists a significant potential in using the generated labels.

3.2 Statistical Analysis of the synthetic data

The synthetic dataset of this study is rendered as a Full HD image sequence. Each image contains 5.2 people per frame on average, totalling more than 30K bounding boxes. The distance of the avatars from the camera in the scene ranged from 5 to 100 meters, which resulted in bounding boxes with dimensions from 0.1 to 400 pixels.

3.3 Worker detection model

This study adopts pre-trained YOLOv5 (Jocher et al., 2022) as the worker detection model. Pre-trained YOLOv5 is trained on the COCO dataset (Lin et al., 2014), where the inputs are labelled in 80 object categories. Although the pre-trained YOLO can detect the persons, the process of fine-tuning the model with synthetic images is essential for multiple reasons. Firstly, the major difference of context in construction-related images with the COCO dataset reduces the performance of the pre-trained model in construction scenes. The construction contexts contain images such as equipment and workers and the backgrounds are usually cluttered. However, the COCO images are mostly captured out of the construction sites, with very controlled illumination and backgrounds (Kim et al., 2019). Secondly, the pre-trained network is not compatible with various camera viewpoints. This is due to the limited variation in the COCO dataset. For example, an image of persons captured with a UAV has a completely different appearance and scale than the person presented in the COCO dataset, which in turn, puzzles the convolutional layers and deteriorates the localization performance of the pre-trained model. To incorporate a wider range of viewpoints, this study fine tunned the model with diversified synthetic images. Another essential factor is the overfitting problem. Since the YOLO model has a very high level of complexity and considering that the COCO dataset is not suitable for construction worker detection, there is a need for a huge, labelled image dataset of construction workers to avoid overfitting. Therefore, this study adopts the pre-trained YOLO as the initial point and fine tunned with synthetic images.

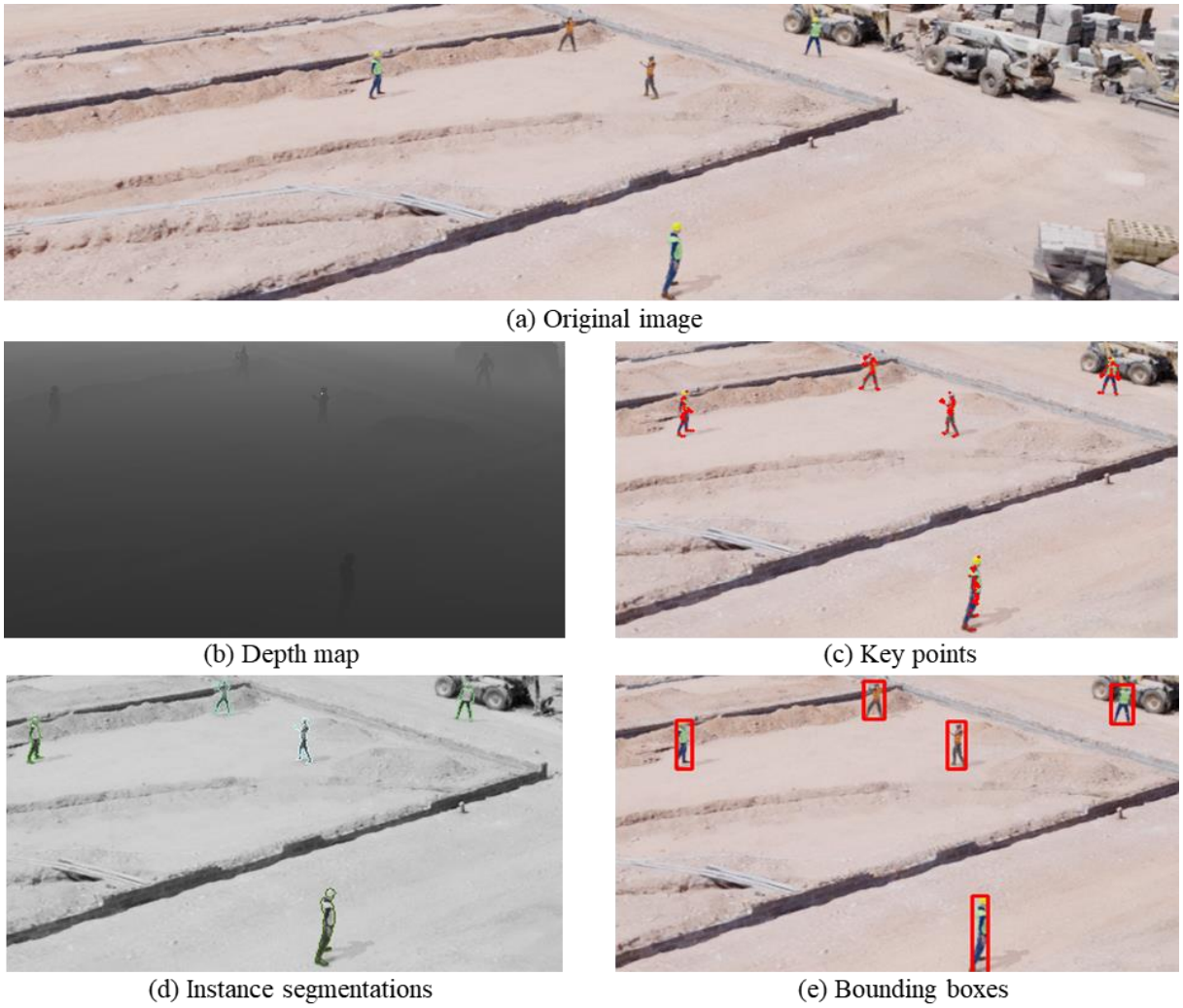

Figure 2. Sample of synthetic data

4. Experimental details and results

The Generalized Intersection of Unions (GIoU) is used as the bounding box regression loss function. GIoU is adapted in YOLOv5 to address the inaccurate calculation of non-overlapping bounding boxes. Further details of the GIoU loss function can be found in (Wang et al., 2021). For training the model, we used Stochastic Gradient Descent (SGD) optimizer. As a fine-tuning process, all of the pre-trained model's layers were completely unfrozen and all the pre-trained weights of the network were used as an initialization point. After unfreezing all the layers, the model is trained for 50 epochs with an initial learning rate of 1e-2 while using cosine annealing the final learning was decreased to 1e-4. The momentum of the optimizer and weight decay was set to 0.937 and 5e-4 respectively. Figure 3 illustrates the learning curve of the model fine-tuned on synthetic data for 50 epochs with a batch size of 4. Decrease of loss by increasing epochs in both training and validation set in the learning curve is an absolute indication that synthetic data enables DNN training.

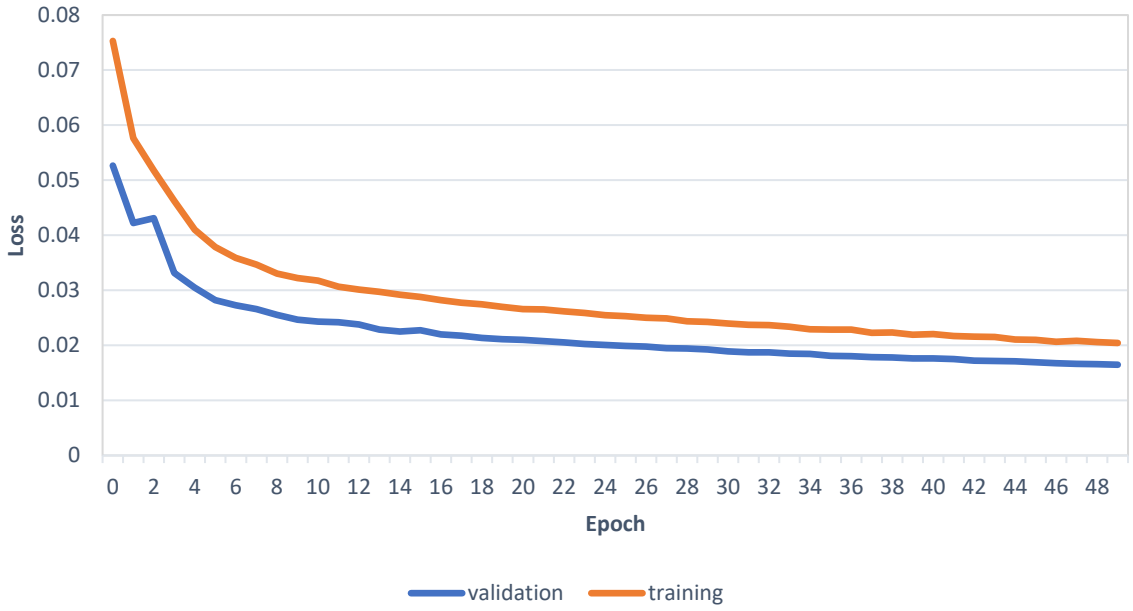

Figure 3. The learning curve of the model trained on the synthetic dataset

For visual verification, this study used the publicly available dataset of Moving Objects in Construction Sites (MOCS) (Xuehui et al., 2021). This dataset contains 41,668 images of 13 categories of construction mobile objects. Concerning the scope of this research, images containing workers are filtered out of the MOCS validation set, which yielded 3,091 images as benchmarking data for this research. Figure 4 illustrates several detection results of the synthetic data-trained model on real construction images (i.e., MOCS validation set). As shown in Figure 4, it successfully detected construction workers in diverse scenarios and at varying scales. Given that this model was trained only with synthetic data, the detection results are an indication of the great potential of synthetic data in DNN training. It should be noted that the fine-tuned model used in this experiment did not reach the optimum performance since it was trained with a very limited amount of synthetic data (i.e., 5,740 images). With the incorporation of a wider range of 3D backgrounds, worker motions, and avatar colours in data generation, the model would have another chance to improve its performance. Our follow-up study will address it, establishing a complete training set of synthetic construction images, and conducting a quantitative evaluation on real test images. This study will lay a stepping stone to non-real but real-looking image-driven DNN training.

Figure 4. Sample prediction for the model trained on synthetic images

5. Conclusion

The productivity and safety of the construction industry have continuously experienced staggering challenges, while the widespread adoption of DNN-based visual AI revolutionized other industries. DNN-powered automation and digitization offer promising solutions; however, the unavailability of high-quality and diversified data prohibitively hinders the construction industry from reaping the benefits. To address the time-consuming manual data collection process and avoid data privacy concerns, this study proposes a novel computational framework that enables synthetic data generation for DNN training. The proposed framework leverages the power of graphical engines to develop a virtual replica of the construction site and renders non-real but real-looking images. To increase the diversity of the synthetic dataset, the proposed framework randomizes critical features of the scene, such as the worker's pose, clothing, camera viewpoint, lighting condition, and sun direction. Although numerous questions remained intact, the results of this study indicated the potential and applicability of computer-generated synthetic images in the development of high-quality, field applicable DNNs. Further research should be conducted to address critical questions such as understanding DNNs' behaviour when trained with synthetic data, exploring the impact of synthetic image realism on DNN performance, examining the feasibility of training DNNs for pose estimation, semantic segmentation, depth estimation and activity recognition, and determining the optimal size of the synthetic dataset. Moreover, the creation of a virtual construction site involved the manual placement of avatars in pre-defined locations in the scene. Although this procedure can be automated, this study limited its scope to the manual placement of avatars in the scene. Leveraging the agent-based simulations can be a suitable approach toward automation of the avatar placement in the scene.

References

Acharya, D., Khoshelham, K., Winter, S. (2019). BIM-PoseNet: Indoor camera localisation using a 3D indoor model and deep learning from synthetic images. ISPRS J Photogramm 150, 245–258. https://doi.org/10.1016/j.isprsjprs.2019.02.020

Akinosho, T.D., Oyedele, L.O., Bilal, M., Ajayi, A.O., Delgado, M.D., Akinade, O.O., Ahmed, A.A. (2020). Deep learning in the construction industry: A review of present status and future innovations. J. Build. Eng. 32, 101827. https://doi.org/10.1016/j.jobe.2020.101827

Alvanchi, A., TohidiFar, A., Mousavi, M., Azad, R., Rokooei, S. (2021). A critical study of the existing issues in manufacturing maintenance systems: Can BIM fill the gap? Comput Ind 131, 103484. https://doi.org/10.1016/j.compind.2021.103484

Amato, G., Ciampi, L., Falchi, F., Gennaro, C., Messina, N. (2019). Learning Pedestrian Detection from Virtual Worlds, in: Ricci, E., Rota Bulò, S., Snoek, C., Lanz, O., Messelodi, S., Sebe, N. (Eds.), Image Analysis and Processing – ICIAP, Springer International Publishing, Cham, pp. 302–312.

Assadzadeh, A., Arashpour, M., Brilakis, I., Ngo, T., Konstantinou, E. (2022). Vision-based excavator pose estimation using synthetically generated datasets with domain randomization. Autom Constr 134, 104089. https://doi.org/10.1016/j.autcon.2021.104089

Ayodele, T., Kahimo-shakantu, K. (2021). Data Sharing in the Construction Industry: An Assessment of Stakeholders' Perception, in: International Conference on Infrastructure Development and Investment Strategies for Africa.

Bang, S., Park, S., Kim, Hongjo, Kim, Hyoungkwan (2019). Encoder–decoder network for pixel-level road crack detection in black-box images. Comput-Aided Civ Inf 34, 713–727. https://doi.org/10.1111/mice.12440

Blender (2022). URL https://www.blender.org/ (accessed 10 April 2022).

Bousmalis, K., Silberman, N., Dohan, D., Erhan, D., Krishnan, D. (2017). Unsupervised pixel-level domain adaptation with generative adversarial networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3722–3731.

Chien, C.-F., Dauzère-Pérès, S., Huh, W.T., Jang, Y.J., Morrison, J.R. (2020). Artificial intelligence in manufacturing and logistics systems: algorithms, applications, and case studies. Int. J. Prod. Res. 58, 2730–2731. https://doi.org/10.1080/00207543.2020.1752488

Delgado, J.M.D., Oyedele, L. (2021). Deep learning with small datasets: using autoencoders to address limited datasets in construction management. Appl. Soft Comput. 112, 107836. https://doi.org/10.1016/j.asoc.2021.107836

Dixit, S., Mandal, S.N., Thanikal, J.V., Saurabh, K. (2019). Evolution of studies in construction productivity: A systematic literature review (2006–2017). Ain Shams Eng. J. 10, 555–564. https://doi.org/10.1016/j.asej.2018.10.010

Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., Van Der Smagt, P., Cremers, D., Brox, T. (2015). Flownet: Learning optical flow with convolutional networks, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 2758–2766.

Fabbri, M., Brasó, G., Maugeri, G., Cetintas, O., Gasparini, R., Ošep, A., Calderara, S., Leal-Taixé, L., Cucchiara, R. (2021). MOTSynth: How Can Synthetic Data Help Pedestrian

Detection and Tracking?, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10849–10859.

Grosse, K., Lee, T., Biggio, B., Park, Y., Backes, M., Molloy, I. (2021). Backdoor Smoothing: Demystifying Backdoor Attacks on Deep Neural Networks.

Hong, Y., Park, S., Kim, Hongjo, Kim, Hyoungkwan (2021). Synthetic data generation using building information models. Autom. Constr. 130, 103871. https://doi.org/10.1016/j.autcon.2021.103871

Hwang, S., Jebelli, H., Choi, B., Choi, M., Lee, S. (2018). Measuring Workers' Emotional State during Construction Tasks Using Wearable EEG. J. Constr. Eng. Manag. 144, 04018050. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001506

Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., NanoCode012, Kwon, Y., TaoXie, Fang, J., imyhxy, Michael, K., Lorna, V, A., Montes, D., Nadar, J., Laughing, tkianai, yxNONG, Skalski, P., Wang, Z., Hogan, A., et al. (2022). ultralytics/yolov5: v6.1 - TensorRT, TensorFlow Edge TPU and OpenVINO Export and Inference. https://doi.org/10.5281/zenodo.6222936

Kim, D., Liu, M., Lee, S., Kamat, V.R. (2019). Remote proximity monitoring between mobile construction resources using camera-mounted UAVs. Autom Constr 99, 168–182. https://doi.org/10.1016/j.autcon.2018.12.014

Kim, Dahun, Woo, S., Lee, J.-Y., Kweon, I.S. (2020). Video panoptic segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9859–9868.

Kim, Daeho, Lee, S., Kamat, V.R. (2020). Proximity Prediction of Mobile Objects to Prevent Contact-Driven Accidents in Co-Robotic Construction. J. Comput. Civ. Eng. 34, 04020022. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000899

Kim, Hongjo, Kim, Hyoungkwan (2018). 3D reconstruction of a concrete mixer truck for training object detectors. Autom. Constr. 88, 23–30. https://doi.org/10.1016/j.autcon.2017.12.034

Koops, B.-J. (2014). The trouble with European data protection law. Int. Data Priv. Law 4, 250–261. https://doi.org/10.1093/idpl/ipu023

Laufer Alexander, Shapira Aviad, Telem Dory (2008). Communicating in Dynamic Conditions: How Do On-Site Construction Project Managers Do It? J Manage Eng 24, 75–86. https://doi.org/10.1061/(ASCE)0742-597X(2008)24:2(75)

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L. (2014). Microsoft COCO: Common Objects in Context, in: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.), European conference on computer vision (ECCV), Springer International Publishing, Cham, Switzerland, pp. 740–755.

Luo, H., Wang, M., Wong, P.K.-Y., Cheng, J.C.P. (2020). Full body pose estimation of construction equipment using computer vision and deep learning techniques. Autom. Constr. 110, 103016. https://doi.org/10.1016/j.autcon.2019.103016

Luo, X., Li, H., Cao, D., Yu, Y., Yang, X., Huang, T. (2018). Towards efficient and objective work sampling: Recognizing workers' activities in site surveillance videos with two-stream convolutional networks. Autom. Constr. 94, 360–370. https://doi.org/10.1016/j.autcon.2018.07.011

Ma, J.W., Czerniawski, T., Leite, F. (2020). Semantic segmentation of point clouds of building interiors with deep learning: Augmenting training datasets with synthetic BIM-based point clouds. Autom. Constr. 113, 103144. https://doi.org/10.1016/j.autcon.2020.103144

Mahmood, B., Han, S., Seo, J. (2022). Implementation experiments on convolutional neural network training using synthetic images for 3D pose estimation of an excavator on real images. Autom. Constr. 133, 103996. https://doi.org/10.1016/j.autcon.2021.103996

Neuhausen, M., Herbers, P., König, M. (2020). Using Synthetic Data to Improve and Evaluate the Tracking Performance of Construction Workers on Site. Appl. Sci. 10, 4948. https://doi.org/10.3390/app10144948

Neythalath, N., Søndergaard, A., Bærentzen, J.A. (2021). Adaptive robotic manufacturing using higher order knowledge systems. Autom. Constr. 127, 103702. https://doi.org/10.1016/j.autcon.2021.103702

Oxford Economics (2021). Future of Construction: A Global Forecast for Construction to 2030. Oxford Economics. https://www.oxfordeconomics.com/resource/future-of-construction/

Rao, T.V.N., Gaddam, A., Kurni, M., Saritha, K. (2022). Reliance on Artificial Intelligence, Machine Learning and Deep Learning in the Era of Industry 4.0, in: Smart Healthcare System Design, John Wiley & Sons, Ltd, pp. 281–299. https://doi.org/10.1002/9781119792253.ch12

Ristani, E., Solera, F., Zou, R., Cucchiara, R., Tomasi, C. (2016). Performance Measures and a Data Set for Multi-target, Multi-camera Tracking, in: Hua, G., Jégou, H. (Eds.), European conference on computer vision (ECCV), Springer International Publishing, Cham, pp. 17–35.

Ros, G., Sellart, L., Materzynska, J., Vazquez, D., Lopez, A.M. (2016). The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3234–3243.

Shorten, C., Khoshgoftaar, T.M. (2019). A survey on Image Data Augmentation for Deep Learning. J. Big Data 6, 60. https://doi.org/10.1186/s40537-019-0197-0

Soltani, M., Zhu, Z., Hammad, A. (2018). Framework for Location Data Fusion and Pose Estimation of Excavators Using Stereo Vision. J. Comput. Civ. Eng. 32, 04018045. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000783

Wang, Z., Wu, Y., Yang, L., Thirunavukarasu, A., Evison, C., Zhao, Y. (2021). Fast Personal Protective Equipment Detection for Real Construction Sites Using Deep Learning Approaches. Sensors 21. https://doi.org/10.3390/s21103478

Xiao, B., Kang, S.-C. (2021). Development of an Image Data Set of Construction Machines for Deep Learning Object Detection. J. Comput. Civ. Eng. 35, 05020005. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000945

Xuehui, A., Li, Z., Zuguang, L., Chengzhi, W., Pengfei, L., Zhiwei, L. (2021). Dataset and benchmark for detecting moving objects in construction sites. Autom. Constr. 122, 103482. https://doi.org/10.1016/j.autcon.2020.103482

Zhou, Z., Goh, Y.M., Li, Q. (2015). Overview and analysis of safety management studies in the construction industry. Saf. Sci. 72, 337–350. https://doi.org/10.1016/j.ssci.2014.10.006