

# UTILIZING 360-DEGREE IMAGES FOR SYNTHETIC DATA GENERATION IN CONSTRUCTION SCENARIOS

*Aqsa Sabir*

Department of Computer Science and Engineering, Chung Ang University, Seoul South Korea

*Rahat Hussain, Syed Farhan Alam Zaidi, Akeem Pedro, Mehrtash Soltani, Dongmin Lee\* & Chansik Park*

ConTI Lab, Department of Architectural Engineering, Chung-Ang University, Seoul, South Korea

*\*Corresponding Author*

**ABSTRACT:** *Computer vision-based safety monitoring requires machine learning models trained on generalized datasets covering various viewpoints, surface properties, and lighting conditions. However, capturing high-quality and extensive datasets for some construction scenarios is challenging at real job sites due to the risky nature of construction scenarios. Previous methods have proposed synthetic data generation techniques involving 2D background randomization with virtual objects in game-based engines. While there has been extensive work on utilizing 360-degree images for various purposes, no study has yet employed 360-degree images for generating synthetic data specifically tailored for construction sites. To improve the synthetic data generation process, this study proposes a 360-degree images-based synthetic data generation approach using Unity 3D game engine. The approach efficiently generates a sizable dataset with better dimensions and scaling, encompassing a range of camera positions with randomized lighting intensities. To check the effectiveness of our proposed method, we conducted a subjective evaluation, considering three key factors: object positioning, scaling in terms of object respective size, and the overall size of the generated dataset. The synthesized images illustrate the visual improvement in all three factors. By offering an improved data generation method for training safety-focused computer vision models, this research has the potential to significantly enhance the automation of the construction safety monitoring process, and hence, this method can bring substantial benefits to the construction industry by improving operational efficiency and reinforcing safety measures for workers.*

**KEYWORDS:** *360-Degree Images, Computer Vision, Synthetic Data Generation, Game Engine, Object Detection, Construction Safety Monitoring*

## 1. INTRODUCTION

Effectively monitoring construction sites in real-time demands robust computer vision (CV) models trained on diverse datasets capturing different viewpoints, surface properties, and lighting conditions (Li et al., 2022; Sami Ur Rehman et al., 2022). This diversity ensures accurate and adaptable surveillance for improved safety and efficiency. However, obtaining such extensive and high-quality datasets poses significant challenges. Furthermore, addressing the complexities of dynamic construction environments with numerous elements and rapidly changing conditions is a recognized challenge. To address these issues, the utilization of smart devices for automated progress and real-time construction monitoring through object detection technology and positional data presents a rapid and precise solution (Rho et al., 2020).

Previous methods attempted to address this issue propose synthetic data generation techniques by introducing 2D background randomization with virtual objects in game-based engines (Zhang et al., 2022). While these techniques have shown promise in controlled settings, they still exhibit shortcomings in dynamic construction work environments (Lee et al., 2022). This limitation might arise from the complexity of recreating the intricate interactions among dynamic elements and the complex spatial relationships found in construction sites. Capturing the subtle aspects of depth perception and object occlusion, which are essential for precise object detection, can pose a challenge when working with synthetic datasets (Choi & Pyun, 2021).

This study presents a pioneering method utilizing 360-degree images to craft synthetic data, leveraging the Unity 3D game engine, to comprehensively address these challenges. By adopting this approach, we aim to bridge the gap between synthetic and real-world data, enabling effective training of CV models for construction scenarios. The proposed method efficiently generates a sizable dataset with better dimensions and scaling, encompassing a range of camera positions with day and night lighting intensities. This not only facilitates improved object detection but also opens avenues for applications in construction progress monitoring and site safety analysis. To validate the proposed method, we conducted a thorough analysis, encompassing the examination of three crucial factors: Object Position, objects' size scaling, and dataset size.

Object positioning factor scrutinizes the precision with which objects are localized within the synthetic data. It involves a meticulous assessment of the alignment and placement of objects in relation to their real-world counterparts. Higher accuracy in object positioning indicates closer adherence to reality, crucial for applications demanding precise spatial understanding. Object size scaling refers to the scale at which objects are represented within synthetic data, a critical dimension. This factor entails a meticulous examination of whether the size relationships between objects are faithfully replicated. Accurate scaling ensures that the synthetic data mirrors the real-world environment, influencing tasks where object proportions are of paramount importance. Dataset size describes the magnitude of the dataset generated through the proposed technique, an instrumental metric in its efficacy. A larger dataset not only provides a more comprehensive representation of the construction scenario but also potentially enhances the performance of machine learning models. It enables a broader spectrum of scenarios to be captured, invaluable in training robust computer vision models.

Following the introduction and background section, which provide a thorough overview of the research problem, Section 2 offers an extensive literature review. Section 3 then elaborates on the proposed synthetic data generation method. Section 4 performs a comparative analysis between the proposed approach and conventional methodologies. Finally, the paper concludes by summarizing its contributions and delineating potential avenues for future research.

## 2. LITERATURE REVIEW

The construction industry encounters formidable challenges in procuring authentic image data, largely stemming from the perilous and dynamic nature of construction sites. In response, synthetic data generated through computer graphics emerges as a promising and cost-effective solution for training machine learning models in the construction sector. This approach proves especially conducive to tasks such as site monitoring, defect detection, and material classification, endowing models with the capability to derive invaluable insights. Consequently, this enhances the efficiency and efficacy of construction processes and outcomes in practical, real-world scenarios.

A suite of techniques, including computer graphics, data simulation, data augmentation, generative adversarial networks (GANs), and synthetic-to-real transfer learning, collectively contribute to the generation of synthetic datasets. Among these, computer graphics entails the creation of three-dimensional (3D) models of objects and scenes through specialized software platforms like Blender, Autodesk 3ds Max, Maya, or the Unity Game Engine. This process culminates in the rendering of these models to generate two-dimensional (2D) images (Frolov et al., 2022). The advantage lies in the controlled environment it affords for data generation, enabling the creation of bespoke datasets with predefined attributes, including controlled variations in lighting, camera angles, or object orientations (Oh et al., 2021; Wong et al., 2019). Data simulation, on the other hand, involves the emulation of the underlying data generation processes and interrelationships between variables. This endeavor yields synthetic images that closely mimic the appearance and geometric characteristics of their real-world counterparts. Concurrently, generative adversarial networks (GANs) represent a neural network architecture comprising a generator network and a discriminator network. This setup empowers the generator to create novel images while the discriminator distinguishes between generated and authentic images, culminating in adversarial training. Moreover, synthetic-to-real transfer learning encompasses the training of a model on synthetic data, followed by fine-tuning on authentic data. This iterative process engenders synthetic data that closely approximates real-world data. These methodologies, whether utilized in isolation or synergy, facilitate the generation of expansive and diverse synthetic datasets for training and assessing machine learning models within the construction domain.

In 2015, Soltani et al. conducted pioneering work in synthetic data generation for excavation tasks, utilizing Autodesk 3ds Max and Google SketchUp in conjunction with Histogram of Gradient (HOG) transformations for precise segmentation (Soltani et al., 2016). Since then, a surge of studies, particularly post-2020, has significantly advanced this field. Neuhausen et al., Kim et al., and Tohidifar et al., turned to Blender software, leveraging its capabilities in constructing intricate 3D models of workers (J. Kim et al., 2022; Neuhausen et al., 2020; Tohidifar et al., 2022). Similarly, Kim et al. extended Blender's utility by creating synthetic images of scaffolds through the integration of point clouds (A. Kim et al., 2022). Barrera-Animas et al., introduced an innovative methodology that combines Blender-based synthetic image generation with automatic labeling, effectively surmounting limitations associated with separate processes (Barrera-Animas & Davila Delgado, 2023).

Wei et al. harnessed Building Information Modeling (BIM) software to create comprehensive 3D models of buildings at various stages of construction. These models were seamlessly integrated with Unreal Engine, facilitating the generation of a synthetic dataset exhibiting diverse light conditions and viewing perspectives (Wei & Akinci, 2022). Sutjaritvorakul et al. and Siu et al. capitalized on Unreal Engine's capabilities to generate images

of workers and sewerage pipes, respectively, catering to object detection through closed-circuit television (CCTV) cameras (Siu et al., 2022; Sutjaritvorakul et al., 2020).

More recently, there has been a notable surge in the utilization of the Unity Perception package for synthetic data generation. Another study harnessed Unity Perception to intricately craft building facades in Koto City, employing manual annotation techniques involving polygons and bounding boxes (Zhang et al., 2022). Similarly, few more efforts towards generating datasets tailored for construction workers and excavation tasks, implemented bounding box auto-annotation methods (Assadzadeh et al., 2022; Lee et al., 2022). Notably, the Unity Perception package was initially introduced by Borkman et al. for synthetic dataset creation in computer vision applications (Borkman et al., 2021). This package emerged as an invaluable resource in generating expansive synthetic data sets for our machine learning models. Its rich array of features prompted us to explore its potential with 360-degree images within the construction domain. Noteworthy functionalities such as the Perception Camera and robust data labeling options facilitated the creation of a diverse and highly accurate dataset, crucial for training effective machine learning models. Indeed, earlier studies were constrained by their narrow focus and reliance on manual annotation methods, which limited their scalability and real-world applicability. Moreover, challenges in faithfully replicating dynamic construction environments remained unaddressed. This study pioneers 360-degree image and Unity 3D-based synthetic data generation for construction scenarios, overcoming challenges, and improving object detection for safety and monitoring.

### 3. METHODOLOGY

The research process outlined in this paper is delineated into five distinct stages, as illustrated in Figure 1. The research process began with Data Collection and Preprocessing, where real-world construction site imagery was gathered and prepared for further analysis. Subsequently, Virtual Construction Site Simulation was conducted to create a digital representation mirroring real-world scenarios. This simulation served as the foundation for generating annotated data through the integration of the Unity Perception package. The package's capabilities, including Perception Camera and data labeling options, played a pivotal role in producing a diverse and accurately annotated dataset. Finally, the Results and Evaluation section scrutinized the performance of the machine learning models trained on the synthetic data, providing valuable insights into their effectiveness in dynamic construction environments. Furthermore, Figure 2 illustrates the complete system architecture.

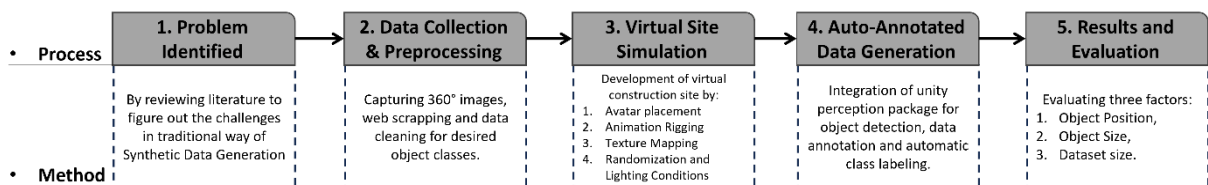


Fig 1: Research Process

#### 3.1 Data collection and preprocessing

The initial phase of constructing our synthetic dataset involved the acquisition of authentic 360-degree images from real-world construction sites. This imperative task was accomplished through on-site visits to various construction locations, where we employed a Ricoh Theta V camera to capture high-fidelity images. A comprehensive set of approximately 49 shots was meticulously taken across three distinct sites. These captures were strategically dispersed across three distinct sites, a deliberate maneuver aimed at ensuring a panoramic representation of both indoor and outdoor contexts. This strategic diversity not only broadens the dataset's scope but also fortifies its ecological fidelity, essential for simulating real-world scenarios accurately.

To uphold the dataset's integrity and applicability, a discerning filtering process was meticulously implemented. This process entailed the deliberate exclusion of instances featuring the target classes or objects slated for subsequent annotation. Moreover, as a supplementary measure to bolster the dataset, an additional 20 images were judiciously selected from reputable online sources. These images underwent a rigorous refinement process, facilitated by Adobe Photoshop, a widely recognized image editing tool. This intervention was crucial in the removal of extraneous elements that might have inadvertently compromised the dataset's fidelity. This meticulous

curation was necessitated by the publicly accessible nature of these images.

By rigorously adhering to these meticulous acquisition and preprocessing protocols, we established a robust foundation for the subsequent stages of annotation and synthetic data generation. This methodological rigor ensures that the dataset remains of the highest quality and relevance, serving as an indispensable resource for the accurate training and evaluation of our computer vision models in dynamic construction environments.

### **3.2 Virtual construction site simulation:**

During this pivotal phase, the synthesis of virtual components was meticulously undertaken to construct a dynamic virtual construction site. This critical stage comprised four foundational elements, each of which played a significant role in enhancing the authenticity and diversity of the synthetic dataset. A situation as a case scenario where workers are actively engaged in ladder climbing was meticulously composed to clearly capture the real atmosphere of a construction site environment.

#### **3.2.1 Avatar Placement:**

The foundational step involved generating 3D avatars with distinct physical attributes and appearances through Ready Player Me. To augment avatar diversity, human parts from SketchFab were incorporated, with customization of attire and features. Avatars were tailored to emulate team members' characteristics using the Unity 3D avatar maker plugin, and adjustments to facial features and joints were made using Blender and Unity 3D. Six avatars were strategically positioned within the hollow cylinder structure. This included four male workers alternating between two ladders switching on/off at predetermined intervals. Helmets, a safety precaution, were optionally worn during simulation. Dataset diversity was amplified by introducing random alterations in clothing color, pattern, skin complexion, and hair shade. This amalgamation of diverse and realistic avatars not only diversifies the synthetic dataset but also augments its capacity to simulate a broad spectrum of personnel scenarios within a construction site environment.

#### **3.2.2 Animation Rigging:**

In the subsequent phase, a climbing animation was seamlessly integrated into the avatars within the unity game engine. A pre-existing climbing animation from Mixamo<sup>1</sup> was utilized and imported into the Unity engine. The animation was then rigged to the avatars, with meticulous adjustment of parameters to ensure perfect synchronization with the ladder model. The seamless integration of climbing animations fortifies the dataset by providing a realistic representation of ladder-related actions. This, in turn, equips ML models with the proficiency to accurately identify and classify such actions within the construction site environment.

#### **3.2.3 Texture Mapping:**

The Texture Mapping phase was dedicated to seamlessly enveloping the 360-degree background images around the virtual cylinder. Distinct materials were assigned to each image, with meticulous organization within a designated folder named MaterialsBG in the Unity project. To infuse the synthetic data with variety, a script was implemented to randomly select materials from the MaterialsBG folder, generating a diverse array of image frames. This strategic process played a pivotal role in cultivating a truly immersive virtual environment. The judicious mapping of the 360-degree images onto the surface of the cylinder heightened the authenticity of the synthetic dataset, furnishing it with meticulously curated training data of superior quality for ML models.

#### **3.2.4 Randomization and Lighting Conditions:**

Dynamic lighting effects were harnessed to infuse our virtual construction site with an authentic and immersive ambiance. The Unity engine's built-in directional light was adeptly employed to replicate a spectrum of natural lighting conditions, spanning dawn, noon, afternoon, and night. A meticulously crafted script dynamically modulated the light intensity, mimicking the natural progression of light over time. This meticulous attention to lighting dynamics further fortified the efficacy of the synthetic dataset, as CV based techniques rely on precise lighting cues for accurate object identification and localization.

---

<sup>1</sup> <https://www.mixamo.com/#/>

### 3.3 Generating annotated data through unity perception PACKAGE INTEGRATION:

The incorporation of the Unity Perception Package stands as a transformative stride in elevating the authenticity and intricacy of our synthetic dataset. This comprehensive toolset within the Unity environment streamlines the process of automatic labeling and annotation of objects based on pre-defined classes, a critical step for training computer vision models.

At the core of this integration lies the Camera Sensor component. This crucial element faithfully simulates the behavior of a real-world camera, capturing images from the scene with a level of fidelity that closely mirrors actual visual acquisition. This technical facet not only enriches the dataset's realism but also empowers our models to process images in a manner akin to their real-world counterparts. Complementing the Camera Sensor is the Labeler Component, an indispensable tool for training computer vision models. This component meticulously annotates objects within the scene, a vital step in providing the models with ground truth information for accurate object detection and classification. This facet significantly enhances the dataset's utility in training robust models capable of discerning and categorizing objects within complex construction environments. Moreover, the Renderer Component stands as another critical element in this integration. This component harnesses the Unity Perception Package's neural rendering pipeline to meticulously render the scene. By doing so, it imbues the generated images with a level of visual fidelity that is essential for training models to recognize and understand complex, real-world scenarios.

The Labeler Component acts like a foundation in the process of training computer vision models. It interfaces seamlessly with a predefined set of classes and automatically identifies objects within the scene, assigning them appropriate labels based on their class membership. This automated labeling process significantly amplifies the dataset's efficacy in training robust models capable of discerning and categorizing objects within the dynamic and intricate environments of construction sites. The Labeler Component operates in tandem with scripts written in C#, leveraging the Perception API provided by the package. These scripts access object information and apply labels to them based on their class. The API includes functions and methods that facilitate this automatic annotation process. Through the orchestrated interplay of the Labeler Component, Perception API, and scripts, our synthetic dataset gains a precise and detailed ground truth, indispensable for training computer vision models. This annotated data forms the cornerstone of our dataset, empowering our models in object detection and environmental perception tasks within the complex and safety-critical context of construction sites. Through the orchestrated interplay of the Labeler Component, Perception API, and scripts, our synthetic dataset gains a precise and detailed ground truth, indispensable for training computer vision models. This annotated data forms the cornerstone of our dataset, empowering our models in object detection and environmental perception tasks within the complex and safety-critical context of construction sites.

As a result of this integration, the Unity Perception Package generates several key JSON files alongside the image dataset. These files, including 'annotation\_definitions.json', 'captures.json', 'metric\_definitions.json', 'metrics.json', and 'sensors.json', hold crucial annotation data, capture details, metric definitions, recorded metrics, and sensor information respectively. They play a pivotal role in enriching our dataset for object detection model training and evaluation. The annotation\_definitions.json file, for instance, provides detailed information about class labels and object attributes, while captures.json contains information regarding the captured scenes. These JSON files collectively form the backbone of our dataset, empowering our models with the necessary information to accurately perceive and interpret the dynamic construction environment.

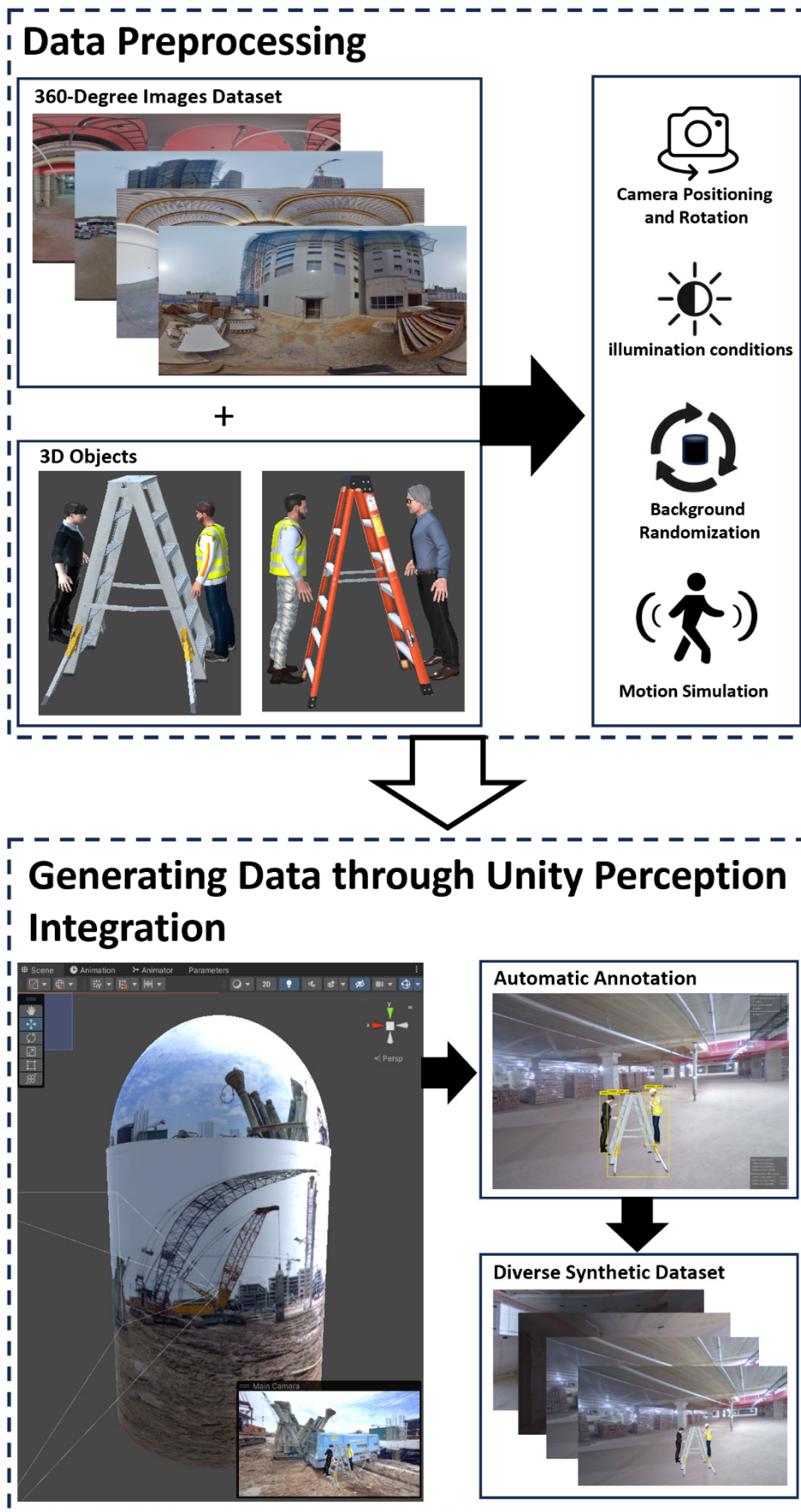


Fig 2: System Architecture explaining whole process of synthetic dataset generation.

## 4. RESULTS

The synthetic data generated through the proposed approach utilizing 360-degree images was focused on three critical factors: Object Position (Accuracy), Object Size (Scaling), and Dataset Size. The results from proposed techniques are shown in Figure 3.

### 4.1.1 Object Positioning:

Upon thorough examination, it became evident that the synthetic data derived from 360-degree images demonstrated excellent accuracy in object position determination. The comprehensive perspective offered by 360-degree images facilitated more precise object localization within the construction environment. This was particularly notable in instances where objects were positioned in close proximity or within complex spatial configurations. The enhanced depth perception afforded by the spherical perspective of 360-degree imagery played a pivotal role in accurately pinpointing object locations. Moreover, the capability to observe objects from diverse angles within a single image frame contributed to a deeper comprehension of their spatial relationships within the construction site. This heightened accuracy in object positioning is a critical advantage, especially in safety-critical environments where precise object localization is paramount for accident prevention and worker safety.

### 4.1.2 Object's Size Scaling:

One of the notable advantages of leveraging 360-degree images was the observed improvement in object size scaling. The spherical perspective offered by 360-degree imagery allowed for a more precise representation of object sizes in relation to their immediate surroundings because it encompassed a comprehensive view of the entire environment without the distortions. This panoramic view allowed for a faithful portrayal of how objects interacted within the construction site, aligning closely with their real-world proportions. This was especially pertinent when considering objects from varying angles or perspectives. The spherical view granted by 360-degree images mitigated distortions, ensuring that object sizes were accurately depicted regardless of the viewing angle. In contrast, data generated from 2D background images often grappled with limitations in faithfully representing object sizes, particularly when objects were viewed from non-standard angles. Additionally, the ability to view objects from multiple angles in a single image frame further contributed to this enhanced accuracy in size representation. This critical improvement in object size scaling with 360-degree imagery contributes significantly to the overall fidelity and accuracy of the synthetic dataset, ultimately enhancing the effectiveness of computer vision models in interpreting real-world construction scenarios.

### 4.1.3 Dataset Size:

The size of dataset played a pivotal role in influencing the performance of computer vision models. The use of 360-degree images facilitated the creation of a larger dataset. A significant contributor to this considerable dataset expansion is the unique capability of 360-degree images. A single 360-degree image can generate multiple frames, each offering different perspectives while maintaining consistent lighting and camera positions. This efficiency greatly enhances the adaptability and effectiveness of models in real-world construction site scenarios. The substantial dataset derived from 360-degree images significantly broadened the ability of models to generalize and make well-informed decisions across a wide range of construction site scenarios. Undoubtedly, this rich training data strengthened the robustness and competence of models, establishing it as a valuable tool for real-time construction site monitoring.


Factor	360-Degree Images
Object Position	<p style="text-align: center;">Higher Accuracy</p> 
Object Size Scaling	<p style="text-align: center;">Better Scaling</p> 
Data Size	<p style="text-align: center;">Larger Dataset</p> 

Fig 3: Results of the proposed approach.

## 5. DISCUSSION AND CONCLUSION

This study introduces a 360-degree image-based approach for generating synthetic data to train CV models for construction site safety monitoring. The 360-degree dataset excels in object positioning due to its ability to provide a comprehensive view of the construction site from multiple perspectives. This advantage is particularly valuable for precise object localization in complex spatial arrangements common in construction settings. Additionally, the spherical perspective inherent in 360-degree images ensures a more accurate representation of object sizes relative to their surroundings. This accuracy is crucial for effective monitoring in construction environments, as it aligns object proportions faithfully with real-world dimensions. Furthermore, the larger dataset size derived from 360-degree images offers extensive and diverse training data, enabling the model to generalize effectively across various construction scenarios. Unlike 2D images that produce only one synthetic image for one scenario, each 360-degree image can generate multiple synthetic images for one scenario.

The advantages of employing 360-degree images as the foundation for synthetic data generation are multifold. The all-encompassing view offered by 360-degree imagery allowed for a more holistic representation of the construction environment. This comprehensive perspective, combined with the ability to accurately depict object positions and sizes, resulted in a dataset that offered a remarkably realistic simulation of real-world scenarios.

It is important to acknowledge that while this approach demonstrates notable progress, it does have certain limitations. A limitation of using 360-degree imagery, versus prior 2D background images, is handling dynamic elements on construction sites. While offering a broad view, 360-degree images may introduce distortion, especially at the edges, impacting object detection accuracy, especially for peripheral objects. Addressing and correcting this distortion during data generation is crucial for precise computer vision model training. Future research aims to expand the dataset, explore advanced model architectures, and assess data efficacy through



computer vision-based model training on actual construction sites. Integration of multimodal information, like combining imagery with LiDAR or other sensors, will enrich the dataset and enhance object detection accuracy. Furthermore, we will objectively measure the effectiveness of the proposed approach by conducting computer vision-based model training and comparing the results with the traditional method of generating synthetic data. This approach holds the potential to serve as a foundational framework for numerous industries, offering a blueprint for the modernization of their safe operations, thereby propelling safety standards and operational prowess to new heights on an international scale.

## ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2022R1A2B5B02002553) and the "National R&D Project for Smart Construction Technology (No.23SMIP-A158708-04)" funded by the Korea Agency for Infrastructure Technology Advancement under the Ministry of Land, Infrastructure and Transport, and managed by the Korea Expressway Corporation.

## REFERENCES

- Assadzadeh, A., Arashpour, M., Brilakis, I., Ngo, T., & Konstantinou, E. (2022). Vision-based excavator pose estimation using synthetically generated datasets with domain randomization. *Automation in Construction*, *134*, 104089. <https://doi.org/10.1016/j.autcon.2021.104089>
- Barrera-Animas, A. Y., & Davila Delgado, J. M. (2023). Generating real-world-like labelled synthetic datasets for construction site applications. *Automation in Construction*, *151*, 104850. <https://doi.org/10.1016/j.autcon.2023.104850>
- Borkman, S., Crespi, A., Dhakad, S., Ganguly, S., Hogins, J., Jhang, Y.-C., Kamalzadeh, M., Li, B., Leal, S., Parisi, P., Romero, C., Smith, W., Thaman, A., Warren, S., & Yadav, N. (2021). *Unity Perception: Generate Synthetic Data for Computer Vision*. <https://doi.org/https://doi.org/10.48550/arXiv.2107.04259>
- Choi, W., & Pyun, S. (2021). Synthetic Training Data Generation for Fault Detection Based on Deep Learning. *Geophysics and Geophysical Exploration*, *24*(3), 89–97. <https://doi.org/https://doi.org/10.7582/GGE.2021.24.3.089>
- Frolov, V., Faizov, B., Shakhuro, V., Sanzharov, V., Konushin, A., Galaktionov, V., & Voloboy, A. (2022). Image Synthesis Pipeline for CNN-Based Sensing Systems. *Sensors*, *22*(6), 2080. <https://doi.org/10.3390/s22062080>
- Kim, A., Lee, K., Lee, S., Song, J., Kwon, S., & Chung, S. (2022). Synthetic Data and Computer-Vision-Based Automated Quality Inspection System for Reused Scaffolding. *Applied Sciences*, *12*(19), 10097. <https://doi.org/10.3390/app121910097>
- Kim, J., Kim, D., Shah, J., & Lee, S. (2022). Training a Visual Scene Understanding Model Only with Synthetic Construction Images. *Computing in Civil Engineering* *2021*, 221–229. <https://doi.org/10.1061/9780784483893.028>
- Lee, J. G., Asce, S. M., Hwang, J., Chi, S., Asce, M., & Seo, J. (2022). Synthetic Image Dataset Development for Vision-Based Construction Equipment Detection. *Computing in Civil Engineering*, *36*(5). [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0001035](https://doi.org/10.1061/(ASCE)CP.1943-5487.0001035)
- Li, Y., Wei, H., Han, Z., Jiang, N., Wang, W., & Huang, J. (2022). Computer Vision-Based Hazard Identification of Construction Site Using Visual Relationship Detection and Ontology. *Buildings*, *12*(6), 857.
- Neuhausen, M., Herbers, P., & König, M. (2020). Using synthetic data to improve and evaluate the tracking performance of construction workers on site. *Applied Sciences (Switzerland)*, *10*(14), 4948. <https://doi.org/10.3390/app10144948>
- Oh, X., Loh, L., Foong, S., Koh, Z. B. A., Ng, K. L., Tan, P. K., Toh, P. L. P., & Tan, U. X. (2021). Initialisation of Autonomous Aircraft Visual Inspection Systems via CNN-Based Camera Pose Estimation. *Proceedings - IEEE International Conference on Robotics and Automation, 2021-May*, 11047–11053. <https://doi.org/10.1109/ICRA48506.2021.9561575>
- Rho, J., Park, M., & Lee, H.-S. (2020). Automated construction progress management using computer vision-based CNN model and BIM. *Korean Journal of Construction Engineering and Management*, *21*(5), 11–19.

Sami Ur Rehman, M., Shafiq, M. T., & Ullah, F. (2022). Automated Computer Vision-Based Construction Progress Monitoring: A Systematic Review. *Buildings*, 12(7), 1037.

Siu, C., Wang, M., & Cheng, J. C. P. (2022). A framework for synthetic image generation and augmentation for improving automatic sewer pipe defect detection. *Automation in Construction*, 137, 104213. <https://doi.org/10.1016/j.autcon.2022.104213>

Soltani, M. M., Zhu, Z., & Hammad, A. (2016). Automated annotation for visual recognition of construction resources using synthetic images. *Automation in Construction*, 62, 14–23. <https://doi.org/10.1016/j.autcon.2015.10.002>

Sutjaritvorakul, T., Vierling, A., & Berns, K. (2020). Data-driven worker detection from load-view crane camera. *Proceedings of the 37th International Symposium on Automation and Robotics in Construction, ISARC 2020: From Demonstration to Practical Use - To New Stage of Construction Robot*, 864–871. <https://doi.org/10.22260/isarc2020/0119>

Tohidifar, A., Fatemeh Saffari, S., & Kim, D. (2022). Fake it till you make it: Training Deep Neural Networks for Worker Detection using Synthetic Data. *The 29th EG-ICE International Workshop on Intelligent Computing in Engineering*, 386–396. <https://doi.org/10.7146/AUL.455.C229>

Wei, Y., & Akinci, B. (2022). Synthetic Image Data Generation for Semantic Understanding in Everchanging Scenes Using BIM and Unreal Engine. *Computing in Civil Engineering 2021*, 934–941. <https://doi.org/10.1061/9780784483893.115>

Wong, M. Z., Kunii, K., Baylis, M., Ong, W. H., Kroupa, P., & Koller, S. (2019). Synthetic dataset generation for object-to-model deep learning in industrial applications. *PeerJ Computer Science*, 2019(10), e222. <https://doi.org/10.7717/peerj-cs.222>

Zhang, J., Fukuda, T., & Yabuki, N. (2022). Automatic generation of synthetic datasets from a city digital twin for use in the instance segmentation of building facades. *Journal of Computational Design and Engineering*, 9(5), 1737–1755. <https://doi.org/10.1093/jcde/qwac086>