HangCon: Benchmark Data Set for Enhanced Detection of Hanging Objects in Construction Sites

Gilsu Jeong, S.M.ASCE[1]; Seongeun Park[2]; Joonseok Lee[3]; Moonseo Park, M.ASCE[4]; and Changbum R. Ahn, A.M.ASCE[5]

Abstract: Lifting operations on construction sites pose significant safety risks due to the potential hazard of falling objects. Effective monitoring of hanging objects is crucial for preventing accidents and ensuring worker safety. However, detecting hanging objects presents unique challenges for existing models, including the invariance in object shapes regardless of their hanging status, complex backgrounds that obscure ropes, and the diversity of hanging objects in terms of size, shape, and texture. To address these challenges, this study introduces HangCon (Hanging Objects in Construction Sites), a novel data set specifically designed for detecting "hanging objects"—loads suspended by tower cranes. HangCon contains 101,381 images, split between 50,842 images of hanging objects and 50,539 images of nonhanging objects, providing detailed annotations and diverse scenes. To evaluate HangCon's effectiveness, this study conducted experiments using 10 benchmark models. The results highlighted the challenges in detecting hanging objects, with the best mAP at 71.63% for hanging objects alone, improving to 76.01% with unified annotations of objects and ropes. These findings highlight the complexity of detecting hanging objects and emphasize the necessity to implement advanced techniques such as semantic segmentation, depth estimation, and improved rope line detection. HangCon serves as a crucial resource for developing and refining detection models tailored to construction environments, significantly contributing to improved safety and operational efficiency on construction sites. By offering a comprehensive and well-annotated collection of images, HangCon facilitates the training and benchmarking of object detection models specifically for construction environments. **DOI: 10.1061/JCCEE5.CPENG-6283.** *© 2025 American Society of Civil Engineers.*

Author keywords: Hanging object; Construction safety; Image data set; Object detection; Deep learning.

Introduction

As construction projects grow in complexity and scale, effective monitoring has become increasingly vital for tasks such as object detection, activity recognition, and cycle time measurement (Ahn et al. 2012; Fang et al. 2018; Sherafat et al. 2020; Xiao and Kang 2021b). Recent research has focused on automated monitoring methods to enhance accuracy and efficiency while reducing time spent on manual monitoring. These methods provide valuable data for optimizing construction processes, thereby improving overall productivity (Azar 2016; Kim et al. 2018a; Roberts and Golparvar-Fard 2019). Advances in vision-based monitoring, particularly through object detection techniques, have transformed the industry by enabling real-time recognition of hazardous activities via video surveillance (Golparvar-Fard et al. 2013; Jeong et al. 2024; Kim et al. 2016). These technologies are essential for enhancing site safety and operational efficiency, making automated monitoring systems crucial for managing complex construction projects.

Monitoring suspended loads in high-risk workplaces is crucial for ensuring both safety and productivity enhancements (Li and Liu 2012; Yang et al. 2014). At construction sites, lifting and transporting heavy materials is common. These activities are frequently performed above workers' height, making the suspended loads invisible to them and significantly increasing the risk of accidents. The potential hazards associated with unseen suspended loads necessitate vigilant and effective monitoring to prevent accidents and ensure worker safety (Ali et al. 2024; Wang et al. 2023; Yong et al. 2023). Fig. 1 illustrates safety issues that workers face during crane operations. Efficient monitoring of lifting activities is also crucial for improving work efficiency, as well-coordinated operations accelerate material transport, positively affecting project timelines (Huang et al. 2021; Roberts and Golparvar-Fard 2019; Zhang and Hammad 2012). Therefore, implementing robust monitoring systems with object detection techniques for hanging objects is critical for maintaining safety standards and optimizing construction project efficiency.

In this study, objects suspended and transported by a tower crane are defined as "hanging objects," and the task of detecting these hanging objects is defined as "hanging object detection." One of the primary challenges in detecting hanging objects on construction sites is their slow or stationary movement, which complicates the use of movement-based detection methods. Additionally, the dynamic environment of construction sites, with moving elements such as wind-blown fabrics, operating machinery, and workers, makes it difficult to accurately distinguish hanging objects from other moving items. Traditional image-based approaches face challenges in accurately identifying hanging objects due to the

[1]Ph.D. Student, Dept. of Architecture and Architectural Engineering, Seoul National Univ., Seoul 08826, Republic of Korea. ORCID: https://orcid.org/0009-0005-2652-2987. Email: jeonggskr@snu.ac.kr

[2]Ph.D. Student, Dept. of Architecture and Architectural Engineering, Seoul National Univ., Seoul 08826, Republic of Korea. ORCID: https://orcid.org/0000-0003-0617-5154. Email: seongeunpark@snu.ac.kr

[3]Assistant Professor, Graduate School of Data Science, Seoul National Univ., Seoul 08826, Republic of Korea. Email: joonseok@snu.ac.kr

[4]Professor, Dept. of Architecture and Architectural Engineering, Seoul National Univ., Seoul 08826, Republic of Korea. Email: mspark@snu.ac.kr

[5]Professor, Dept. of Architecture and Architectural Engineering, Seoul National Univ., Seoul 08826, Republic of Korea (corresponding author). ORCID: https://orcid.org/0000-0002-6733-2216. Email: cbahn@snu.ac.kr

Note. This manuscript was submitted on June 27, 2024; approved on October 1, 2024; published online on January 9, 2025. Discussion period open until June 9, 2025; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Computing in Civil Engineering***, © ASCE, ISSN 0887-3801.

Fig. 1. Example images illustrating safety issues of workers during object lifting by tower crane. (Images by authors.)

diverse range of objects and their consistent physical appearance regardless of whether or not they are hanging. This requires recognizing subtle differences in a spatial context beyond the scope of conventional shape recognition, presenting an interesting and unique challenge to existing object detection algorithms. Consequently, developing advanced detection techniques that can handle these complexities is essential to enhance safety and efficiency in construction site operations. Despite this necessity, very few attempts in existing research have been made to recognize hanging objects. Additionally, although extensive research exists on object detection for construction equipment and workers, with many data sets available as online open sources, data sets for hanging objects are scarce. This lack of data sets restricts the research on hanging objects and remains a significant challenge.

To address this research gap, this study addresses the novel challenge of hanging object detection and provides a tailored image data set—**HangCon** (**Han**ging Objects in **Con**struction Sites)—for benchmarking purposes. The objective of this study is to develop a data set for detecting hanging objects on construction sites, as illustrated in Fig. 2. This paper describes the data collection process, the preprocessing techniques applied, the annotation methods used, and the benchmarking results of widely used object detection models. This study is anticipated to advance the development in this specialized field by offering a data set specifically for detecting hanging objects on construction sites. It also explores the fundamental aspects and challenges of hanging object detection, a topic not fully addressed by current object detection methods. The HangCon data set is expected to serve as a critical reference for ongoing research in this field. By deepening the understanding of the unique challenges associated with detecting hanging objects, HangCon could support the development of more effective safety measures and operational efficiencies on construction sites.

Background and Related Work

Object Detection Algorithms

Object detection is a core task in computer vision, involving the identification and localization of objects within an image or video. Over the years, significant advancements have been made to improve detection accuracy and efficiency. Early models such as Faster R-CNN (Ren et al. 2015), introduced the Region Proposal Network (RPN) for better detection, whereas Cascade R-CNN (Cai and Vasconcelos 2018) enhanced accuracy through multistage refinement. Single-stage models such as RetinaNet (Lin et al. 2017) addressed class imbalance with focal loss. The YOLO (You Only Look Once) (Redmon et al. 2016) series, including YOLOv5 and YOLOv8 Version 8.0.0, optimized real-time detection at impressive speed and accuracy. Keypoint-based anchor-free models such as CornerNet (Law and Deng 2018) and CenterNet (Zhou et al. 2019) innovated by focusing on object corners and centers, respectively. ATSS (Adaptive Training Sample Selection) (Zhang et al. 2020) improved robustness through adaptive sample selection. EfficientDet (Tan et al. 2020), balanced accuracy and efficiency using an optimized backbone and scalable feature fusion. Transformer-based models such as DETR (detection transformer) (Carion et al. 2020) and deformable-DETR (Zhu et al. 2021) introduced self-attention mechanisms that directly predict bounding boxes and handle varied object shapes and sizes. These models collectively advanced object detection, offering robust solutions for diverse applications, particularly in complex environments such as construction sites.

Vision-Based Monitoring in Construction

Vision-based monitoring in construction has become a modern alternative to traditional manual methods, significantly enhancing

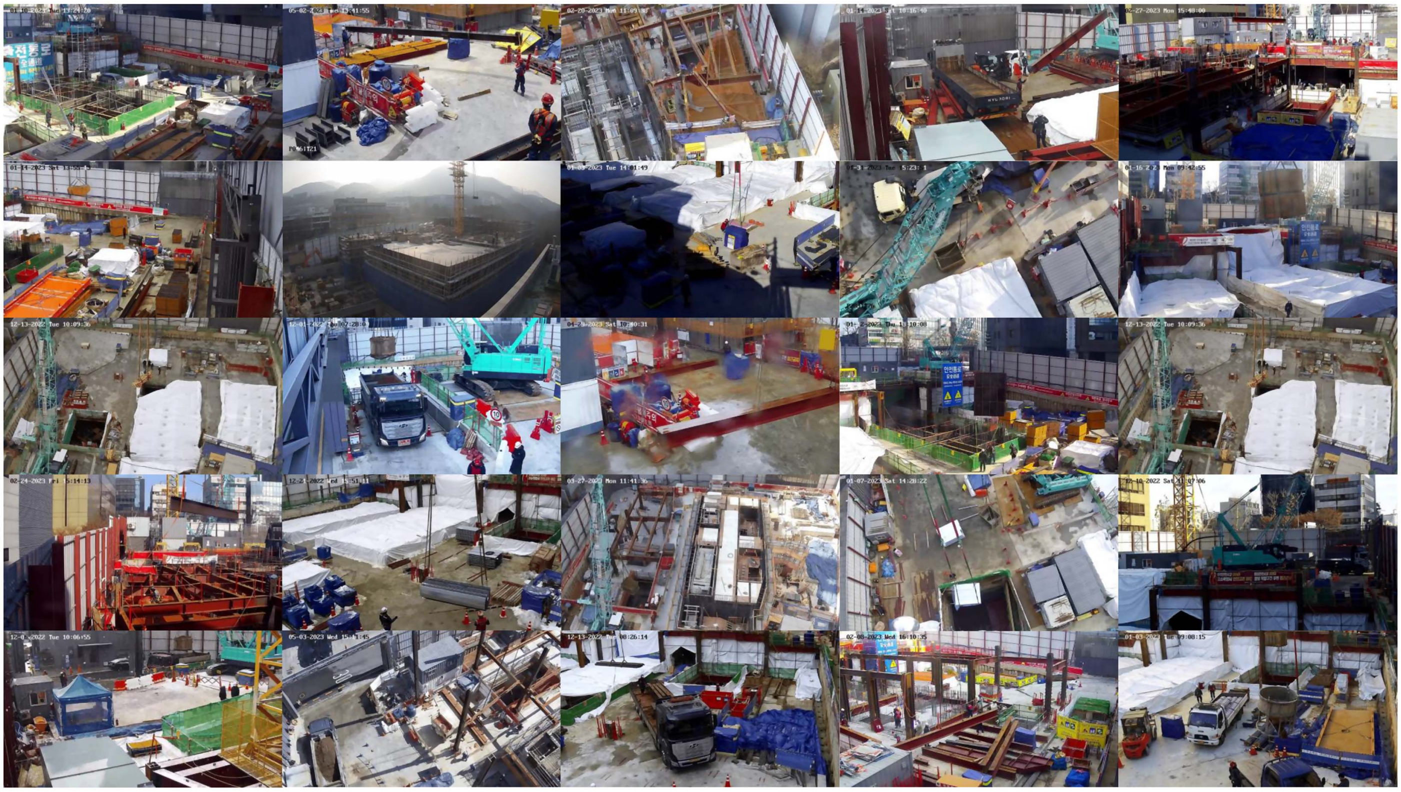

Fig. 2. Example images of HangCon data set. (Images by authors.)

safety and productivity. Object detection algorithms trained on specialized data sets have been widely used to detect and track construction-related objects, including equipment, environments, and workers. For example, Fang et al. (2018) and Lin et al. (2021) used improved R-CNN models to detect heavy equipment on construction sites, whereas Chen et al. (2023) applied zero-shot learning for activity recognition in earthmoving tasks. Additionally, Kim et al. (2018b) combined region-based fully convolutional networks (R-FCN) with transfer learning for robust equipment detection. Roberts and Golparvar-Fard (2019) introduced a deep learning-based method for detecting excavator activities. Lin et al. (2024) introduced a real-time struck-by hazards detection system using far-field surveillance videos, tailored for small- and medium-sized construction sites. Kim et al. (2023b) developed a small object detection system based on YOLOv5, enhancing real-time accuracy and flexibility in monitoring various object sizes on construction sites.

Regarding worker safety, several studies have focused on detecting construction workers and ensuring compliance with safety regulations. Tang et al. (2020) proposed a method to classify worker-tool interactions for safety compliance. Nath et al. (2020) and Wang et al. (2021) researched verifying PPE (personal protective equipment) compliance to enhance safety. Jeelani et al. (2021) also developed an automated system for real-time detection of hazardous conditions, including worker localization and hazard detection. Additionally, Khan et al. (2022) developed a smart safety hook monitoring system integrating computer vision and Internet of Things (IoT) to prevent falls from height. Kim et al. (2023a) introduced a real-time struck-by hazards detection system using far-field surveillance videos, tailored for small- and medium-sized construction sites. Despite the extensive research in vision-based monitoring for various construction activities, there has been a notable gap in studies focused on detecting hanging objects suspended by cranes. This specific area has not received as much attention, highlighting the need for dedicated research to improve safety and efficiency in construction sites where hanging objects pose significant risks.

Image Data Set for Vision-Based Monitoring in Construction

General image data sets such as Modified National Institute of Standards and Technology (MNIST) (Deng 2012), Pattern Analysis, Statistical Modeling, and Computational Learning Visual Object Classes (PASCAL) (Everingham et al. 2010), Microsoft Common Objects in Context (COCO) (Lin et al. 2015), and ImageNet (Deng et al. 2009) are renowned and expansive in the realm of object recognition and detection, encompassing a vast array of categories and applications. However, despite their extensive scope and acclaim, these data sets often lack categories specifically pertinent to the construction field. They do not adequately represent the unique aspects and requirements of construction environments, highlighting the need for specialized data sets in the construction domain.

Several construction-specific image data sets have been developed. The Harhat Wearing Detection (GDUT-HWD) (Wu et al. 2019) data set focuses on hardhats, reflecting the importance of head protection. The Color Helmet and Vest (CHV) (Wang et al. 2021) data set specializes in personal protective equipment and personnel, such as helmets and vests. The Moving Objects in Construction Sites (MOCS) (Xuehui et al. 2021) data set emphasizes moving objects across 13 categories, highlighting the dynamic nature of construction sites. The AI-Created Image Detection (ACID) (Xiao and Kang 2021a) data set targeted machines with 10 categories, underscoring their role in construction processes. Lastly, the SODA (Duan et al. 2022) data set included a broader

Table 1. Object detection data set and details in construction

Data set	Benchmark models	Categories	Classes	Images
GDUT-HWD (2019) (Wu et al. 2019)	Faster R-CNN SSD YOLOv3	Hardhats	1	3,174
CHV (2021) (Wang et al. 2021)	YOLOv3 YOLOv4 YOLOv5	Helmets Vest Person	6	1,330
MOCS (2021) (Xuehui et al. 2021)	YOLOv3 SSD RetinaNet FCOS NAS-FPN Faster R-CNN TridentFast	Moving objects	13	41,668
ACID (2021) (Xiao and Kang 2021a)	YOLOv3 Inception-SSD Faster-RCNN-ResNet101 R-FCN-ResNet101	Machines	10	10,000
SODA (2022) (Duan et al. 2022)	YOLOv3 YOLOv4	Workers Materials Machines	15	19,846

range of 15 categories, covering workers, materials, and machines. Table 1 summarizes these data sets, detailing their categories and image counts. These specialized data sets are invaluable for developing and benchmarking object detection models tailored to the unique challenges of construction environments, providing the essential data for accurate monitoring and analysis. Due to the critical importance of data sets, recent studies have also focused on generating synthetic data to overcome the challenges of constructing large, high-quality data sets for construction monitoring (Hwang et al. 2023; Lee et al. 2022; Shi et al. 2024). For vision-based monitoring research in the construction field, data set development is essential, and the establishment of benchmark data sets enables continuous and progressive research and development in this area.

Hanging Object Detection in Construction

Research on object detection in construction has primarily focused on equipment and workers, with less attention to the challenge of detecting hanging objects. Although extensive research exists on detecting aerial entities such as airplanes, unmanned aerial vehicles (UAVs), and drones (Ouerteteni et al. 2022; Ghosh et al. 2023; Rozantsev et al. 2015). These studies provide valuable insights and methodologies that can be adapted for different contexts, particularly for detecting objects in the air. Most algorithms developed for aerial object detection rely on motion-based techniques, utilizing a sequence of images to track movement over time. However, these methods face significant challenges when applied to construction settings in which hanging objects often remain stationary or exhibit minimal movement, making it difficult to identify them through motion alone. Detecting hanging objects in construction environments presents unique requirements that highlight the necessity for single-image and shape-based object detection algorithms. Traditional CNN architectures struggle with detecting small or camouflaged objects, particularly in the complex backgrounds typical of construction sites. This makes the detection of hanging objects more challenging compared to aerial entities.

In the construction field, studies have applied approaches to recognize objects suspended by tower cranes. Chian et al. (2022) attempted to apply computer vision techniques to detect crane loads and identify fall zones or areas in which the load could fall during an accident. Jeong et al. (2023) conducted research to detect curtain walls suspended by tower cranes and track their installation process. However, these studies were limited to specific sites at which camera positions and ground information were known or targeted specific pretrained objects and did not apply methods to detect any hanging object present in the image. To address these challenges, this study introduces the HangCon data set as a foundational step toward improving hanging object detection in construction environments.

Problem Formulation

The task of hanging object detection is fundamentally similar to regular object detection, which involves predicting the location and type of all detected objects predefined in the classes. However, the key difference is that the target classes are specifically hanging objects or hanging ropes, irrespective of the type or shape of the object. For example, a box is considered a positive detection (hanging object) if it is suspended in the air; however, it is no longer positive once it is on the ground (see Fig. 3). This distinction highlights the importance of context and the object's interaction with elements such as cranes or ropes. Thus, the task is not merely about recognizing the visual shape of an object but about understanding its status based on these interactions.

Formally, each hanging object is represented by a bounding box $\mathrm{B} = (x, y, w, h)$, where x and y denote the center coordinates of the bounding box, and w and h represent its width and height, respectively. The primary goal of this task is to precisely predict the set of bounding boxes $B_1, B_2, \ldots, B_n$ for each hanging object within a given image I. This task involves not only detecting the presence of such objects but also accurately determining their spatial location within the complex and dynamic environment of construction sites. This added complexity necessitates advanced detection techniques capable of discerning subtle contextual differences to ensure accurate and reliable monitoring in construction settings.

(a) (b)

Fig. 3. Example images illustrating object shape invariance: (a) object on ground; and (b) object hanged from rope. (Images by authors.)

The task of detecting hanging objects on construction sites presents several significant challenges in existing object detection models.

1. Object shape invariance regardless of hanging status: When the same object is either on the ground or suspended in the air, its visual shape often remains unchanged, even though its status is different. This "shape invariance" presents a significant challenge for general object detection models, which typically rely on shape to classify objects. As a result, models may struggle to distinguish between stationary and hanging objects. In the context of hanging object detection, however, models cannot depend solely on shape to determine whether an object is stationary or hanging. Instead, models must consider the object's spatial relationship with its surroundings to accurately differentiate between stationary and suspended objects. As shown in Fig. 3, the same object appears on the ground in one instance and hanging in another, underscoring the need for a contextual understanding rather than relying solely on shape.
2. Complex backgrounds: Environments with numerous overlapping objects, varying textures, and similar colors or shapes are a significant challenge in general object detection tasks. These complex backgrounds hinder the detection process, particularly for identifying ropes, which often blend into the background or are obscured by objects with similar shapes, colors, or textures. The presence of multiple overlapping objects, varying lighting conditions, and the similarity of ropes to other elongated elements such as cables, wires, or structural components further exacerbate the difficulty. Fig. 4 illustrates these challenges, showing two images where hanging objects and ropes are not clearly visible due to the complex background.
3. Diversity in hanging objects: The "hanging object" class includes a broad range of items commonly found on construction sites, such as steel frames, materials, boxes, rebar, soil, and exterior components. This category's diversity introduces significant challenges because objects with varying shapes, sizes, and textures all need to be accurately classified under a single category. Traditional object detection methods, which rely on recognizing specific shapes or textures, struggle to address this level of variability. The significant variation within the "hanging object" class poses a substantial challenge, highlighting the need for advanced detection algorithms capable of generalizing across different forms and appearances.

These challenges highlight the need for specialized approaches in hanging object detection. The complexity of construction sites and the unique nature of hanging objects require more sophisticated models, such as improved background segmentation, context-aware object recognition, and enhanced feature representation. Developing these techniques is vital for improving safety and efficiency on construction sites. To effectively develop and research hanging object detection models, a dedicated image data set is essential. A large-scale, high-quality data set with detailed annotations is necessary to address the unique challenges of hanging object detection. This data set should capture a wide range of construction site environments, objects, and ropes under various

Fig. 4. Example images where hanging objects and ropes are not clearly visible due to complex background. (Images by authors.)

lighting conditions and angles to ensure robust and generalizable models. To address these challenges in hanging object detection tasks on construction sites, this study has developed a specialized image data set called HangCon. This large-scale, high-quality data set includes detailed annotations and captures a variety of construction site environments, scenarios, and diverse objects and ropes.

HangCon Data Set

It begins with the data collection process, followed by preprocessing steps, the annotation procedure, and the data splitting step.

Data Collection

The data collection process for the HangCon data set considered several key factors to ensure authenticity and relevance to real-world construction settings. Images were required to depict complex backgrounds typical of actual construction sites, avoiding simplistic or clear sky backdrops. Only real photographic images were included, excluding synthetic or illustrated ones. Additionally, it was essential that the objects in the images were hanging to accurately represent real construction scenarios.

The data set was collected using six surveillance cameras installed across two different construction sites for 102 days. At the first site, five cameras were positioned at various views and angles to capture detailed scenes of activities at the construction site, whereas the remaining camera was installed in a different building at the second site to provide a broader perspective. Although data were collected from only two sites, each camera was strategically placed at different heights and angles, allowing for the capture of a wide range of construction activities, backgrounds, and objects given various weather conditions, including sunny, rainy, and cloudy days. This diverse setup enhanced the data set's comprehensiveness by capturing seasonal variations and a variety of environmental conditions.

Data collection took place during typical construction hours, specifically from 7:00 a.m. to 5:00 p.m., Monday to Friday, with no recordings on holidays due to the absence of construction activity. This allowed for the inclusion of diverse lighting conditions and shadows, further contributing to the data set's authenticity. The cameras focused on various hanging objects, varying in size and shape, to ensure a diverse and representative data set. These considerations during the data collection phase ensure that the HangCon data set accurately reflects the complex and dynamic nature of construction sites, capturing the diversity of real-world environments. The data collection process resulted in a rich and diverse data set of 101,381 images (50,842 hanging objects and 50,539 nonhanging objects), divided into training (57,704 images), validation (6,457 images), and test (37,220 images) sets. The varied camera angles and extended collection period provided a robust foundation for developing and testing advanced hanging object detection models, ensuring the data set's applicability to real-world scenarios. A rich and diverse data set of 101,381 images was produced by the data collection process. The varied camera angles and extended collection period provided a robust foundation for developing and testing advanced hanging object detection models, ensuring the data set's applicability to real-world scenarios. The data set encompasses approximately 30 object classes, representing a wide range of objects commonly lifted on construction sites, including steel frames, materials, boxes, rebar, soil, and exterior finishes. As illustrated in Fig. 5, these classes cover various types of hanging objects captured in the HangCon data set.

Data Preprocessing

Data preprocessing involved several steps to ensure the data set's quality and usability. First, images were extracted from the raw surveillance videos at 30-s intervals to capture a broad range of scenarios. From the videos recorded each day, approximately 1,000 images (approximately 5% of all extracted images) contained at least one hanging object. To balance the data set, an equal number of images without any hanging objects was also extracted for each day. To ensure the accuracy of the data set, the presence of hanging objects in the images was manually verified by professional construction engineers. The initial resolution of images from cameras A, B, C, D, and F was $1{,}920 \times 1{,}080$, whereas camera E captured images at $1{,}280 \times 720$. For processing consistency and to reduce computational load during model training, all images were resized to 640×360.

Data Annotation

The extracted images were manually annotated for three classes: hanging object, rope, and hanging object with rope. A "hanging object" is defined as any type of object suspended by ropes from cranes. The "rope" class refers to the rope attached to the crane and holding the hanging object, with the bounding box marked from the connected part of the object to the hook. The "hanging object with rope" class includes both the hanging object and the rope, marked by bounding boxes around the edges of both classes. This combined annotation is performed after the initial annotation of the two individual classes using simple mathematical calculations.

The annotation process was carried out by 13 annotators using DarkLabel version 2.4, labeling software. Each class was annotated with bounding box coordinates (x center, y center, width, height) and class labels (hanging object, rope, hanging object with rope). To ensure compatibility with various object detection frameworks, the annotations were converted into both COCO and VOC formats. The COCO format includes comprehensive metadata, such as the image ID, category ID, bounding box coordinates, and segmentation information, which are highly useful for advanced object detection and segmentation tasks. The VOC format is simpler, providing essential information including the image ID, object class, and bounding box coordinates, and is widely used for basic object detection tasks. This thorough annotation process provided a detailed and structured data set ready for training and evaluating hanging object detection models.

Fig. 6 shows examples of the annotation for images, illustrating the three classes: hanging object, rope, and hanging object with rope. As shown in the figure, the hanging object is identified with a bounding box around the object itself. The rope is marked with a bounding box that extends from the connected part of the object to the hook; if the rope is long and cut off at the edge of the image, the bounding box extends to the image's end. The hanging object with a rope class is shown with a bounding box encompassing both the hanging object and the rope.

Data Split

The HangCon data set, consisting of a total of 101,381 images collected over 102 days using seven cameras, was divided for training, validation, and testing purposes. The data set was split into 57,704 images for training, 6,457 images for validation, and 37,220 images for testing. The train and validation data sets were randomly divided in a 9:1 ratio. A period-based split was applied to ensure robust evaluation. For five cameras (A to E), the first 77 days of data were used for training and validation, whereas the remaining 25 days, after maintaining a one-month gap, were used for testing.

Fig. 5. Various types of hanging objects in HangCon. (Images by authors.)

Fig. 6. Annotation examples: hanging object, rope, and hanging object with rope. (Images by authors.)

For one camera (F), the entire 102 days of data were used exclusively for testing. This strategic separation of test data prevents the model from memorizing object shapes by using images from different construction stages and cameras, ensuring a thorough and robust evaluation of the data set. Fig. 7 shows the number of images by camera class for the data set split and illustrates the distribution of training, validation, and test images across the different cameras.

Benchmark Design

Experiments were conducted to evaluate the quality and effectiveness of the proposed hanging object data set "HangCon," These experiments involved training and testing representative benchmark object detection models on the HangCon data set to assess their performance in detecting hanging objects. Additionally, an image classification task was performed to determine the presence or absence of hanging objects in the images. These experiments provided a comprehensive evaluation of the HangCon data set's utility and effectiveness for hanging object detection in construction site environments.

Evaluation Protocol for Object Detection

For the experiments with object detection, a variety of benchmark models were selected to evaluate the performance of the HangCon

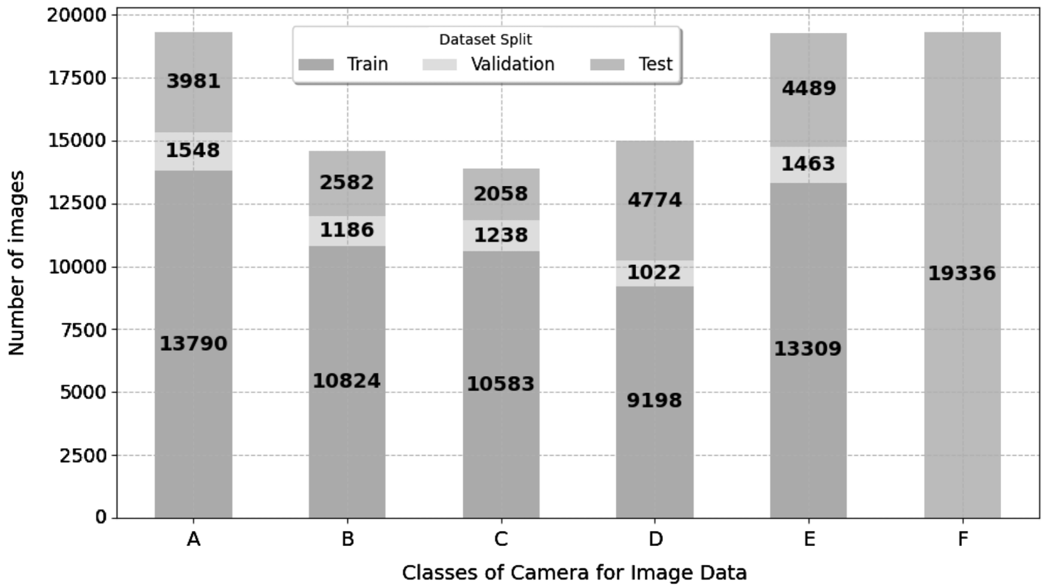

Fig. 7. Number of images by camera class for data set split.

data set. These models include Faster R-CNN (Ren et al. 2015), RetinaNet (Lin et al. 2017), Cascade R-CNN (Cai and Vasconcelos 2018), CornerNet (Law and Deng 2018), CenterNet (Zhou et al. 2019), DETR (Carion et al. 2020), ATSS (Zhang et al. 2020), YOLOv5, EfficientDet (Tan et al. 2020), Deformable-DETR (Zhu et al. 2021), and YOLOv8 Version 8.0.0. The chosen benchmark models encompass a diverse range of base backbones and represent different categories, such as anchor-based one-stage, anchor-based two-stage, anchor-free, and transformer-based models. This selection ensures a comprehensive evaluation of the HangCon data set across different state-of-the-art object detection frameworks. Table 2 provides detailed information on the hyperparameters used for training the selected object detection models. The table includes specifics on the base backbone, input image size, number of epochs, batch size, learning rate, and optimizer for each model. These hyperparameters were carefully chosen to ensure a fair comparison between models.

The benchmark models were trained using three different experimental setups to evaluate their performance under various annotation conditions. Experiment 1 used annotations for hanging objects only, and experiment 2 utilized annotation for both hanging objects and ropes. Experiment 3 focused on annotations for the combined class of hanging objects with ropes. To evaluate the performance of the object detection models, mean average precision (mAP) was used as the primary metric. The mAP was calculated over Intersection over Union (IoU) thresholds of 0.5, 0.75, and 0.95 for each class. These evaluation metrics provided a comprehensive understanding of the models' accuracy and precision in detecting and localizing objects. The evaluation could ensure that the performance of the models on the HangCon data set was thoroughly assessed.

Evaluation Protocol for Image Classification

The purpose of this section is to conduct image classification tasks to determine the presence or absence of hanging objects in images to assess the robustness and accuracy of existing models. This task involves evaluating the ability of the models to classify images based on the presence of hanging objects and understanding the contextual relationship between hanging objects and ropes. The data set used for these classification tasks is the same as the detection data set. The classification tasks involve two main criteria: confidence-based and overlap-based classification tasks.

First, for the confidence-based classification task, an image is classified as "hanging" if any "hanging object" is detected with a confidence score greater than or equal to 0.5. If the detection confidence score is less than this threshold, the image is labeled "no hanging object." Second, for the overlap-based classification task, an image is labeled "hanging" if there is an overlap between a detected "hanging object" with a confidence score of 0.5 or higher

Table 2. Information of hyperparameters for training selected object detection models

Models	Base backbone	Input image size	Epoch	Batch size	Learning rate	Optimizer
Faster RCNN (2015)	ResNet-50	640 × 360	20	16	0.0025	SGD
RetinaNet (2017)	ResNet-50	640 × 360	20	16	0.0025	SGD
Cascade RCNN (2018)	ResNet-50	640 × 360	20	16	0.0025	SGD
CornerNet (2018)	HourglassNet	640 × 360	20	16	0.001	Adam
CenterNet (2019)	ResNet-18	640 × 360	20	16	0.0025	SGD
Detr (2020)	ResNet-50	640 × 360	20	16	0.0025	SGD
ATSS (2020)	ResNet-50	640 × 360	20	16	0.0025	SGD
YOLOv5 (2020)	YOLO	640 × 640	20	16	0.01	SGD
EfficientDet (2020)	EfficientNet	640 × 360	20	16	0.0025	SGD
Deformable-Detr (2021)	ResNet-50	640 × 360	20	4	0.0025	AdamW
YOLOv8 (2023)	YOLO	640 × 640	20	16	0.01	SGD

Table 3. Performance result with mAP by object detection models for Experiment 1 and 2

	Exp.1 only object			Exp.2 object and rope					
	Hanging object mAP (%)			Hanging object mAP (%)			Rope mAP (%)		
Models	@0.5	@0.75	@0.5:0.95	@0.5	@0.75	@0.5:0.95	@0.5	@0.75	@0.5:0.95
Faster RCNN	55.65	46.29	38.50	54.75	43.76	36.68	50.50	30.65	28.91
RetinaNet	44.64	34.64	29.50	45.83	33.61	28.47	45.42	31.29	22.46
Cascade RCNN	53.81	45.38	37.58	54.14	43.58	36.51	52.41	32.67	30.47
CornerNet	63.41	54.36	47.12	65.02	56.04	48.26	63.21	49.23	43.72
CenterNet	45.77	27.04	25.66	47.06	26.33	25.39	44.11	18.67	21.38
Detr	50.86	39.50	25.08	47.99	24.02	20.91	45.38	21.69	18.78
ATSS	47.24	37.85	32.54	46.65	37.63	31.93	46.76	28.87	27.62
YOLOv5	61.10	49.44	42.08	62.65	48.44	41.79	62.81	40.40	37.75
EfficientDet	54.45	40.03	34.73	55.59	34.06	31.35	48.85	27.21	27.15
YOLOv8	71.63	59.61	51.46	72.81	58.54	50.55	72.25	55.86	49.16

and a "rope" with a confidence score of 0.3 or higher. If no overlap is detected, the image is labeled "no hanging object." This means that if a "hanging object" is detected with a high confidence score but there is no corresponding "rope" detected with a confidence score of 0.3 or higher, the image is still labeled as "no hanging object." The lower threshold of 0.3 for rope detection is set in anticipation of the challenge in accurately detecting ropes in construction environments due to the prevalence of similar objects, such as electrical cables, hoses, chains, wires, flexible tubing, straps, webbing, and steel frames, which can complicate the detection process.

The performance of the image classification tasks is assessed using classification accuracy, precision, recall, and F1 score. These metrics provide a comprehensive evaluation of the models' effectiveness in accurately classifying the presence or absence of hanging objects, considering both detection precision and the contextual understanding of the objects within the images.

Results

Results of Object Detection

The experiments conducted using the HangCon data set aimed to evaluate its effectiveness by training models with three versions of the data set: (1) with annotations for only hanging objects, (2) with annotations for both hanging objects and ropes, and (3) with annotations for a combined box that includes both the object and the rope. Tables 3 and 4 presents the performance results for Experiments 1 to 3, highlighting the mAP values for the combined class of hanging objects with ropes across different IoU thresholds.

Across the three versions of the data set, the range of mAP values at the 0.5 IoU threshold varied significantly. In Experiment 1, in which only hanging objects were annotated, the mAP values for hanging objects ranged from 44.64% to 71.63%. In Experiment 2, with annotations for both hanging objects and ropes, the mAP values for hanging objects ranged from 45.83% to 72.81% and from 45.38% to 72.25% for ropes. Experiment 3, which used combined annotations for hanging objects with ropes, showed mAP values for the combined class ranging from 51.09% to 76.01%. The mAP varied significantly across different detection models. Models such as YOLOv8 consistently exhibited higher mAP values across all versions and thresholds, whereas others such as ATSS and DETR showed more variability and generally lower performance.

A comparison of Experiments 1 and 2 generally showed a slight increase in AP values when the rope was included. For instance, YOLOv8 showed a subtle increase from 71.63% in Experiment 1 to 72.81% in Experiment 2, indicating that the presence of the rope might aid in better defining the context of the object, slightly enhancing detection performance. In comparison with Experiment 3, the detection models generally exhibited higher AP values when the elements were annotated as a combined class. This suggests that training models on combined annotations can enhance their ability to understand and detect interconnected objects more effectively. For example, YOLOv8's performance peaked at 76.01% at 0.5 IoU, showcasing significant improvement over its performance in previous versions.

Results of Image Classification

This section presents the results of the experiments conducted to classify the presence of hanging objects in images and the overlap between hanging objects and ropes. These experiments aimed to evaluate the models' performances in detecting hanging objects and understanding their contextual relationship with ropes. To classify the presence of hanging objects, an image was classified as "hanging" if any detected "hanging object" had a confidence score of 0.5 or higher. Table 5 presents the results of this experiment, showing F1 scores ranging from 62.96% to 78.66% for Experiment 1 (without rope) and from 60.50% to 80.74% for Experiment 2 (with rope). Among the various methods tested, YOLOv8 emerged as the top performer, demonstrating the highest F1 scores. The comparison between Experiment 1 (without rope) and Experiment 2 (with rope) shows that models incorporating rope labeling generally exhibited slightly higher accuracy, suggesting that the additional context provided by ropes may help in better delineating

Table 4. Performance result with mAP by object detection models for Experiment 3

	Exp.3 hanging object with rope		
	mAP (%)		
Models	@0.5	@0.75	@0.5:0.95
Faster RCNN	58.01	48.92	43.20
RetinaNet	54.72	46.59	41.12
Cascade RCNN	58.05	49.86	44.12
CornerNet	68.46	53.81	46.23
CenterNet	51.09	38.26	33.4
Detr	53.66	47.01	44.69
ATSS	56.60	48.00	43.13
YOLOv5	65.95	53.89	48.24
EfficientDet	55.66	42.71	40.36
YOLOv8	76.01	68.07	61.08

Table 5. Performance result with confidence-based classification task

	Exp.1 only object (%)				Exp.2 object and rope (%)			
Models	Accuracy	Precision	Recall	F1 score	Accuracy	Precision	Recall	F1 score
Faster RCNN	64.05	60.72	77.76	68.19	64.72	61.57	76.67	68.30
RetinaNet	62.97	62.01	65.28	63.61	64.84	66.23	59.29	62.57
Cascade RCNN	61.10	58.09	77.24	66.31	63.25	59.97	77.75	67.71
CornerNet	62.71	60.27	73.43	66.04	64.27	62.59	71.24	66.19
CenterNet	67.92	63.75	68.00	67.75	68.99	69.58	66.52	68.01
Detr	52.31	51.02	64.71	66.32	53.47	57.13	61.55	66.11
ATSS	66.10	68.66	58.13	62.96	66.17	71.80	52.27	60.50
YOLOv5	76.93	89.60	60.48	72.21	77.44	91.13	60.36	72.62
EfficientDet	65.61	64.47	68.20	66.29	68.04	70.22	61.68	65.67
YOLOv8	80.41	89.76	69.42	78.66	83.68	92.81	73.06	80.74

Table 6. Performance result with overlap-based classification task

	Classification according to overlap (%)			
Models	Accuracy	Precision	Recall	F1 score
Faster RCNN	66.66	55.46	71.84	62.60
RetinaNet	70.90	62.88	57.43	60.03
Cascade RCNN	69.16	58.92	67.95	63.11
CornerNet	68.91	59.09	65.74	61.91
CenterNet	73.51	67.56	61.17	64.21
Detr	57.57	47.20	78.24	58.88
ATSS	63.92	61.36	59.17	59.21
YOLOv5	81.74	88.04	61.31	72.29
EfficientDet	74.85	72.71	56.41	63.53
YOLOv8	86.11	88.19	66.67	75.72

hanging objects, thus improving the model's ability to correctly classify images.

In the experiment of classifying images based on the overlap between detected hanging objects and ropes, it was assumed that hanging objects are typically connected to ropes, resulting in overlapping bounding boxes in the image data. An image was classified as "hanging" if there was an overlap between the object and rope (with the object detected at a confidence threshold of 0.5 or higher, and the rope at 0.3 or higher) and "nonhanging" otherwise. Table 6 provides the classification results, which show F1 scores ranging from 58.88% to 75.72%. YOLOv8 stood out as the top performer, achieving the highest F1 score of 75.72%.

Discussion

This study employed various benchmark models to investigate the detection of hanging objects using HangCon data set. Despite the advanced capabilities of these models, they frequently encountered challenges in accurately distinguishing hanging objects due to their diverse shapes and the complex backgrounds typical of construction sites. This complexity arises from the variability in the appearance of objects and the dynamic environments in which they are situated. To address these challenges, the potential of rope detection as a supplementary method was explored, positing that identifying ropes associated with hanging objects could provide additional contextual information to improve detection accuracy. However, as detailed in Table 3 comparing Experiments 1 and 2, incorporating a rope class during training did not significantly enhance performance. The primary obstacles in rope detection include variations in rope thickness, colors that blend into the background, and frequent partial occlusions, leading to higher rates of false positives or missed detections.

Conversely, Experiment 3, in which both hanging objects and ropes were annotated as a single class, showed a notable improvement in detection performance. This unified approach leverages the contextual relationship between ropes and objects, significantly enhancing the model's ability to recognize and accurately classify hanging objects. Combining the detection of objects and ropes into a single classification criterion significantly improved the accuracy of the image classification tasks, as shown in Tables 5 and 6. Classifying images as "hanging" based on the overlap between hanging objects and ropes resulted in better performance metrics. These findings highlight the advantage of incorporating contextual relationships into classification models, particularly in construction environments in which such contextual elements are prevalent. By detecting ropes in conjunction with hanging objects, models gain additional contextual information that enhances their ability to differentiate between complex background elements and the objects of interest.

Based on the experimental results using the HangCon data set, this study proposes several improvements to enhance the detection of hanging objects in construction environments.

1. Scene understanding with semantic segmentation and depth estimation: Integrating semantic segmentation and depth estimation can significantly enhance detection accuracy by better distinguishing between objects and their backgrounds. Semantic segmentation helps differentiate hanging objects from the background, whereas depth estimation provides spatial context, improving the overall detection precision.
2. Improved rope detection: The experiments highlight the need for more precise rope detection models. Current methods using bounding boxes capture unnecessary background details. Developing models that specifically detect rope lines would yield cleaner and more accurate data for object detection.
3. Recognizing structural connections: Enhancing algorithms to interpret the physical connections between hanging objects and their supporting structures can improve detection accuracy. Understanding these structural relationships is crucial for accurately classifying and detecting hanging objects.
4. Data set augmentation using synthetic data: To further improve detection across diverse object shapes and sizes, augmenting the HangCon data set with synthetic data is recommended. Doing so would broaden the range of detectable objects and enhance model performance by providing a more varied training set.

Implementing these improvements would lead to more robust and accurate detection systems, enhancing safety and efficiency on construction sites. The suggested advancements, such as enhanced scene understanding, improved rope line detection, recognizing structural connections, and data set augmentation, collectively elevate the precision of hanging object detection. The HangCon data set could serve as a foundational benchmark for this research, providing a comprehensive resource for training and evaluating models. This data set could be instrumental in developing and refining detection systems, ensuring they can handle the complexities of real-world construction environments. Consequently, these advancements could pave the way for automated monitoring systems capable of real-time analysis and response, further contributing to site safety and operational efficiency. The HangCon data set represents a crucial first step, and its continued refinement and application could play a pivotal role in advancing the field of hanging object detection.

Conclusion

This study addresses hanging object detection within construction environments, a domain that differs significantly from general object detection due to the varied sizes, shapes, and complex settings of the objects involved. To overcome these challenges, the HangCon data set was introduced, comprising 101,381 images that capture the varied conditions and contexts of hanging objects at construction sites. To validate the applicability and effectiveness of HangCon, experiments were conducted using 10 benchmark object detection models for both detection and classification tasks. The experimental results showed that when only hanging objects were annotated, the best mAP value was 71.63%. In contrast, using unified annotations of hanging objects and ropes, the best mAP value improved to 76.01% at 0.5 IoU. Additionally, classification experiments indicated enhanced accuracy in identifying the presence of hanging objects when contextual information from ropes was included. In the classification of the presence of hanging objects, the best F1 score improved from 78.66% without rope annotations to 80.74% with rope annotations.

HangCon represents a novel contribution to the field of construction monitoring as the first data set specifically curated for hanging object detection. This data set offers a meticulously assembled collection of annotated images that capture the intricate details of hanging objects and their associated ropes. The specialized focus of HangCon distinguishes it from existing construction data sets, establishing it as a critical benchmark for the training and evaluation of detection models. The comprehensive scope of the data set ensures that it encompasses the diverse and complex conditions typical of real-world construction sites, rendering it an invaluable resource for advancing research in this domain. By providing detailed annotations and encompassing a wide range of construction scenarios, HangCon lays a solid foundation for future investigations aimed at enhancing detection accuracy, thereby significantly improving safety and operational efficiency on construction sites. The HangCon data set has the potential to significantly enhance existing construction monitoring systems by enabling a more accurate detection of hanging objects. This capability could improve the effectiveness of safety monitoring by helping to identify and manage hazardous zones and could facilitate more efficient coordination of lifting operations. As a result, integrating HangCon into these systems may lead to safer working environments and more streamlined on-site operations.

Future work should focus on both improving the HangCon data set and developing specialized detection models using this data set. Enhancements to the HangCon data set should include expanding its scope beyond specific construction sites, increasing camera usage, and extending the data collection period. Additionally, the data set should be continuously enhanced by adding new subcategories and objects. This expansion will involve selecting subsets of the data set based on factors such as environmental conditions, object characteristics, and the unique challenges of construction sites. Additionally, using HangCon, future study should aim to develop advanced detection models specifically tailored for hanging objects. These models should integrate enhanced scene understanding through semantic segmentation and depth estimation, improve rope line detection, and recognize structural connections between hanging objects and their supports. By addressing these areas, future study could create more robust and accurate detection systems, significantly enhancing safety management and operational efficiency on construction sites.

Data Availability Statement

Some or all data, models, or code generated or used during the study are available in a repository online in accordance with funder data retention policies. The HangCon data set and related resources can be accessed at https://github.com/gilsujeong/HangCon.

Acknowledgments

This research was supported by the BK21 FOUR (Fostering Outstanding Universities for Research) project in 2024 (No. 4120200113771), Artificial Intelligence Institute at Seoul National University, and the Institute of Engineering Research at Seoul National University. The authors express their gratitude for the support.

References

Ahn, C. R., S. Lee, and F. Peña-Mora. 2012. “Monitoring system for operational efficiency and environmental performance of construction operations using vibration signal analysis.” In *Proc., Construction Research Congress 2012: Construction Challenges in a Flat World*, 1879–1888. Reston, VA: ASCE. https://doi.org/10.1061/9780784412329.189.

Ali, A. H., T. Zayed, R. D. Wang, and M. Y. S. Kit. 2024. “Tower crane safety technologies: A synthesis of academic research and industry insights.” *Autom. Constr.* 163 (Jul): 105429. https://doi.org/10.1016/j.autcon.2024.105429.

Azar, E. R. 2016. “Construction equipment identification using marker-based recognition and an active zoom camera.” *J. Comput. Civ. Eng.* 30 (3): 04015033. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000507.

Cai, Z., and N. Vasconcelos. 2018. “Cascade R-CNN: Delving into high quality object detection.” In *Proc., 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, 6154–6162. New York: IEEE.

Carion, N., F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko. 2020. “End-to-end object detection with transformers.” Preprint, submitted May 26, 2020. https://arxiv.org/abs/2005.12872.

Chen, C., B. Xiao, Y. Zhang, and Z. Zhu. 2023. “Automatic vision-based calculation of excavator earthmoving productivity using zero-shot learning activity recognition.” *Autom. Constr.* 146 (Feb): 104702. https://doi.org/10.1016/j.autcon.2022.104702.

Chian, E. Y. T., Y. M. Goh, J. Tian, and B. H. W. Guo. 2022. “Dynamic identification of crane load fall zone: A computer vision approach.” *Saf. Sci.* 156 (Dec): 105904. https://doi.org/10.1016/j.ssci.2022.105904.

Deng, J., W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. 2009. “ImageNet: A large-scale hierarchical image database.” In *Proc., 2009 IEEE Conf. on Computer Vision and Pattern Recognition*, 248–255. New York: IEEE.

Deng, L. 2012. “The MNIST database of handwritten digit images for machine learning research [best of the web].” *IEEE Signal Process. Mag.* 29 (6): 141–142. https://doi.org/10.1109/MSP.2012.2211477.

Duan, R., H. Deng, M. Tian, Y. Deng, and J. Lin. 2022. “SODA: A large-scale open site object detection dataset for deep learning in construction.” *Autom. Constr.* 142 (Oct): 104499. https://doi.org/10.1016/j.autcon.2022.104499.

Everingham, M., L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. 2010. "The pascal visual object classes (VOC) challenge." *Int. J. Comput. Vision* 88 (2): 303–338. https://doi.org/10.1007/s11263-009-0275-4.

Fang, Q., H. Li, X. Luo, L. Ding, T. M. Rose, W. An, and Y. Yu. 2018. "A deep learning-based method for detecting non-certified work on construction sites." *Adv. Eng. Inf.* 35 (Jan): 56–68. https://doi.org/10.1016/j.aei.2018.01.001.

Ghosh, S., J. Patrikar, B. Moon, M. M. Hamidi, and S. Scherer. 2023. "AirTrack: Onboard deep learning framework for long-range aircraft detection and tracking." In *Proc., 2023 IEEE International Conference on Robotics and Automation (ICRA)*, 1277–1283. New York: IEEE.

Golparvar-Fard, M., A. Heydarian, and J. C. Niebles. 2013. "Vision-based action recognition of earthmoving equipment using spatio-temporal features and support vector machine classifiers." *Adv. Eng. Inf.* 27 (4): 652–663. https://doi.org/10.1016/j.aei.2013.09.001.

Huang, C., W. Li, W. Lu, F. Xue, M. Liu, and Z. Liu. 2021. "Optimization of multiple-crane service schedules in overlapping areas through consideration of transportation efficiency and operational safety." *Autom. Constr.* 127 (Jul): 103716. https://doi.org/10.1016/j.autcon.2021.103716.

Hwang, J., J. Kim, and S. Chi. 2023. "Site-optimized training image database development using web-crawled and synthetic images." *Autom. Constr.* 151 (Jul): 104886. https://doi.org/10.1016/j.autcon.2023.104886.

Jeelani, I., K. Asadi, H. Ramshankar, K. Han, and A. Albert. 2021. "Real-time vision-based worker localization & hazard detection for construction." *Autom. Constr.* 121 (Jan): 103448. https://doi.org/10.1016/j.autcon.2020.103448.

Jeong, G., M. Jung, S. Park, M. Park, and C. R. Ahn. 2024. "Contextual multimodal approach for recognizing concurrent activities of equipment in tunnel construction projects." *Autom. Constr.* 158 (Feb): 105195. https://doi.org/10.1016/j.autcon.2023.105195.

Jeong, I., J. Hwang, J. Kim, S. Chi, B.-G. Hwang, and J. Kim. 2023. "Vision-based productivity monitoring of tower crane operations during curtain wall installation using a database-free approach." *J. Comput. Civ. Eng.* 37 (4): 04023015. https://doi.org/10.1061/JCCEE5.CPENG-5105.

Khan, M., R. Khalid, S. Anjum, S. V.-T. Tran, and C. Park. 2022. "Fall prevention from scaffolding using computer vision and IoT-based monitoring." *J. Constr. Eng. Manage.* 148 (7): 04022051. https://doi.org/10.1061/(ASCE)CO.1943-7862.0002278.

Kim, H., S. Bang, H. Jeong, Y. Ham, and H. Kim. 2018a. "Analyzing context and productivity of tunnel earthmoving processes using imaging and simulation." *Autom. Constr.* 92 (Aug): 188–198. https://doi.org/10.1016/j.autcon.2018.04.002.

Kim, H., H. Kim, Y. W. Hong, and H. Byun. 2018b. "Detecting construction equipment using a region-based fully convolutional network and transfer learning." *J. Comput. Civ. Eng.* 32 (2): 04017082. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000731.

Kim, H., K. Kim, and H. Kim. 2016. "Data-driven scene parsing method for recognizing construction site objects in the whole image." *Autom. Constr.* 71 (Nov): 271–282. https://doi.org/10.1016/j.autcon.2016.08.018.

Kim, H., J. Seong, and H.-J. Jung. 2023a. "Real-time struck-by hazards detection system for small- and medium-sized construction sites based on computer vision using far-field surveillance videos." *J. Comput. Civ. Eng.* 37 (6): 04023028. https://doi.org/10.1061/JCCEE5.CPENG-5238.

Kim, S., S. H. Hong, H. Kim, M. Lee, and S. Hwang. 2023b. "Small object detection (SOD) system for comprehensive construction site safety monitoring." *Autom. Constr.* 156 (Dec): 105103. https://doi.org/10.1016/j.autcon.2023.105103.

Law, H., and J. Deng. 2018. "CornerNet: Detecting objects as paired keypoints." In *Proc., European Conf. on Computer Vision (ECCV)*, 734–750. Dordrecht, Netherlands: Springer.

Lee, J. G., J. Hwang, S. Chi, and J. Seo. 2022. "Synthetic image dataset development for vision-based construction equipment detection." *J. Comput. Civ. Eng.* 36 (5): 04022020. https://doi.org/10.1061/(ASCE)CP.1943-5487.0001035.

Li, Y., and C. Liu. 2012. "Integrating field data and 3D simulation for tower crane activity monitoring and alarming." *Autom. Constr.* 27 (Nov): 111–119. https://doi.org/10.1016/j.autcon.2012.05.003.

Lin, S. S., S. L. Shen, A. Zhou, and Y. S. Xu. 2021. "Risk assessment and management of excavation system based on fuzzy set theory and machine learning methods." *Autom. Constr.* 122 (Feb): 103490. https://doi.org/10.1016/j.autcon.2020.103490.

Lin, T.-Y., P. Goyal, R. Girshick, K. He, and P. Dollar. 2017. "Focal loss for dense object detection." Preprint, submitted August 7, 2017. https://arxiv.org/abs/1708.02002.

Lin, T.-Y., M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár. 2015. "Microsoft COCO: Common objects in context." Preprint, submitted May 1, 2014. https://arxiv.org/abs/1405.0312.

Lin, X., Z. Guo, H. Guo, and Y. Zhou. 2024. "Vision-based real-time posture tracking for multiple construction workers." *J. Comput. Civ. Eng.* 38 (4): 04024023. https://doi.org/10.1061/JCCEE5.CPENG-5816.

Nath, N. D., A. H. Behzadan, and S. G. Paal. 2020. "Deep learning for site safety: Real-time detection of personal protective equipment." *Autom. Constr.* 112 (Apr): 103085. https://doi.org/10.1016/j.autcon.2020.103085.

Ouerteteni, M. E. H., A. Zaafouri, T. Delleji, M. Bouchouicha, Z. Chtourou, and M. Sayadi. 2022. "Flying objects classification using trajectory images and convolutional neural network." In *Proc., 2022 IEEE Information Technologies & Smart Industrial Systems (ITSIS)*, 1–5. New York: IEEE.

Redmon, J., S. Divvala, R. Girshick, and A. Farhadi. 2016. "You only look once: Unified, real-time object detection." Preprint, submitted June 8, 2015. https://arxiv.org/abs/1506.02640.

Ren, S., K. He, R. Girshick, and J. Sun. 2015. "Faster R-CNN: Towards real-time object detection with region proposal networks." In *Proc., Advances in Neural Information Processing Systems*. New York: Curran Associates.

Roberts, D., and M. Golparvar-Fard. 2019. "End-to-end vision-based detection, tracking and activity analysis of earthmoving equipment filmed at ground level." *Autom. Constr.* 105 (Sep): 102811. https://doi.org/10.1016/j.autcon.2019.04.006.

Rozantsev, A., V. Lepetit, and P. Fua. 2015. "Flying objects detection from a single moving camera." In *Proc., 2015 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 4128–4136. New York: IEEE.

Sherafat, B., C. R. Ahn, R. Akhavian, A. H. Behzadan, M. Golparvar-Fard, H. Kim, Y.-C. Lee, A. Rashidi, and E. R. Azar. 2020. "Automated methods for activity recognition of construction workers and equipment: State-of-the-art review." *J. Constr. Eng. Manage.* 146 (6): 03120002. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001843.

Shi, M., C. Chen, B. Xiao, and J. Seo. 2024. "Vision-based detection method for construction site monitoring by integrating data augmentation and semisupervised learning." *J. Constr. Eng. Manage.* 150 (5): 04024027. https://doi.org/10.1061/JCEMD4.COENG-14388.

Tan, M., R. Pang, and Q. V. Le. 2020. "EfficientDet: Scalable and efficient object detection." Preprint, submitted November 20, 2019. https://arxiv.org/abs/1911.09070.

Tang, S., D. Roberts, and M. Golparvar-Fard. 2020. "Human-object interaction recognition for automatic construction site safety inspection." *Autom. Constr.* 120 (Dec): 103356. https://doi.org/10.1016/j.autcon.2020.103356.

Wang, J., Q. Zhang, B. Yang, and B. Zhang. 2023. "Vision-based automated recognition and 3D localization framework for tower cranes using far-field cameras." *Sensors* 23 (10): 4851. https://doi.org/10.3390/s23104851.

Wang, Z., Y. Wu, L. Yang, A. Thirunavukarasu, C. Evison, and Y. Zhao. 2021. "Fast personal protective equipment detection for real construction sites using deep learning approaches." *Sensors* 21 (10): 3478. https://doi.org/10.3390/s21103478.

Wu, J., N. Cai, W. Chen, H. Wang, and G. Wang. 2019. "Automatic detection of hardhats worn by construction personnel: A deep learning approach and benchmark dataset." *Autom. Constr.* 106 (Oct): 102894. https://doi.org/10.1016/j.autcon.2019.102894.

Xiao, B., and S.-C. Kang. 2021a. "Development of an image data set of construction machines for deep learning object detection." *J. Comput.*

Civ. Eng. 35 (2): 05020005. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000945.

Xiao, B., and S.-C. Kang. 2021b. "Vision-based method integrating deep learning detection for tracking multiple construction machines." *J. Comput. Civ. Eng.* 35 (2): 04020071. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000957.

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

Yang, J., P. Vela, J. Teizer, and Z. Shi. 2014. "Vision-based tower crane tracking for understanding construction activity." *J. Comput. Civ. Eng.* 28 (1): 103–112. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000242.

Yong, Y. P., S. J. Lee, Y. H. Chang, K. H. Lee, S. W. Kwon, C. S. Cho, and S. W. Chung. 2023. "Object detection and distance measurement algorithm for collision avoidance of precast concrete installation during crane lifting process." *Buildings* 13 (10): 2551. https://doi.org/10.3390/buildings13102551.

Zhang, C., and A. Hammad. 2012. "Improving lifting motion planning and re-planning of cranes with consideration for safety and efficiency." *Adv. Eng. Inf.* 26 (2): 396–410. https://doi.org/10.1016/j.aei.2012.01.003.

Zhang, S., C. Chi, Y. Yao, Z. Lei, and S. Z. Li. 2020. "Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection." In *Proc., 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9756–9765. New York: IEEE.

Zhou, X., D. Wang, and P. Krähenbühl. 2019. "Objects as points." Preprint, submitted April 16, 2019. https://arxiv.org/abs/1904.07850.

Zhu, X., W. Su, L. Lu, B. Li, X. Wang, and J. Dai. 2021. "Deformable DETR: Deformable transformers for end-to-end object detection." Preprint, submitted October 8, 2020. https://arxiv.org/abs/2010.04159.