

Article

Experimental Study on Using Synthetic Images as a Portion of Training Dataset for Object Recognition in Construction Site

Jaemin Kim , Ingook Wang and Jungho Yu *

Department of Architecture Engineering, Kwangwoon University, Seoul 01897, Republic of Korea; znsl10@naver.com (J.K.); ingookwang11@naver.com (I.W.)

* Correspondence: myazure@kw.ac.kr; Tel.: +82-2-940-5564; Fax: +82-2-909-5546

Abstract: The application of Artificial Intelligence (AI) across various industries necessitates the acquisition of relevant environmental data and the implementation of AI recognition learning based on this data. However, the data available in real-world environments are limited and difficult to obtain. Construction sites represent dynamic and hazardous environments with a significant workforce, making data acquisition challenging and labor-intensive. To address these issues, this experimental study explored the potential of generating synthetic data to overcome the challenges of obtaining data from hazardous construction sites. Additionally, this research investigated the feasibility of hybrid dataset in securing construction-site data by creating synthetic data for scaffolding, which has a high incidence of falls but low object recognition rates due to its linear object characteristics. We generated a dataset by superimposing scaffolding objects, from which the backgrounds were removed, onto various construction site background images. Using this dataset, we produced a hybrid dataset to assess the feasibility of synthetic data for construction sites and to evaluate improvements in object recognition performance. By finding the optimal composition ratio with real data and conducting model training, the highest accuracy was achieved at an 8:2 ratio, with a construction object recognition accuracy of 0.886. Therefore, this study aims to reduce the risk and labor associated with direct data collection at construction sites through a hybrid dataset, achieving data generation at a low cost and high efficiency. By generating synthetic data to find the optimal ratio and constructing a hybrid dataset, this research demonstrates the potential to address the problems of data scarcity and data quality on construction sites. The improvement in recognition accuracy of the construction safety management system is anticipated, suggesting that the creation of synthetic data for constructing a hybrid dataset can reduce construction safety-accident issues.

Keywords: synthetic data; hybrid dataset; object detection; construction sites



Citation: Kim, J.; Wang, I.; Yu, J. Experimental Study on Using Synthetic Images as a Portion of Training Dataset for Object Recognition in Construction Site. *Buildings* **2024**, *14*, 1454. <https://doi.org/10.3390/buildings14051454>

Academic Editor: Jurgita Antucheviciene

Received: 1 April 2024
Revised: 9 May 2024
Accepted: 14 May 2024
Published: 17 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Background and Purpose of Research

The application of AI in construction is gaining momentum, with AI being utilized not only in architecture and structural engineering but also in intelligent disaster prevention and reduction [1]. To effectively deploy AI across various settings, it is crucial to gather data specific to each environment. For instance, developing an AI solution for detecting cracks in concrete requires a vast array of crack images collected from different settings [2]. However, in reality, data are finite, and obtaining training data from real-world environments demands considerable time and effort [3]. Consequently, there is a growing emphasis on the importance of and demand for securing training data, especially in construction sites, where such efforts are increasingly recognized. Nonetheless, existing research on acquiring training data for construction sites faces several limitations. First, construction sites are inherently dynamic and hazardous, characterized by frequent changes and a large workforce. The traditional approach to collecting AI training data at construction sites, i.e., direct onsite data acquisition, poses significant risks. Given the construction industry's

reputation as one of the most hazardous industries [4], gathering data directly from these environments is fraught with safety concerns. An alternative, gathering data in a virtual environment instead of through onsite photography or videography, offers a safer and more quantifiable method. However, data obtained from existing virtual reality typically suffer from limited backgrounds due to the labor-intensive and time-consuming costs associated with background creation. Consequently, they fail to reflect the complex and dynamic background characteristics of actual construction sites, and they do not capture the diversity and continuous change inherent to construction-site environments. According to an existing study, changing backgrounds make it challenging to identify objects consistently [5]. Therefore, the efficacy of object recognition can greatly vary with the dynamics of the site's background, underscoring the need for data collection methods that incorporate the authentic and variable backgrounds of construction sites to achieve meaningful outcomes.

This study aimed to address the limitations of previous research and capture the dynamics of construction sites by separating and synthesizing objects and backgrounds to generate synthetic data of scaffolding, which is often a cause of high incidences of fall accidents on construction sites. Throughout this process, the effectiveness of synthetic data creation, the potential to resolve data scarcity issues using synthetic data, and the exploration of various composition ratios to enhance object recognition performance were investigated. Additionally, by creating backgrounds that match the object's environment, the goal was to secure data suitable for each site and reduce the need for on-site real data collection through the optimal mix of synthetic and real data within a hybrid dataset. By constructing hybrid datasets and assessing the accuracy across different composition ratios, this research sought to empirically validate how hybrid datasets can address data scarcity and quality issues and aimed to realize a cost-effective and efficient data construction method through the generation of synthetic data.

1.2. Research Scope and Method

The objective of this study was to acquire data through synthesis, effectively distinguishing between the object and background to capture the specific characteristics of the construction site. The background represents the construction environment, while the focal object is defined as the frame scaffolding. Frame scaffolding is extensively utilized in construction sites, and falls from such structures are a common occurrence. The imperative of AI-based recognition for identifying frame scaffolding lies in its potential to enhance worker safety and detect potential errors in scaffolding assembly. Additionally, acquiring training data for frame scaffolding is essential to enhance AI recognition performance. Given the linear structure of frame scaffolding, as opposed to a planar one, it allows for the visibility of all scenes positioned behind it. Consequently, the recognition of frame scaffolding, which is a linear object, is anticipated to pose greater difficulty compared to other objects.

The dataset utilized for the benchmarking training performance was captured through video recordings at the construction site. This study's methodology and procedures are illustrated in Figure 1. In the initial phase, the frame scaffolding, identified as the object of interest, was segmented from its background using the real-world images acquired. In terms of the background, a comprehensive training dataset was developed by gathering photographs and videos that accurately depict the construction site's atmosphere. These elements were then synthesized with the construction site backgrounds, integrating the isolated objects. This synthesized training dataset was further augmented using Roboflow, an advanced platform designed for image dataset management and enhancement of model development processes. Roboflow offers key features, such as image annotation, labeling, and augmentation. Following the augmentation, the enriched synthetic dataset underwent labeling and was processed using the YOLOv5 algorithm for training. The research included an analysis of the outcomes by modifying the proportion of synthetic to real data within the dataset, evaluating the dataset's efficacy at various ratios. The structure of the remainder of

this document is as follows. Section 2 engages in a literature review pertinent to training datasets. Section 3 details the methodology employed for generating a synthetic dataset specific to construction sites, encompassing strategies for data collection, synthesis, labeling, and augmentation. Section 4 outlines the results from the performance evaluation of the hybrid dataset. Section 5 concludes the document with a summary of our findings and their implications.

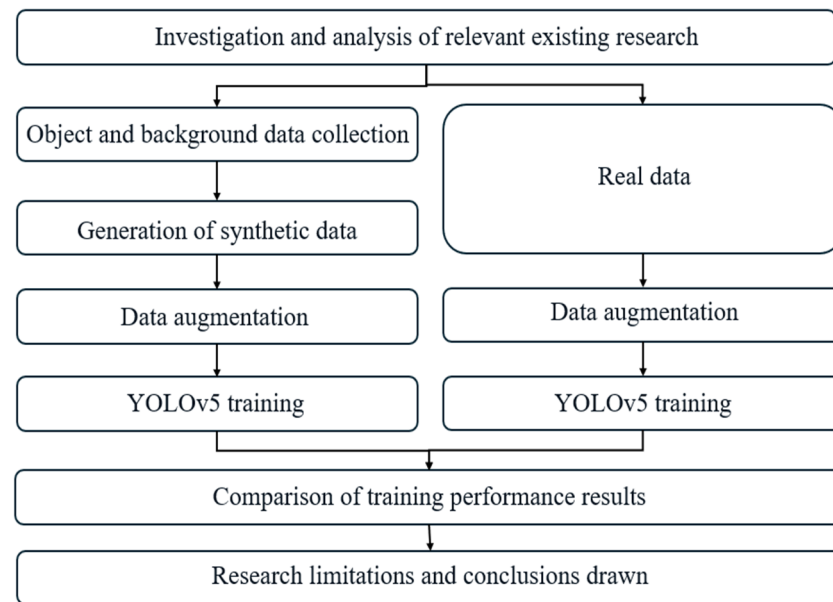


Figure 1. Research progress and methodology.

2. Literature Review

Research is being conducted to improve object recognition performance in construction sites by developing neural network-based architectures and by generating synthetic data. In [6,7], a network that predicts parameters like the center point, width, and angle of construction vehicles to detect them accurately and efficiently was proposed. Additionally, ref. [8] introduced a detection method based on the Enet neural network for identifying fenced at construction sites. The algorithm requires reconstruction to suit new objects, and network design also demands a significant variety of data. Thus, this study focused on generating synthetic data to improve recognition performance and conducted a literature review to explore this method. The construction of an AI training dataset is categorized as illustrated in Figure 2. It primarily involves two approaches: collecting data from real environments and generating virtual training data through synthesis. In the case of data collected from real environments, the process can be challenging due to the dynamic nature of the surroundings or the presence of hazardous conditions. Consequently, significant research is focused on the synthesis of data within virtual environments as a strategy to amass adequate training data, even under adverse conditions.

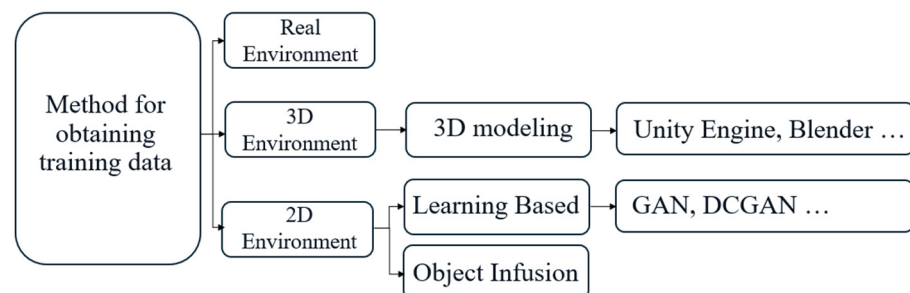


Figure 2. Method for obtaining training data.

Table 1 summarizes the methods for generating synthetic data as identified in previous studies. The key term ‘Synthetic Image Data Generation’ was used to conduct searches on Google Scholar and SCIENCE ON. The studies listed in the table span from 2017 to 2023, indicating that they were conducted within the past seven years. These papers are categorized based on the synthetic data generation method employed. There are three primary synthesis approaches: in a 3D environment, in a 2D environment, and a hybrid of both. Research related to 3D environment synthesis utilizes three-dimensional (3D) programs and game engines. For instance, data from tools like the Unity Engine, commonly used as a game development platform, are transformed from 3D graphical images or videos into 2D image data for use as training material. One notable study that focused on image synthesis for underwater object recognition using Convolutional Neural Networks (CNNs) involved rendering the objects after 3D CAD modeling [9]. However, employing game engines or 3D programs for synthesis is time-consuming, technically demanding, and thus costly. To address these constraints, ‘object infusion’ and ‘learning-based’ synthesis methods have been developed. For example, research on generating synthetic data involves modeling objects in a 3D environment and integrating them into 2D background images [10]. Nevertheless, this approach is limited to generating synthetic data only for the objects that have been modeled, and it lacks realism, failing to accurately represent physical objects. The ‘learning-based’ method, particularly using Generative Adversarial Networks (GANs), offers a solution by allowing for the generation of synthetic images from existing data. This approach not only facilitates the expansion of insufficient training datasets through GANs but also significantly increases the sample size of the dataset, showcasing its potential for enhancing data diversity and quality.

Table 1. Comparative table of research related to synthetic data generation.

Method	Ref.	Contents		
		Method Description	Model Used	Performance Metric
3D Modeling	[9]	Utilized 3D CAD models in Blender for underwater object detection.	MASK R-CNN	mAP: 0.9
	[11]	Constructed 3D environment with Blender for food identification inside refrigerators.	CCN	Synthetic4000 mAP: 24
3D + 2D	[10]	Generated objects using 3D modeling programs, integrating them into 2D backgrounds for construction site analysis to detect worker.	DNN	Real5000 +syn5000 67.8%
	[12]	Combined 3D BIM programs with GAN for interior building analysis.	RCNN	71.6%, 84.9%
	[13]	Integrated CAD modeling and 3D program rendering with GAN for object modeling.	YOLOv4 PVNet	Synthetic300 mAP: 0.93
GAN	[14]	Applied the SinGAN-Seg model for tumor detection in visceral imaging.	UNet++	IOU 0.79
	[15]	Utilized AE-GAN for jellyfish detection in the sea.	FCN	71.4%
2D	[16]	Background removal and object insertion technique for generic object analysis.	YOLOv5	mAP 0.593
	[17]	Enhanced object insertion with background removal and shadow effect for person detection.	YOLOv4, YOLOv5	synthetic mAP 0.873, 0.864
	[18]	Applied background removal, object insertion, and blur processing for traffic bucket detection on water.	CNN	-
	Our research	Developed a novel method combining object insertion, generation hybrid dataset and training for scaffolding detection at construction sites.	YOLOv5	Hybrid dataset mAP 0.892

The initial exploration of data training utilizing Generative Adversarial Networks (GANs) led to the creation of a synthetic image dataset for medical technology applications [14]. Similarly, a previous study generated synthetic data for jellyfish in marine environments using deep learning techniques [15]. Moreover, numerous studies, such as the generation of dental image data employing Deep Convolutional GANs (DCGANs) [19], are underway. These efforts are significant, as they address the shortage of training data for deep learning by enlarging the dataset and enhancing efficiency. The strategy of deploying algorithms like GAN and DCGAN to enrich images and amplify the sample size through modifications in object characteristics primarily aims to increase data volume, rather than its diversity. Achieving diversity through extensive alterations, such as variations in the background environment and object positioning, proves challenging when augmenting with a constrained dataset. In contexts like construction sites, not only is a large sample size vital but also data from diverse environments. As such, synthesizing varied backgrounds and objects through the ‘insert object’ method becomes crucial for attaining significant diversity.

The ‘insert object’ technique allows for the incorporation of objects into pre-existing background images without the necessity for configuring specific settings via a 3D program. This approach facilitates the acquisition of a substantial volume of training data and the enhancement of data diversity, circumventing the need for background 3D modeling, even in scenarios of data scarcity. For instance, in a study applying the object insertion technique, data were synthesized by isolating traffic cones from images of roads submerged under water [18]. By leveraging the object insertion method, it is feasible to generate novel data by synthesizing backgrounds and objects, thereby achieving diversity through the alteration of object types and locations.

Table 2 contrasts the processes involved in constructing synthetic datasets. The asterisk (*) in the table denotes tasks that required time and labor investment. The approach of generating a training dataset via deep learning techniques, such as GAN, primarily aims to enlarge the sample size, often at the expense of diversity. Furthermore, this method is characterized by significant investments in time and technical expertise. Hence, for a construction site, with its constantly evolving environment, a method that prioritizes diversity, like object insertion, becomes crucial. The generation of synthetic data necessitates data that align with the specific environmental requirements, suggesting that AI applications for construction sites demand training data that accurately represent the site’s visual characteristics. Previous research has been limited by a lack of consideration for the diversity of construction environments and backgrounds in the ‘object insertion’ method, and there is an absence of studies on creating synthetic data for linear objects like scaffoldings, which are prone to high rates of fall accidents. In the constantly changing environment of construction sites, securing various training data through the ‘object insertion’ method appears increasingly necessary. Therefore, this study explored the creation of synthetic data using the object insertion method for scaffolding, which has a high rate of fall incidents but low object recognition rates due to its linear object characteristics, to confirm the potential of hybrid datasets in obtaining data for construction sites.

Table 2. Synthetic dataset construction process (* Tasks that required time and labor investment).

Method	Process Summary	Labor-Intensive Aspects
Generative Adversarial Network (GAN)	- The method followed by the creation of the GAN architecture and training the model to generate synthetic images.	- Generation of GAN architecture * - Training of GAN model *
3D Modeling platform	- The process involves constructing a detailed 3D environment, specifying object properties, and then generating images.	- Construction of 3D environment * - Specification of object properties *

Table 2. Cont.

Method	Process Summary	Labor-Intensive Aspects
3D Modeling platform + GAN	<ul style="list-style-type: none"> - This approach includes GAN structure creation and training, and it concludes with the rendering of images and the production of multi-angle views. 	<ul style="list-style-type: none"> - Modeling * - Generation of GAN architecture * - Training of GAN Model * - Rendering *
2D Object Insertion	<ul style="list-style-type: none"> - This method involves collecting real-world data, removing backgrounds, synthesizing new images, and augmenting the dataset for greater diversity. 	<ul style="list-style-type: none"> - Background removal

3. Construction Site Hybrid Dataset Creation

Figure 3 illustrates the concept of generation synthetic data and construction hybrid dataset. For hybrid dataset construction, both real and synthetic datasets are needed. In Figure 3, the blue indicates the dataset used for generating the synthetic dataset, and the gray indicates the real dataset. The first stage involves data collection and synthesis, focusing on the synthesis of scaffold objects and construction background images. The second stage encompassed data labeling and augmentation, while the third stage involves constructing hybrid dataset.

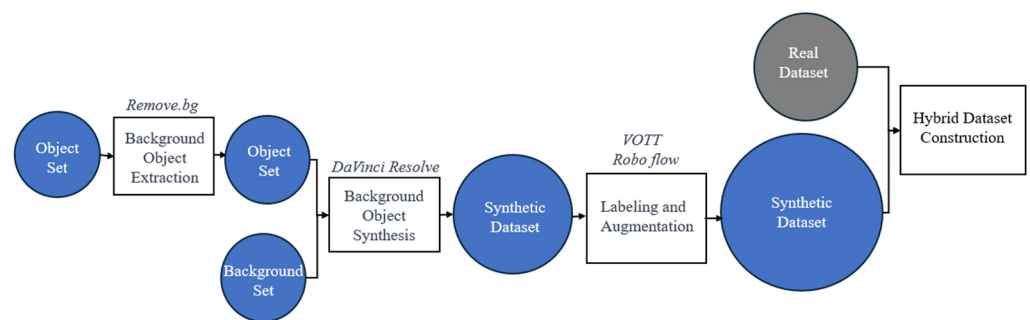


Figure 3. Concept of generation synthetic data and construction hybrid dataset.

3.1. Data Collection and Synthesis

The synthesis of data is detailed in Figure 4. Initially, necessary data comprising construction site backgrounds and construction objects for synthetic data creation were compiled. Background data, representing construction sites, were sourced from photographs taken at actual sites. With regard to frame scaffolding, to ensure diversity reflective of actual construction site usage, datasets with redundant shapes were eliminated, favoring a collection of varied configurations. To enhance the diversity of object types, we utilized different types of scaffolding as objects. Figure 4 illustrates the method of generating synthetic data, where simple images depict scaffold images used for object synthesis. The background data consist of construction site images where scaffolding can be placed on the floor. Simplified images were produced by separating the backgrounds from the collected data using ‘remove.bg’, an AI-powered online service tool designed for automatic background removal from images. ‘Remove.bg’ is a free online tool that facilitates the quick and easy removal of backgrounds from objects. However, when removing backgrounds from scaffolding, which consists of thin lines, there is a high likelihood of improper separation if the color of the scaffolding closely resembles that of the background. Therefore, this study primarily utilized scaffolding data with white backgrounds. If an algorithm capable of effectively separating objects from their backgrounds is developed in the future, it could significantly enhance both realism and convenience. The total number of construction site background images collected was 302. Synthetic data were then generated by superimposing objects, from which backgrounds were removed, onto these images. This object insertion process

utilized the ‘DaVinci Resolve’ software, version 18, renowned for its advanced image- and video-editing features.

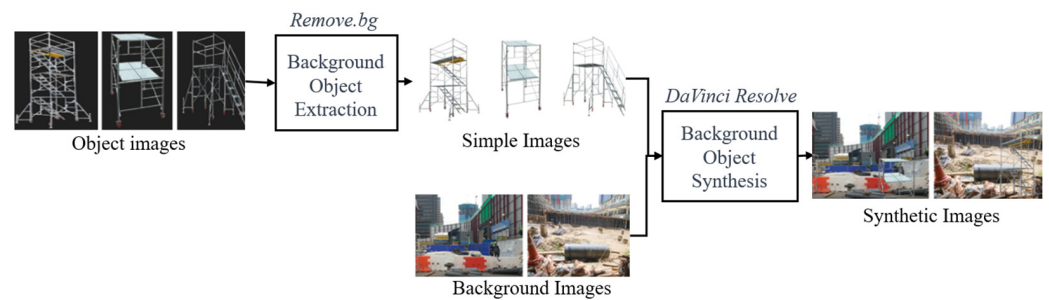


Figure 4. Synthetic data-generation method.

Deep-learning models process data at the pixel level. Consequently, when an image is enlarged or reduced, it is perceived as a distinct image by the model. Enhancing the training dataset through image-processing techniques is instrumental in improving deep learning-model performance. This study sought to augment the volume of data through data synthesis and augmentation techniques. In the data creation process, synthetic images were generated by resizing and repositioning, resulting in the production of 280 synthetic scaffolding frame images. This approach facilitated an increase in diversity using the existing limited dataset, without the need for additional data collection. The synthesized data, presented in Figure 5, were created in video format, which contains more information than static images, indicating a high capability for generating substantial training data from a single source.



Figure 5. Frame scaffolding synthesis data.

3.2. Data Labeling and Augmentation

To evaluate performance, image data synthesized from construction sites and objects were annotated and labeled frame by frame using VOTT (<https://github.com/microsoft/VoTT>) (Visual Object Tagging Tool developed by the Commercial Software Engineering group at Microsoft in Israel), a freely available open-source software framework designed for image annotation and labeling. Figure 6 illustrates the labeled image data and the yellow box indicates synthesized frame scaffolding. After labeling 280 images using synthetic data, the dataset underwent an augmentation process in Roboflow. Securing data in various environmental conditions such as brightly lit or dark areas is essential. However, the labor

required to acquire such data in the real world is significant. To overcome the challenges of obtaining diverse training data in realistic environments, adjustments in saturation, brightness, and exposure were made to enhance environmental diversity. Additionally, to further expand the dataset, augmentation techniques such as flip, crop, noise, and cutout were employed, with the augmented data displayed in Figure 7. The yellow box in Figure 7 indicates synthesized and augmented frame scaffolding. VOTT also enables the conversion of video data into image data, which are subsequently transformed into textual data via labeling. As a result, this process yielded 2000 frames of image data for scaffolding synthesis.



Figure 6. Labeled synthetic data images.

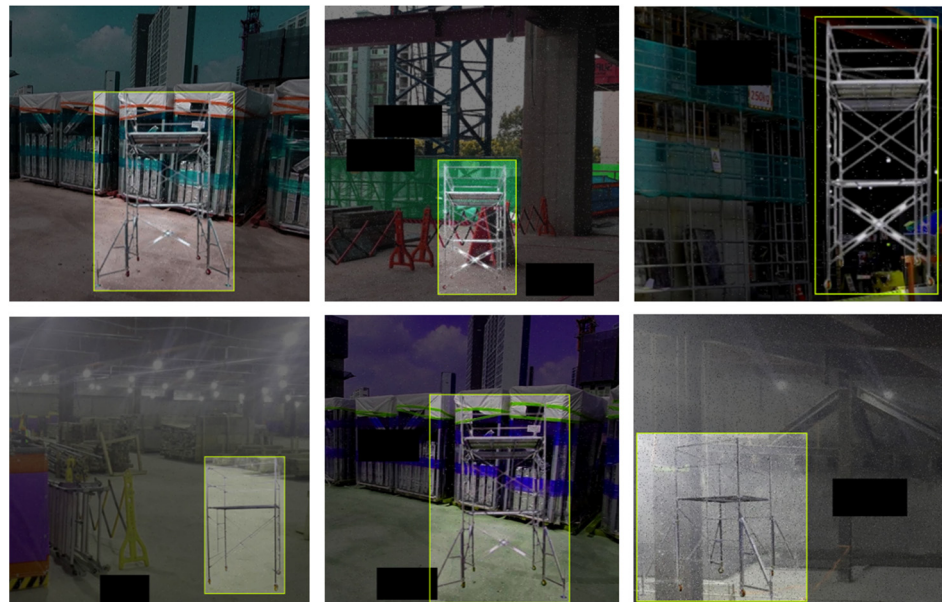


Figure 7. Labeled synthetic data using augmentation techniques.

3.3. Construction Hybrid Dataset

To experiment with the highest object-recognition performance across different ratios, the study constructed a hybrid dataset by adjusting the proportion of real and synthetic data. The dataset's classification by type is presented in Table 3. The 'real dataset' comprises

1100 images for training, 830 for validation, and 317 for testing, totaling 2247 images. The test dataset consists of 317 images, utilizing the same real data to evaluate the performance across different composition ratios. The synthetic dataset includes 1100 training images and 950 validation images. The Hybrid Dataset 2:8 splits into 360 real images (20%) and 1440 synthetic images (80%) from a total of 1800 images. The ‘Hybrid Dataset 5:5’ is composed of 850 real images (50%) and 850 synthetic images (50%), totaling 2017 images, including the testing set. Of these, hybrid datasets are allocated into 1300 for training and 400 for validation. All four datasets were tested using the same set of 317 images. This test set contained images from the real dataset not used in training or validation.

Table 3. Classification table by dataset type.

Dataset Type	Dataset			Total Data
	Training Set	Validation Set	Testing Set	
Real dataset	1100	830	317	2247
Synthetic dataset	1100	950	317	2367
Hybrid dataset (2 real/8 synthetic)	1200	600	317	2117
Hybrid Dataset 5:5	1300	400	317	2017

4. Validation

4.1. Validation Scenario

For data validation, the labeled synthetic data were used to train the YOLOv5 model, facilitated by a specially developed algorithm. YOLOv5 developed by Ultralytics, was selected for model training and performance evaluation due to its excellent balance of speed and accuracy in object detection task. The YOLO model has shown significant improvements in inference speed and model size. In the context of object recognition at construction sites, real-time detection is crucial for future applications. Therefore, YOLOv5, with its superior balance of speed and accuracy, was utilized for this purpose. This algorithm manages the training process and evaluates the model’s performance metrics, such as accuracy and loss. The training efficacy of both synthetic and real data was assessed using this algorithm. The YOLOv5s model was deployed in Python 3.10, on the Google Colab server. The datasets were classified prior to model training. During the training phase, the model underwent 100 epochs with a batch size of 16, and images were resized to 416×416 to optimize GPU memory usage.

4.2. Results of Validation

In this study, precision and mean average precision were selected as the evaluation metrics. Precision measures the proportion of positive predictions that are actually correct. In fields such as scaffolding, where safety is paramount and numerous fall accidents occur, minimizing the number of incorrect object detections by the model is crucial. Therefore, by assessing precision, the aim is to reduce false alarms and ensure reliability in real danger situations. The performance of data-driven algorithms is often assessed using precision, recall, and mean average precision (mAP). Precision, as defined by Formula (1), represents the algorithm’s accuracy in identifying true-positive results among all positive predictions made. mAP, the mean of the average precision, serves as an overall indicator of the algorithm’s object detection performance, with AP representing the area under the precision–recall (P-R) curve. The closer an algorithm’s AP is to 1, the better its detection performance. Therefore, precision and recall values are instrumental in evaluating an algorithm’s training efficacy.

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

In this experiment, we tested the object-detection capabilities of the synthetic data. Table 4 presents the test results based on the dataset classification. The test revealed that the dataset comprised entirely of real data achieved the highest training performance, with a mAP of 0.989. In contrast, the dataset consisting entirely of synthetic data showed a lower performance, with a mAP of 0.518. Nonetheless, the YOLOv5 model's performance can be improved by integrating real and synthetic data. Among the hybrid datasets, a ratio of 8:2 (synthetic to real) yielded better training results than a 5:5 ratio, with metrics nearing those of the real data. The performance metrics for a hybrid dataset (2 real/8 synthetic) are presented in Table 5. The results indicate an accuracy of 0.712, an F1-score of 0.830, a recall of 0.787, and a precision of 0.879. These metrics demonstrate the efficacy of integrating synthetic data with real data to train detection models. Figure 8 illustrates the test graph for the hybrid dataset (2 real/8 synthetic), including (a) the precision curve and (b) the mAP graph. In the graph, the light line shows the actual metric values at each step, while the bold line, which applies a smoothing factor of 0.6 to reduce short-term noise and exceptional fluctuations in the data. Both (a) precision and (b) mAP values saw significant improvement after 20 epochs, stabilizing at 1 after 70 epochs.

Table 4. Test results by dataset classification.

Dataset Type	P (Precision)	mAP50
① Real dataset	0.983	0.989
② Synthetic dataset	0.892	0.518
③ Hybrid dataset (2 real/8 synthetic)	0.879	0.892
④ Hybrid dataset (5 real/5 synthetic)	0.862	0.779
⑤ Hybrid dataset (8 real/2 synthetic)	0.852	0.886

Table 5. Performance metrics of hybrid dataset (2 real/8 synthetic).

Dataset	Accuracy	F1-Score	Recall	Precision
Hybrid Dataset (2 real/8 synthetic)	0.712	0.830	0.787	0.879

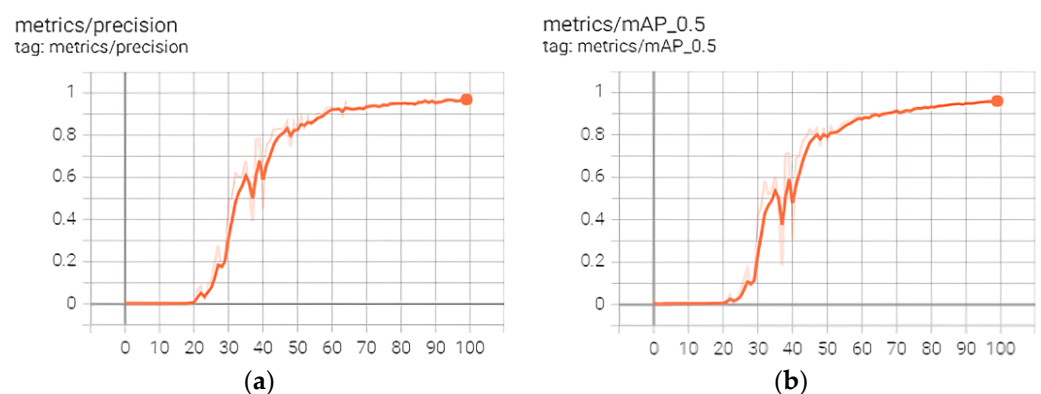


Figure 8. ③ Hybrid dataset (2 real/8 synthetic) test graph: (a) precision curve and (b) mAP curve.

Figure 9 explores how varying the mix ratio of real to synthetic data in the frame scaffolding dataset affects the training performance. The graph indicates that accuracy does not linearly increase with a higher proportion of real data, highlighting that the performance boost is not unconditional with an increase in real data. Thus, when creating a hybrid dataset through synthetic data generation, finding the optimal balance between

real and synthetic data is essential. The graph demonstrates that, while the dataset with only synthetic data had a low mAP of 0.518, adjusting the ratio with real data improved the performance. The dataset combining real and synthetic data at a 2:8 ratio achieved a mAP of 0.892, marking an approximate 0.38 increase from the synthetic-only dataset. With a composition ratio of 2:8 between real and synthetic data, the dataset significantly improved object recognition accuracy on construction sites. Despite its slightly lower accuracy compared to the dataset with only real data, the 2:8 ratio dataset showed enhanced feasibility for data collection, utilizing the efficiency of image synthesis to rapidly gather extensive data. Synthetic data can be combined with real data to construct a training dataset, thereby reducing the labor involved in collecting actual data. However, differences exist between real data and synthetic data, with the probability being higher that data recognized in practice will resemble real data more closely than synthetic data. It is not merely the quantity of real data that increase accuracy. Rather, adjusting the synthetic ratio to enhance performance with an optimally balanced dataset is crucial. Thus, the ratio of synthetic to real data is significant.

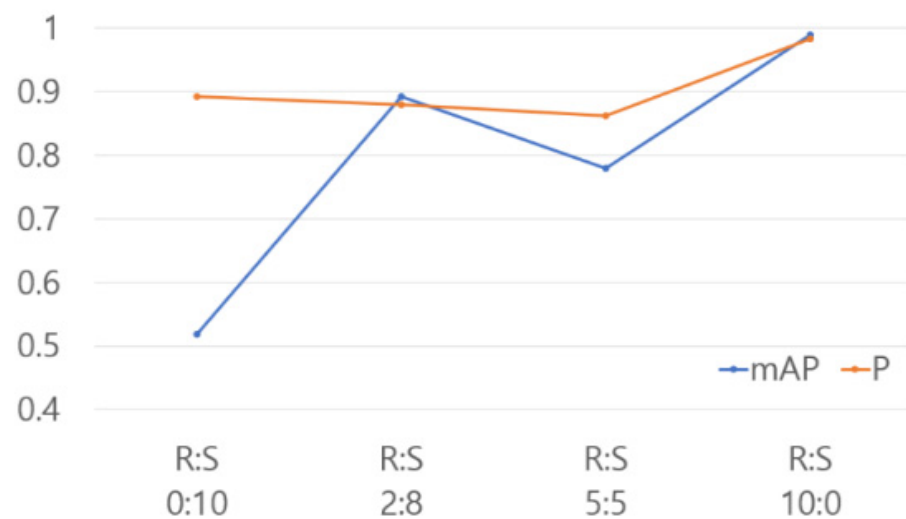


Figure 9. Performance based on the ratio variation between real data and synthetic data.

Figure 10 presents the results of frame scaffolding object recognition; the red box indicates detected scaffolding at a construction site. It demonstrates successful recognition of both the frame scaffolding in the augmented image and the non-augmented original image. The process of building a frame scaffolding synthetic dataset comprising 2000 images through synthesis and augmentation took approximately 40 min. However, it is worth noting that the proficiency of the tool may influence the presence of errors. Real data collected from construction sites often face challenges in securing diverse training data due to limitations in time and space. In contrast, synthetic data offer the advantage of generating various datasets by manipulating factors such as object position and size, even with a limited number of construction objects and backgrounds. Synthetic data can achieve a satisfactory performance with fewer real data compared to traditional methods, potentially easing the challenges of data collection at construction sites. Increasing the composition ratio of synthetic data while decreasing the proportion of real data seems to mitigate the challenges associated with actual data collection in construction sites, thereby offering a promising approach to address the difficulties in obtaining real data in such environments.

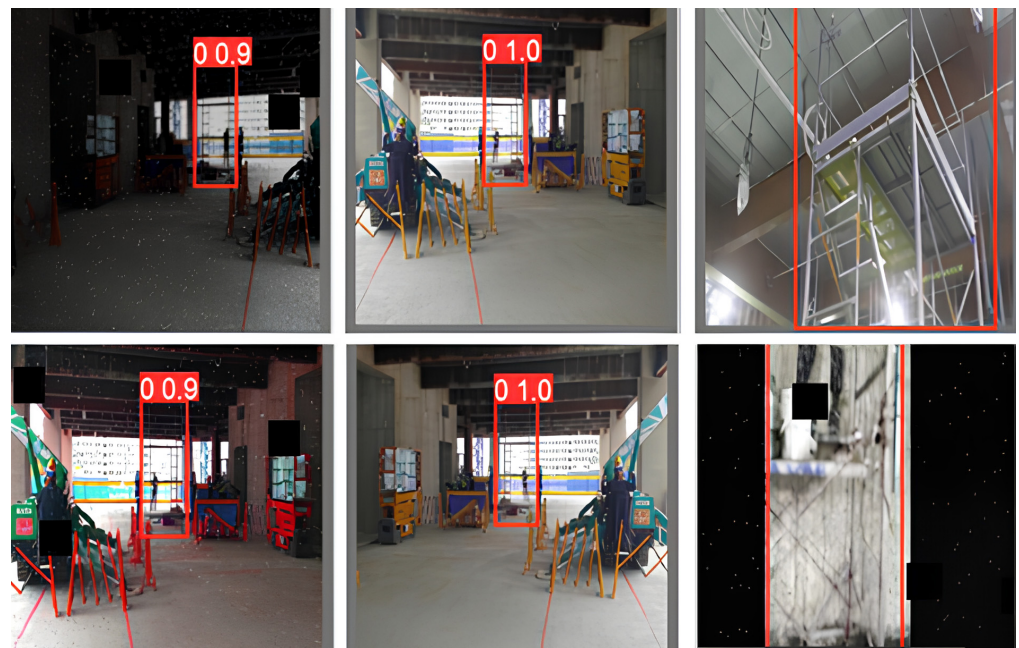


Figure 10. Frame scaffolding recognition results.

5. Conclusions

This study involved an experimental investigation into the potential of synthetic data generation for scaffolding, one of the contributing factors to fall accidents on construction sites, by using linear objects as the subject. The aim was to find the optimal ratio of mixing synthetic and real data to achieve a high object recognition performance and to verify the effectiveness of the hybrid dataset through experimentation. Our aim was to overcome the challenges of collecting training data at construction sites by creating a training dataset with an optimal composition ratio using synthetic data. We generated meaningful data at a low cost by synthesizing 2D backgrounds and objects. Training with a hybrid dataset, comprising real and synthetic data at a 2:8 ratio, yielded a performance similar to that of real data, validating the efficacy of the synthesized data. By reducing the proportion of data requiring direct collection at the construction site by 80% through the construction of a hybrid dataset, we mitigated the risk associated with data collection. The limitations of this study include the difficulty in making accurate comparisons among research papers due to the varied topics of each study, such as synthesized backgrounds, objects, and data quantities. Additionally, a limitation of this study is the small size of the dataset and the fact that the proportions of synthetic to real data in the hybrid dataset do not cover all cases. Future research aims to increase the size of the dataset and diversify the mix ratio of real and synthetic data to more precisely determine the impact of these ratios on object recognition performance. Despite these challenges, this study confirms the impact of hybrid datasets on performance improvement by creating synthetic data, including linear objects like scaffolds commonly found at construction sites. As there are limitations to collecting all necessary data directly from construction sites, the significance lies in generating low-cost, high-efficiency data by reducing the need for actual data collection through synthetic data creation. Through this research, it is evident that synthetic data generation, combined with real data at appropriate ratios, can reduce the labor involved in direct data collection on-site, particularly for data that are challenging to obtain due to linear characteristics. Future studies are needed to improve the accuracy of object recognition using synthetic datasets. For the future research, we plan to enhance the detection performance and increase data diversity by incorporating various types of object data collected at construction sites, as well as background data differentiated by environmental factors. We anticipate that the inclusion of more diverse background data, tailored to specific environmental requirements and object types, will make the data more suitable for construction sites. If synthetic

data closely resemble real data, they are expected to enhance accuracy, necessitating the application of additional imaging technology to make synthetic data closer to reality. Image-processing techniques can adjust shadows, contrast, and color changes to make synthetic data more similar to real data. Furthermore, since real objects are 3D, recognizing images can vary depending on the object's position and angle. To address this, synthetic object data collected from various angles rather than a single angle are expected to yield higher object recognition rates. Moreover, combining AI with ontology to assess the safety or unsafety of recognized objects can significantly enhance AI's applicability [20]. Simultaneous research on situation judgment, incorporating synthetic data generation and ontology, could lead to a more accurate situation judgment. Beyond object recognition, if we can judge the safety or danger of a situation that occurs during work on an object, we can expect to reduce the occurrence rate of accidents at construction sites.

Author Contributions: Conceptualization, J.K., I.W. and J.Y.; methodology, J.Y.; software, J.K. and I.W.; validation, J.K. and I.W.; formal analysis, J.K. and I.W.; investigation, J.K. and I.W.; resources, J.K., I.W. and J.-H.; data curation, J.K. and I.W.; writing—original draft preparation, J.K. and I.W.; writing—review and editing, J.K.; visualization, J.K. and I.W.; supervision, J.Y.; project administration, J.Y.; funding acquisition, J.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by authors on request.

Acknowledgments: The present research was conducted by the excellent researcher support project of Kwangwoon University in 2024.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Xu, Y.; Qian, W.; Li, N.; Li, H. Typical advances of artificial intelligence in civil engineering. *Adv. Struct. Eng.* **2022**, *25*, 3405–3424. [\[CrossRef\]](#)
2. Jung, S.-Y.; Sohn, C.-B.; Yu, J.-H. A study on the meaningful characteristic variables of the image for estimating the depth of concrete cracks. *Korea Facil. Manag. Assoc.* **2021**, *16*, 43–51.
3. Yang, H. Synthetic Image Dataset Generation for Defense using Generative Adversarial Networks. *J. Korea Inst. Mil. Sci. Technol.* **2019**, *22*, 49–59. [\[CrossRef\]](#)
4. Yeo, C.J.; Yu, J.H.; Kang, Y. Quantifying the Effectiveness of IoT Technologies for Accident Prevention. *J. Manag. Eng.* **2020**, *36*, 04020054. [\[CrossRef\]](#)
5. Xiao, K.; Engstrom, L.; Ilyas, A.; Madry, A. Noise or Signal: The Role of image Backgrounds in Object Recognition. *arXiv* **2020**, arXiv:2006.09994.
6. Guo, Y.; Xu, Y.; Niu, J.; Li, S. Anchor-free arbitrary-oriented construction vehicle detection with orientation-aware Gaussian heatmap. *Comput.-Aided Civ. Infrastruct. Eng.* **2023**, *38*, 907–919. [\[CrossRef\]](#)
7. Guo, Y.; Xu, Y.; Li, S. Dense construction vehicle detection based on orientation-aware feature fusion convolutional neural network. *Autom. Constr.* **2020**, *112*, 103124. [\[CrossRef\]](#)
8. Zhang, X.; Tang, T.; Wu, Y.; Quan, T. Construction Site Fence Recognition Method Based on Multi-Scale Attention Fusion ENet Segmentation Network. In Proceedings of the 35th International Conference on Software Engineering and Knowledge Engineering, Online, 1–10 July 2023.
9. Jeon, M.; Lee, Y.; Shin, Y.S.; Jang, H.; Yeu, T.; Kim, A. Synthesizing image and automated annotation tool for CNN based under water object detection. *J. Korea Robot. Soc.* **2019**, *14*, 139–149. [\[CrossRef\]](#)
10. Kim, J.; Kim, D.; Lee, S.; Chi, S. Hybrid DNN training using both synthetic and real construction images to overcome training data shortage. *Autom. Constr.* **2023**, *149*, 104771. [\[CrossRef\]](#)
11. Rajpura, P.S.; Bojinov, H.; Hegde, R.S. Object detection using deep CNNs trained on synthetic image. *arXiv* **2017**, arXiv:1706.06782.
12. Hong, Y.; Park, S.; Kim, H.; Kim, H. Synthetic data generation using building information models. *Autom. Constr.* **2021**, *130*, 103871. [\[CrossRef\]](#)
13. Yang, X.; Fan, X.; Wang, J.; Lee, K. Image translation based synthetic data generation for industrial object detection and pose estimation. *IEEE Robot. Autom. Lett.* **2022**, *7*, 7201–7208. [\[CrossRef\]](#)
14. Thambawita, V.; Salehi, P.; Sheshkal, S.A.; Hicks, S.A.; Hammer, H.L.; Parasa, S.; Lange, T.D.; Halvorsen, P.; Riegler, M.A. SinGAN-Seg: Synthetic training data generation for medical image segmentation. *PLoS ONE* **2022**, *17*, e0267976. [\[CrossRef\]](#) [\[PubMed\]](#)

15. Kim, K.; Myung, H. Autoencoder-combined generative adversarial networks for synthetic image data generation and detection of jellyfish swarm. Digital Object Identifier. *IEEE Access* **2018**, *6*, 54207–54214. [[CrossRef](#)]
16. Zhang, X.; Zhao, C.; Luo, H.; Zhao, W.; Zhong, S.; Tang, L.; Peng, J.; Fan, J. Automatic learning for object detection. *Neurocomputing* **2022**, *484*, 260–272. [[CrossRef](#)]
17. Sung, J.; Kim, H.; Kim, M.; Mok, Y.; Park, C.; Paik, J. Synthetic image generation for data augmentation to train an unconscious person detection network in a UAV environment. *IEIE Trans. Smart Process. Comput.* **2022**, *11*, 156–161. [[CrossRef](#)]
18. Jiang, J.; Qin, C.Z.; Yu, J.; Cheng, C.; Liu, J.; Huang, J. Obtaining urban waterlogging depths from video images using synthetic image data. *Remote Sens.* **2020**, *12*, 1014. [[CrossRef](#)]
19. Shin, H.; Kim, D.; Chang, S.; Kim, K.-w.; Jeong, J.; Son, Y. A Study on Generation Model of DCGANs-based Teeth Image Data for Effective Identification. *J. Korean Inst. Intell. Syst.* **2020**, *30*, 265–271. [[CrossRef](#)]
20. Lee, S.-K.; Yu, J.-H. Ontological inference process using AI-based object recognition for hazard awareness in construction sites. *Autom. Constr.* **2023**, *153*, 104961. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.