Project Leadership and Society 5 (2024) 100149

Contents lists available at ScienceDirect

Project Leadership and Society

journal homepage: www.sciencedirect.com/journal/project-leadership-and-society

Empirical Research Paper

Deep learning-based computer vision in project management: Automating indoor construction progress monitoring

Biyanka Ekanayake [a,*], Johnny Kwok Wai Wong [a], Alireza Ahmadian Fard Fini [a], Peter Smith [a], Vishal Thengane [b]

[a] School of Built Environment, University of Technology Sydney, 15 Broadway, Ultimo, NSW, 2007, Australia
[b] School of Computer Science and Electronic Engineering, Alan Turing Building, University of Surrey, Guildford, Surrey, United Kingdom

ARTICLE INFO

Keywords:
Computer vision
Construction projects
Deep learning
Digitalization
Indoor construction progress monitoring
Project management

ABSTRACT

Progress monitoring is crucial for effective project management, particularly in construction projects. The adoption of computer vision with deep learning expedites automation, accuracy, and efficiency in construction progress monitoring by overcoming the challenges of laborious, and error prone manual methods. While there is growing attention on developing computer vision based deep learning models for construction progress monitoring, deployment platforms for project managers are lacking. Using computer vision, this study develops a Mask Recurrent Convolutional Neural Network deep learning model. It utilizes progress images of drywall construction from two indoor construction sites and tests the model on a third indoor site in Sydney, Australia. The model is capable of automated as-built visual detection and work-in-progress measurement. The study also provides an understanding on the deployment process of the deep learning model on a cloud-based platform called Streamlit. By developing a model tailored for automatically quantifying work-in-progress of indoor construction elements and detailing the process of deploying that model on a cloud-based platform, this study significantly advances digitalization of construction project management. Project managers, stand to benefit from these advancements by gaining access to more accurate and automated construction progress monitoring for better decision-making.

1. Introduction

In project management, effective progress monitoring is crucial for cost-effective resource use, milestone achievement, deadline adherence, risk mitigation, and informed decision-making (Navon, 2007; Raymond and Bergeron, 2008). Progress monitoring in construction projects, involves collecting, analyzing, reporting, and visualizing progress information of construction elements and activities (Bohn and Teizer, 2010; Golparvar-Fard et al., 2015). Inefficient and inaccurate work-in-progress monitoring significantly contributes to delay and budget overruns in construction projects (Omar et al., 2018). Previous research reveals that more than 53% of construction projects encounter delays and more than 66% of projects undergo budget overruns (Han et al., 2015)

Traditional progress monitoring processes are error prone, time consuming, and visually unfriendly (Golparvar-Fard et al., 2015; Yang et al., 2015). The digital transformation in project management has accelerated the adoption of computer vision technologies, providing greater accuracy and efficiency compared to traditional methods (Bednar and Welch, 2020; Li et al., 2021; Reja et al., 2022). Computer vision-based progress monitoring involves capturing site images with cameras and analyzing them using algorithms. Deep learning, a subset of machine learning, plays a crucial role in this process (Szeliski, 2010; Zhang et al., 2009; Moragane et al., 2024). Recent studies by Li et al. (2021), Wei et al. (2022), and Zhao et al. (2023) demonstrate effective computer vision applications with deep learning in automated construction progress monitoring. Despite these achievements, automated progress monitoring in indoor construction activities continues to present unique challenges for project managers in relation to recognizing the as-built state of construction elements and calculating work-in-progress (Wong et al., 2024). These difficulties arise from obstructions, clutter, variable lighting conditions, and the achromatic appearance of indoor components (Deng et al., 2020; Ekanayake et al., 2021a). Despite growing interest in training deep learning models for construction progress monitoring, there is also a notable lack of research

* Corresponding author.
E-mail address: biyanka.ekanayake@uts.edu.au (B. Ekanayake).

https://doi.org/10.1016/j.plas.2024.100149
Received 25 December 2023; Received in revised form 6 August 2024; Accepted 20 August 2024
Available online 22 August 2024

on how to deploy such advanced computer vision-based deep learning models cost effectively and efficiently in practice (Ekanayake et al., 2024; Li et al., 2021). Deployment involves making such trained models available for real-world applications on cloud or local computer applications to track progress, which is essential for project managers (Khorasani et al., 2022).

The current study first explores how computer vision can be leveraged to automatically recognize and calculate work-in-progress of indoor construction elements using a deep learning model, which is built upon Mask Recurrent Convolutional Neural Network (Mask R-CNN) (He et al., 2017) and trained on Google Colaboratory platform (Google Colab). The model uses progress images focusing on interior drywalls installation from two construction sites and is tested on a third site in Sydney, Australia. Then this study provides an understanding on the deployment process of the deep learning model on a cloud-based platform called Streamlit (2024), to facilitate the transition from computer vision models to practical applications. Therefore, this study intends to address the research question: How can deep learning-based computer vision techniques be effectively leveraged to automatically quantify work-in-progress for indoor construction elements?

Followed by section 1 on the introduction, section 2 provides a literature review to the study. Section 3 presents the research methods and section 4 describes the development of the computer vision and deep learning-based model and the platform on which it is deployed. In section 5, a discussion is presented, followed by theoretical contributions and practical implications in section 5.1. The paper ends with the conclusions and future directions in section 6.

2. Literature review

This section explores the challenges of traditional construction progress monitoring for the project managers, and how state-of-the-art progress monitoring approaches have increased automation, accuracy, and efficiency. This literature review also delves into training and deploying computer vision-based deep learning models.

2.1. Challenges of traditional construction progress monitoring

Traditional construction progress monitoring relies heavily on daily site reports and manual inspections (Wei et al., 2022), involving walk-through assessments (Ham et al., 2016). These methods depend on approximate visual assessments for recognizing as-built states and measuring work completion (Golparvar-Fard et al., 2015; Zhang et al., 2009), leading to subjective interpretations and inaccurate reporting (Golparvar-Fard et al., 2015). Errors and inefficiencies in traditional methods contribute to schedule overruns of 20% and budget overruns of up to 80% (Agarwal et al., 2016). Enhancing digitalization in progress monitoring with automated approaches ensures accurate, timely decision-making and reduces errors (Zhang et al., 2009).

2.2. State-of-the-art construction progress monitoring approaches

The integration of digital technologies in construction sites has marked a significant shift towards more efficient and accurate progress monitoring methods (Kopsida et al., 2015; Wong et al., 2024). The use of digital cameras, initially for site security, has laid the groundwork for the adoption of site images based visual inspection techniques, enabling construction progress monitoring through computer vision (Ahmadian Fard Fini et al., 2022; Hamledari et al., 2017). Beyond visual data, the construction industry has leveraged 3D reconstructions from laser scanning to enhance volumetric data collection (Bosché et al., 2015). Additionally, construction object tracking using radio wave-based methods, such as Radio Frequency Identification (RFID) (Valero et al., 2015) and geospatial technologies (El-Omari and Moselhi, 2011) to capture positional data are other methods of construction progress monitoring. The Internet of Things (IoT) technologies have also enabled the development of digital twins for enhanced construction progress monitoring (Qureshi et al., 2020; Soman et al., 2017; Sacks et al., 2020).

Recent studies in automated construction progress monitoring have been using building information models (BIM) to track discrepancies with 4D BIM focusing on time management and with 5D BIM emphasising on cost management (Kopsida et al., 2015; Reja et al., 2022). Off-the-shelf products such as “Reconstruct” leverage computer vision integration with 4D and 5D BIM to provide real-time, actionable insights (Reconstruct, 2024). However, these advancements face several challenges that hinder widespread adoption. For 4D BIM and 5D BIM, challenges include the dependency on accurate and continuous data input to maintain up-to-date schedules, which can be disrupted by site conditions and unforeseen delays (Rao et al., 2022; Sacks et al., 2020). Both 4D and 5D BIM integrations face limitations in data accuracy and the need for line-of-sight in mapping, which can be problematic in cluttered or indoor environments (Rao et al., 2022). Where sensors are used to collect data, high computational costs associated with processing sensory data for BIM integration further complicate real-time applications of these technologies (Sacks et al., 2020). More importantly, while BIM excels in as-planned modelling and visualizing progress deviations, it often lacks comprehensive coverage of all construction stages, particularly indoor activities (Pal et al., 2022; Wei et al., 2022).

2.3. The use of computer vision for indoor construction progress monitoring

The adoption of computer vision-based technology for construction progress monitoring, particularly in indoor environments, represents a significant technological advancement, offering distinct advantages over traditional monitoring methods and other technologies. Leveraging 2D images (Hamledari et al., 2017; Deng et al., 2020) and 3D point clouds (Armeni et al., 2016; Zhu et al., 2023), computer vision emulates human visual inspection but with greater accuracy and efficiency. By analyzing these images through sophisticated image processing algorithms, computer vision technologies monitor construction progress (Moragane et al., 2024).

Computer vision-based approach is increasingly favoured for indoor construction sites due to its ability to overcome the limitations posed by physical constraints in the cluttered and enclosed indoor environments, with low and variable lighting conditions that typically challenge other data capturing technologies such as laser scanning, radio-based and geo spatial-based techniques (Hamledari and McCabe, 2016). Despite these advantages, implementing computer vision indoors is not without its challenges. Technical issues such as occlusions, clutter, and the varying intensity of artificial lighting significantly impact the quality of visual input data, requiring advanced preprocessing to mitigate noise and ensure accurate feature extraction and object recognition (Ekanayake et al., 2021b). Additionally, the practical limitations concerning the use of unmanned aerial vehicles (UAVs) or unmanned ground vehicles (UGVs) in densely populated indoor spaces and uncertainties in unsupervised camera movements affect object recognition (Wong et al., 2023). Ekanayake et al. (2021a) categorised all these challenges into technical challenges related to indoor objects, lighting conditions and camera positioning. When these challenges are effectively addressed, computer vision stands out as a superior method for indoor construction progress monitoring.

2.4. Deep learning algorithms for computer vision approaches

Current computer vision-based approaches, particularly using deep learning models such as convolutional neural networks (CNNs), excel in object detection, classification, localization, and segmentation from images (Kardovskyi and Moon, 2021; Chollet, 2017; LeCun et al., 2015). The recent developments in computer vision-based indoor construction progress monitoring have harnessed the detection and classification ability of Mask R-CNN for recognizing indoor elements with irregular shapes using 2D images (Ying and Lee, 2019). The segmentation ability of Mask R-CNN has been used to quantify work completion (Wei et al., 2022) with the use of depth images. The contributions of Armeni et al. (2016) and Zhu et al. (2023) play significant roles in advancing the field of indoor construction progress monitoring through their focus on 3D semantic parsing, multi-object relocalization, and reconstruction in changing 3D environments. Due to its high detection accuracy and instance segmentation performance, Mask R-CNN, proposed by He et al. (2017), remains a widely used algorithm (He et al., 2022; Xi et al., 2020).

2.5. Training and deploying of deep learning models

Training deep learning models with thousands of images necessitates high computational resources such as GPUs, RAM, CPU, and storage (O'Mahony et al., 2019; Wang et al., 2021). Advances in cloud computing have made virtualization platforms available to overcome hardware and configuration issues (Fang et al., 2018). Virtualization allows for creating high-performance virtual machines with dedicated resources borrowed from host computers or cloud providers to run deep learning models (Canesche et al., 2021). Google Colab (Google Research, 2024), a freely available virtual machine hosted by Google, offers a cost-effective and popular solution for deep learning model training, utilizing web-based Jupyter notebooks stored in Google Drive (Carneiro et al., 2018; Ohkawara et al., 2021). After training, deploying deep learning models makes them accessible to users without programming backgrounds (Li et al., 2020a,b). Deployment involves making trained models available for real-world applications (Khorasani et al., 2022). The freely available Streamlit open-source application framework allows for creating interactive applications with simple Python scripts and deploying them directly to Streamlit Cloud or running them locally, supporting collaborative purposes (Prapas et al., 2021; Li et al., 2020a,b; Shukla et al., 2021). By deploying a deep learning-based construction progress monitoring model built upon Mask R-CNN, project managers without programming expertise can make informed decisions (Streamlit, 2024).

The adoption of such technological innovations is critical in advancing project management practices (Forget et al., 2022) because integration of advanced technologies into project management processes can significantly enhance efficiency and decision-making capabilities (Papadonikolaki and Galera Zarco, 2023). The use of cloud-based and virtualized environments, such as Google Colab, facilitates scalable and accessible model training, aligning with the industry's move towards digital transformation (Li et al., 2020a,b). Moreover, the deployment of user-friendly applications like those developed with Streamlit makes sophisticated tools accessible to a wider range of non-technical stakeholders, facilitating broader adoption and utilization (Kebao and Jinling, 2021). Despite the advancements in computer vision for construction progress monitoring, no prior studies have specifically demonstrated how virtual environments such as Colab for training and Streamlit for deployment can be utilized for indoor progress monitoring. This study's methodological approach, described in section 3, addresses the technical challenges of training and deploying complex deep learning models.

3. Research methodology

Design science research (DSR) framework has been selected as the methodological framework for this study because the study focuses on addressing a practical problem by developing and deploying a technology-based artefact (Hevner et al., 2004; Iivari and Venable, 2009). The research process of this study adheres to the principles of DSR, incorporating four key cycles within the framework: i) relevance, ii) rigor, iii) design and iv) change and impact cycles (Drechsler and Hevner, 2016).

i) *The relevance cycle:* The relevance cycle starts with identifying the practical need for automated progress monitoring tools in the construction industry, specifically the indoor construction environment. This need is articulated through a comprehensive literature review (in Section 2), which reveals the inefficiencies and inaccuracies associated with traditional progress monitoring methods, the state-of-the art approaches, computer vision and deep learning-based solutions for indoor construction and the importance of training and deploying such solutions. The requirements for the artefact, including the ability to handle various lighting conditions, occlusions, and clutter in indoor environments, are established based on this identified need and are presented in Section 3.1.

ii) *The rigor cycle:* The rigor cycle then ensures that the artefact's design is well founded on existing knowledge bases integrating methodologies from established works of computer vision and deep learning. The research design for training and deploying the deep learning model is presented in Section 3.2.

iii) *The design cycle:* In the design cycle, the artefact of this study, i.e. the computer vision based indoor progress monitoring model, built upon Mask R-CNN deep learning model, and deployed on Streamlit is developed. This involves the process of designing, training, optimizing, and fine tuning the Mask R-CNN model on Colab based on the images collected from two indoor sites as presented in Section 4. The model is then tested on indoor images collected from a third project, ensuring that it meets the specified requirements linking back to the relevance cycle.

iv) *The change and impact cycle:* The change and impact cycle acknowledges the dynamic and evolving nature of construction sites. This cycle captures the broader organizational and societal changes triggered by the deployment of the artefact. The deployment process of the model proposed in this study aims to improve the accuracy and efficiency of progress monitoring, thereby enhancing overall project management practices. This cycle also considers the potential for broader adoption and integration of the technology within the construction industry, contributing to digital transformation and improved project outcomes.

The current study primarily aligns with Level 1 of Gregor and Zwikael's (2024) DSR approach for project management research, which outlines three levels of design knowledge: Level 1 (Artefact development), Level 2 (Nascent design principles), and Level 3 (Theory development). This alignment is due to the study's focus on developing a deep learning-based computer vision model to automate the progress monitoring of indoor construction elements. This automation addresses the challenges of manual progress monitoring and provides project managers with more accurate data, enhancing decision-making. Additionally, the study incorporates some elements of Level 2 for establishing design principles that can guide the broader application of computer vision models in construction project management. These principles are informed by the challenges of deploying deep learning-based computer vision models in dynamic environments such as indoor construction sites and are articulated in section 3.1 and 3.2.

3.1. Data collection from case projects

Researchers have used case projects to demonstrate, develop, and test prototypes in real-life situations (Fellows and Liu, 2015; Stake, 2013). In automated construction progress monitoring studies, the use of case projects for site images collection is the standard practice for algorithmic models training and demonstration (Kim et al., 2018; Wong et al., 2023).

Three indoor construction sites were selected as case projects to collect progress images for training and demonstrating the performance of the deep learning model. All sites were located in Sydney, Australia.

Images of the construction process of framing, insulation, and drywall installation of indoor walls were captured as a representative set of indoor construction progress images to train the deep learning model. Indoor walls are a significant part of interior construction and delays related to them can have costly consequences (Kropp et al., 2018). The progress data capturing, and information extraction related to indoor walls can aid many trades such as site managers, framers, insulation installers and drywall installers. Therefore, indoor walls are perhaps the most significant obstacle in automating indoor construction progress monitoring due to anticipated occlusions, clutter, materials diversity, and varying lighting during their construction process. To infer the as-built state and work-in-progress for an indoor wall, information related to installation of components such as frames, insulation blankets and plasterboards is required (Hamledari et al., 2017).

The Human Ethics Research Committee of the University of Technology Sydney approved the ethics application (ETH20-5459) for data collection. Project 1 is a residential building project with 2 floor levels. A total of 703 images were collected for the training dataset. Project 2 is a commercial building project with fit-out work for office workspace, with 3 floor levels. A total of 675 images were collected for the training dataset. Project 3 is alterations and additions to an existing dwelling, with 2 floor levels and a total of 100 images were collected. A total of 1378 images (703 from Project 1 and 675 from Project 2) were collected for training. The images from Project 3 were used exclusively for testing. Some images captured from the indoor sites are presented in Fig. 1a, b and 1c respectively.

The method for selecting the case study projects in this study aligns with the approach adopted in Deng et al. (2020); Ying and Lee (2019) and Wei et al. (2022). These studies on indoor construction progress monitoring indicate that the nature and number of case projects do not affect the model's training effectiveness, provided that a sufficient number of images, up to 2000 are used to capture the features for object detection. For example, Deng et al. (2020) used an indoor tiling image (1000 images) set from an indoor site. In the study by Wei et al. (2022), Mask R-CNN deep learning model was employed for indoor construction progress monitoring, utilizing 500 images of indoor bricklaying, and plastering scenario from the same site for both training and testing the model. Past research works on deep learning, including Chollet (2017), Goodfellow et al. (2016), LeCun et al. (2015), emphasized that the primary goal of training a deep learning model is to capture the relevant features of objects. If the dataset is sufficiently large and diverse, it allows the model to learn a comprehensive representation of the features, regardless of the specific nature or number of projects (Chollet, 2017; Goodfellow et al., 2016; LeCun et al., 2015).

To capture image, two Brinno TLC200 Pro time-lapse cameras were installed at each site, positioned to account for different lighting conditions. Cameras were installed at the best vantage points considering the availability of the contractor's resources, site facilities and surrounding structures. One camera was placed to monitor areas with artificial lighting, and the other captured images where natural sunlight seeped through openings. This setup ensures that the dataset includes variations in lighting conditions. Additionally, the cameras were strategically placed at different angles to cover various perspectives of the construction sites, capturing a comprehensive range of visual data. This approach also helped in documenting the presence of occlusions, such as construction materials or equipment temporarily blocking parts of the view. These are presented in Fig. 2. To further mitigate overfitting by adding variations to training image set, data augmentation techniques were applied using the "imgaug" library (imgaug, 2022). This involved photometric distortions (adjusting hue, saturation, brightness, contrast, and noise) and geometric distortions (random cropping, flipping, scaling, and rotating) to create a more varied training dataset (Bochkovskiy et al., 2020; Shorten and Khoshgoftaar, 2019). An example of an augmentation is illustrated in Fig. 3. Such augmentations help ensure that the model can generalize well to new images by simulating a wide range of possible variations and conditions that might be encountered in real-world scenarios (Rice et al., 2020). An additional 200 images from the image dataset were randomly augmented making the training dataset 1,578 images. These strategies ensured that training image dataset incorporate diverse features.

While deep learning algorithms are proven to be accurate and efficient for feature extraction compared to conventional machine learning algorithms (O'Mahony et al., 2019), pre-processing with a low level of manual intervention may be required depending on the visual quality of the image to address the challenges in indoor construction sites (Wei et al., 2022). In this study, two pre-processing algorithmic techniques, perspective transformation and low illumination enhancement were employed.

Because of the positioning of the camera, the region of interest of the site images are affected by distortions thus causing an inclined and angular view. Using the *cv2.getPerspectiveTransform* and then *cv2.warpPerspective* functions of OpenCV (OpenCV, 2022), the original perspective can be modified to obtain an accurate region of interest. The low illumination problem in indoor sites contributes to creating dark images from which the feature extraction is challenging unless the images are corrected with lighting enhancements. The dual channel (dark and bright) based method for low illumination image enhancement proposed by Shi et al. (2018) was used in this study. Some of the results are presented in Fig. 4.

In addition to the strategies to ensure that the training image dataset is diverse, implementing algorithmic techniques to pre-process the images from the indoor construction sites contributes to robust feature extraction. The methodological choice of optimized Mask R-CNN model trained on data collected from two indoor sites and tested on a third site

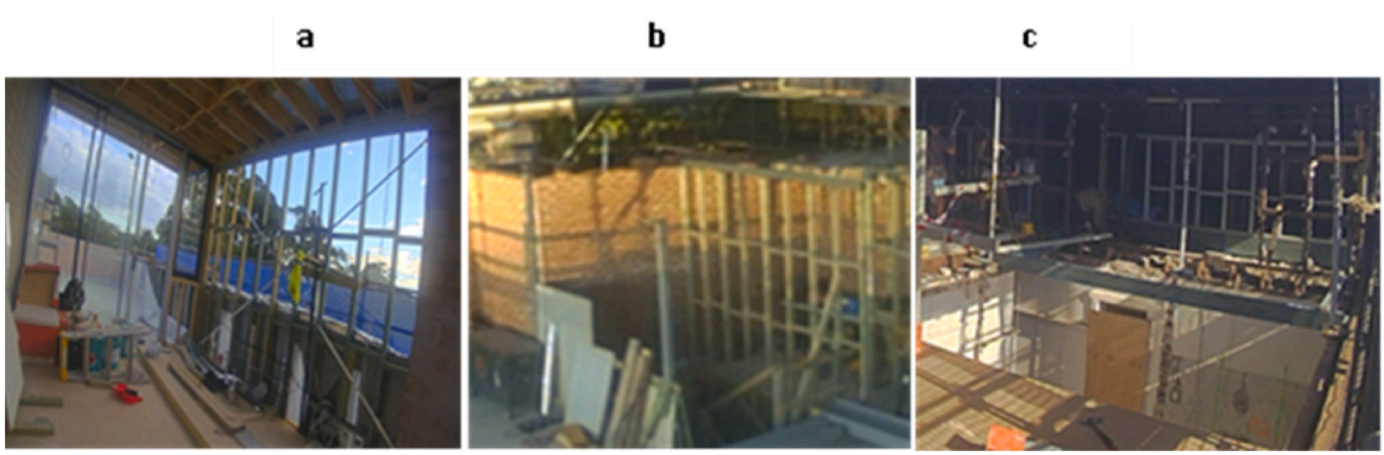

Fig. 1. Case projects: a) Project 1; b) Project 2; c) Project 3.

Fig. 2. Indoor location with a) Natural lighting; b) Artificial lighting c) Construction personnel and material creating occlusions.

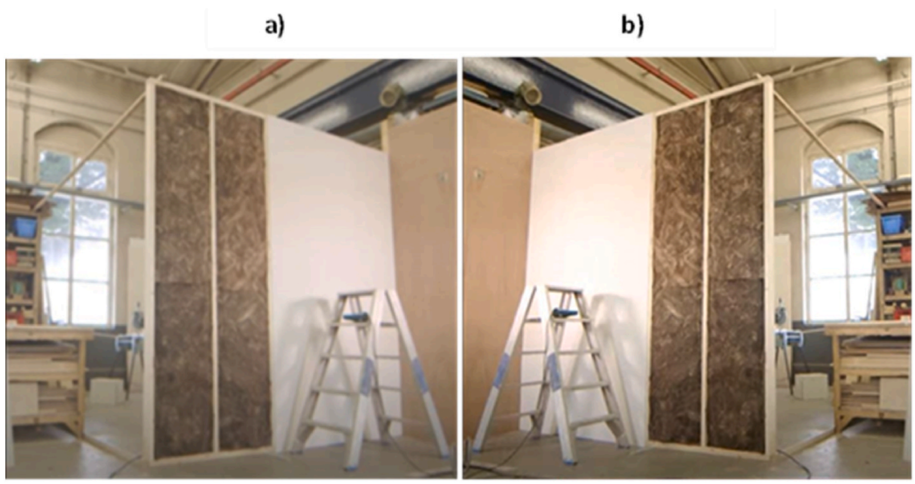

Fig. 3. a) Original image; b) Image flipped vertically and contrast enhanced.

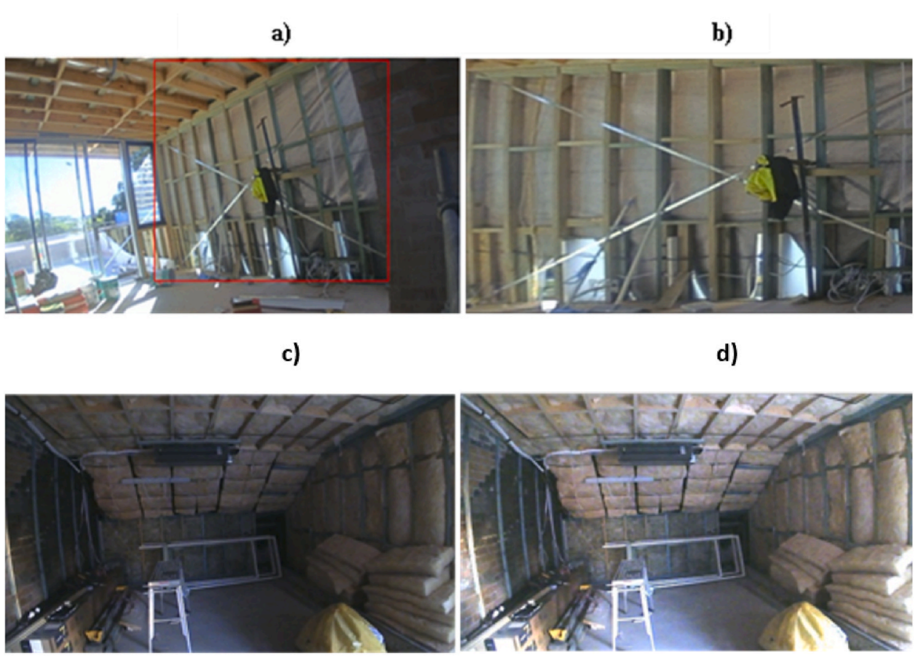

Fig. 4. Perspective transformation: a) Original image; b) Corrected image and Low illumination enhancement: c) Original image; d) Enhanced image.

ensures that the model's performance is tested in a completely new environment. Thereby, this serves as a form of validation. Validation in this context refers to testing the model's performance on an unseen dataset to ensure it can accurately recognize and segment objects that it has not encountered during training. Generalizability refers to the model's ability to perform well across a wide range of diverse scenarios and conditions beyond the specific datasets it was trained and tested on. In the current study, data augmentation techniques and dataset with diverse lighting conditions, occlusions, and various angles were incorporated to expose the model to a broad spectrum of variations, thereby enhancing its ability to generalize to new and different indoor construction scenarios. While testing the model on a third site and the strategies mentioned above indicate model's effectiveness on unseen data, it does not conclusively prove its generalizability across all possible indoor construction environments.

3.2. Research design for training and deploying the deep learning model

The deep learning-based approach presented in this study was built on a Mask R-CNN model. Mask R-CNN predicts segmentation masks in addition to the existing branches to predict object classification and bounding box generation-based detection (He et al., 2017). The network architecture of Mask R-CNN, with the key elements is illustrated in Fig. 5.

The Mask R-CNN model is built upon the ResNet architecture, specifically utilizing the ResNet-FPN backbone for feature extraction. The ResNet-FPN backbone's top-down architecture with lateral connections creates a feature pyramid from a single-scale input, enhancing its ability to recognize partially obscured objects (He et al., 2017). This ResNet architecture, known for its robustness in object detection and instance-based segmentation tasks, effectively manages occlusions (Chilukuri et al., 2022).

The neural network of the Mask R-CNN model is built with different hyperparameters. These hyperparameters can be fine-tuned depending on the requirement of training efficiency and model accuracy for optimization (Goodfellow et al., 2016). Programmers generally conduct many experiments to determine the optimum set of hyperparameters that improve the accuracy of the model (Smith, 2018). In this study, the hyperparameters that were experimented to optimize the Mask R-CNN base model are the learning rate (LR), learning rate scheduler and optimizer because they are responsible for the "learning process" of the deep learning model. The learning rate of the base Mask R-CNN model is 0.02 and the default optimizer is stochastic gradient descent (SGD) (He et al., 2017). The learning rate scheduler, Multiple steps learning rate (MultiStepLR) has been used in the Mask R-CNN optimization. The cosine annealing learning rate scheduler (CosineLR) (Loshchilov and Hutter, 2016) was experimented for the current study. The Adam optimizer (Kingma and Ba, 2014) is an extension to SGD that has recently been adopted in deep learning models to improve optimization.

After devising the method to build the deep learning model, the training of the deep learning model is a crucial step. As discussed in Section 2.5, Colab was used as the training platform. To train deep learning models on Colab, initially, the annotated training images with the annotation files are uploaded to Google Drive. Then, Google Drive is connected to the Colab notebook. In the next step, the training command is executed. Because of the fast GPU and RAM performance in Colab, the training process is not as time consuming as a deep learning model trained on a local computer. Test images are uploaded to Google Drive for inferring detection, classification, and segmentation results.

After the deep learning model is trained, certain metrices are used to evaluate the performance of the model. The current study used the mean average precision *(mAP)* to evaluate the detection accuracy of deep learning model (Chollet, 2017). In addition to mAP, when training deep learning models, loss value indicates how well a certain model behaves after each iteration. The reduction of loss after each or several iterations is an indication of the higher accuracy of the deep learning model (Goodfellow et al., 2016).

To deploy the developed deep learning model, Streamlit framework can be used either to run the deep learning model on the local machine or on the Stremlit cloud (Khorasani et al., 2022). With Streamlit cloud, programmers can simply connect their GitHub repository and then click deploy because Streamlit operates by reading code directly from a GitHub repository, that is made available to the public (Khorasani et al., 2022). The GitHub repository containing the construction site images and the code related to the deep learning model was not made available to the public because the UTS ethics application approval for this project ensures the confidentiality of the data collected. However, the web application can be built on a local Streamlit server for prototyping and testing through a Python runtime environment with all the dependencies. The research process described in Sections 3.1 and 3.2. from site image capture to the generation of useful outputs for project managers is displayed as a graphical summary in Fig. 6.

4. Results and findings

This section explains how the optimized Mask R-CNN model was

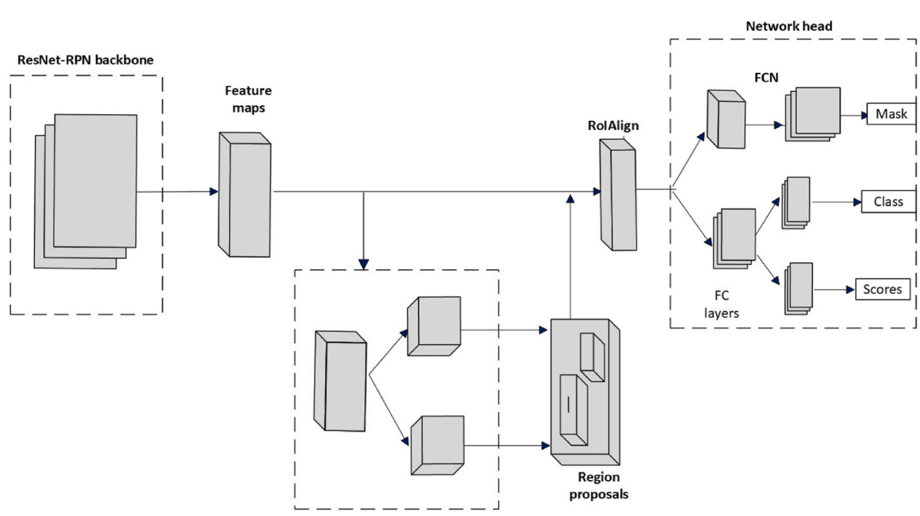

Fig. 5. Mask R-CNN network architecture.
Adapted from: He et al. (2017).

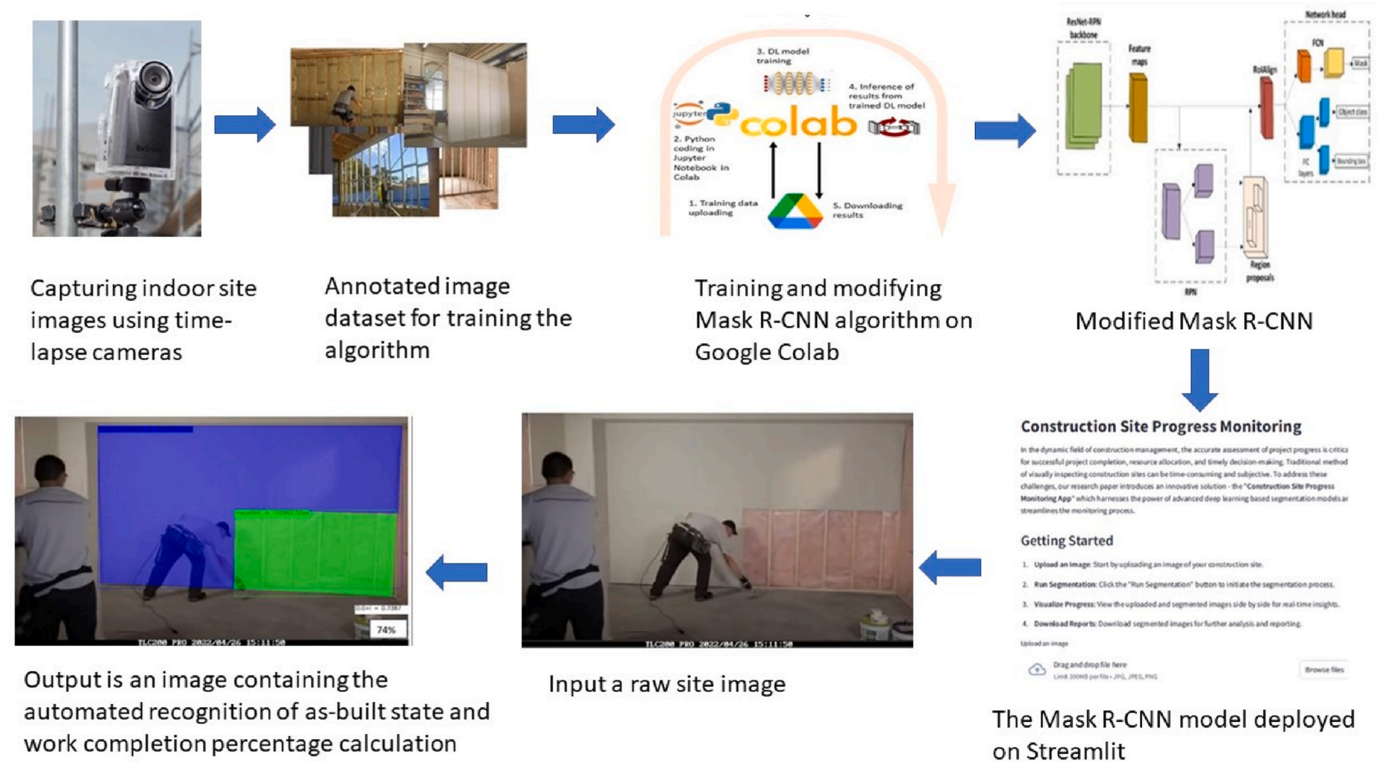

Fig. 6. Summary of the research process.

developed for automated visual recognition of as-built elements and their work completion percentage calculation for indoor progress monitoring. Deploying this Mask R-CNN model on the Streamlit platform is also presented.

4.1. The workflow of optimizing mask R-CNN and training on Colab

The Mask R-CNN model was optimized in terms of the detection, classification and segmentation accuracy for the custom dataset comprised of framing, insulation, and drywall installation images of indoor walls. This optimized model was trained on Google Colab to reduce the high computational resource requirements associated with deep learning models. The hyperparameters that were changed to optimize the Mask R-CNN model are the learning rate, learning rate scheduler and optimizer. Many experiments were conducted to determine which learning rate provides the best performance with the Adam optimizer, compared to the learning rate of the base model with the SGD optimizer. The learning rate scheduler was also changed from MultiStepLR to cosine annealing (CosineLR). In addition to the learning rate, learning rate scheduler and optimizer, which are directly associated with the optimization of the deep learning model, other hyperparameters that affect the efficiency of training of the model were also changed as below.

- Number of images per batch based on the GPU = 2
- Size or the pixel resolution of the images = 800x800
- Iterations = 2000
- Batch size per image = 128
- Number of classes to be detected (framed, insulated, drywall_installed) = 3

The Mask R-CNN model was trained using the “Detectron2” library with the “PyTorch” framework. The technical algorithm representing the workflow of optimizing the Mask R-CNN model is shown in Fig. 7.

During the training process of this deep learning model, for training and validation, choosing a 90%–10% split in the dataset was a strategic decision informed by the common practices in this field of research which is predominantly based on the complexity of the image dataset. On one hand, the 90% training set aligns with established practices in machine learning, where maximizing the amount of training data can improve model performance. By using a larger training set, the model is exposed to a more comprehensive range of scenarios, which enhances its ability to learn robust representations of the input space (Goodfellow et al., 2016). This comprehensive exposure is particularly crucial for complex datasets, as it helps the model to generalize better to unseen data (Bishop, 2006). Additionally, a larger training set helps to mitigate the risk of overfitting, especially when combined with regularization techniques (Srivastava et al., 2014). On the other hand, the 10% validation set is commonly used as it provides a sufficient, yet reliable estimate of the model’s performance. It allows for effective hyperparameter tuning and model selection by providing a separate dataset to evaluate the model’s ability to generalize (Chollet, 2017). The validation set thus serves as a crucial component in ensuring that the model’s performance is robust. Therefore, the decision to use a 90%–10% split is supported by theoretical foundations in machine learning literature, highlighting the benefits of larger training sets for complex datasets and the importance of a validation set for reliable model evaluation (Bishop, 2006; Goodfellow et al., 2016; Srivastava et al., 2014).

Once the optimized Mask R-CNN model was trained, the progress images from the third construction site were used to test whether this deep learning model is capable of automatically recognizing the as-built states of framed, insulated and drywall installed. The testing of indoor images using the optimized Mask R-CNN model is demonstrated in Fig. 8.

Mask R-CNN is capable of extracting the as-built area of the walls in pixel level masks in the indoor site images as red for framing, green for insulation and blue for the drywall. The optimized Mask R-CNN model was post-processed to calculate the area of the pixel masks corresponding to each as-built state detected. The work-in-progress completion of an indoor wall can be calculated using the ratio between the as-

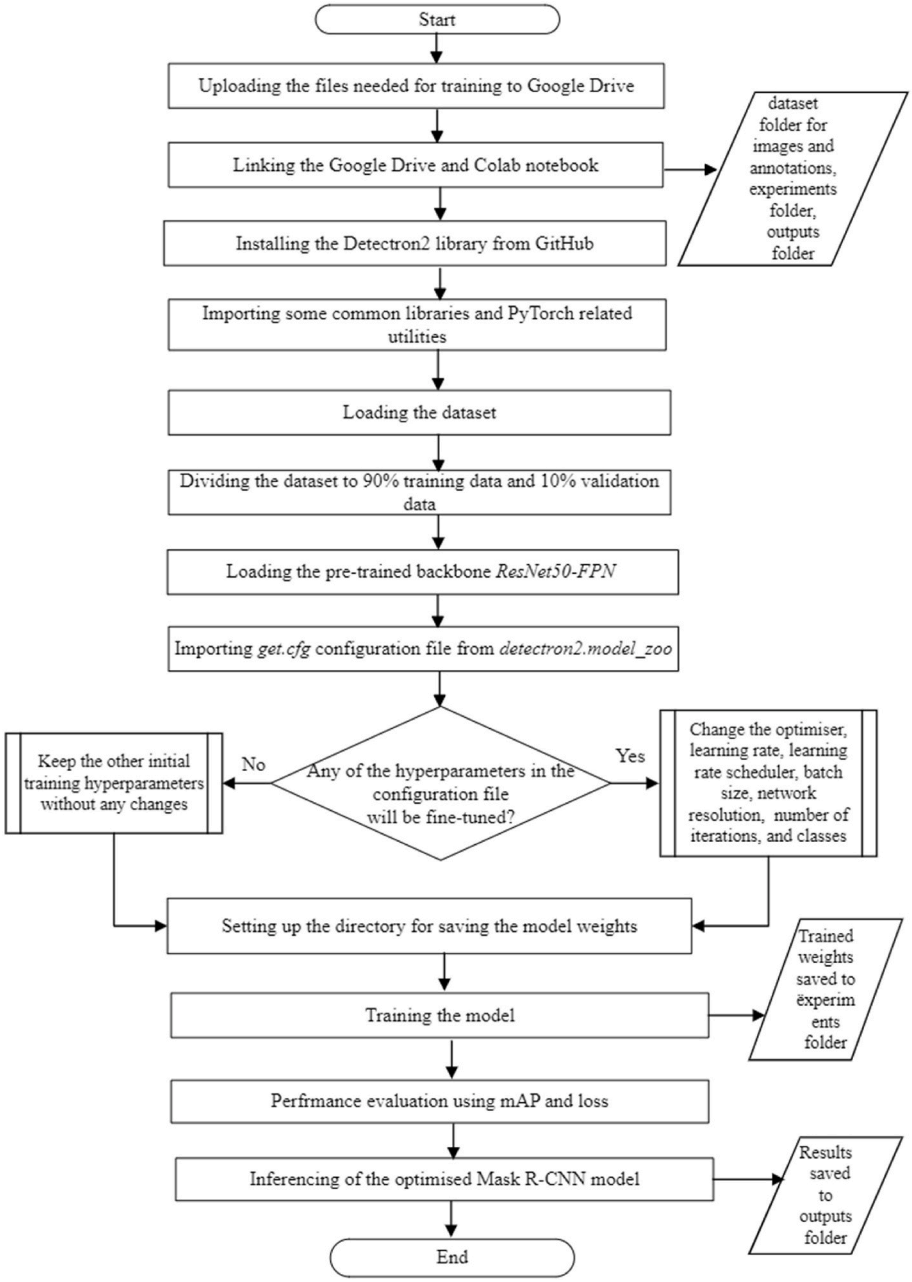

Fig. 7. The technical algorithm.

built area of the state detected (framing, insulation, and drywall installation) on a given day with the physical wall area obtained from 3D construction models. Extraction of the area of the pixel masks is essential to map the relationship between the pixel areas of the detected elements' as-built state and the as-built area of the wall completed on a given date. The underlying idea is to present the indoor site images with their detected as-built state and their estimated work completion ratio. This relationship is mapped as shown in Equation (1) and Equation (2).

$$\frac{Pd}{Pw} = \frac{Sd}{Sw} \quad \text{Equation 1}$$

$$Sd = \frac{Pd}{Pw} * Sw \quad \text{Equation 2}$$

Where,

Pd = pixel mask area of the detected as-built state on a given date d_i (in pixels)
Pw = pixel mask area of the entire wall detected on a given date d_i (in pixels)
Sd = as-built area of the wall completed on a given date d_i (in m^2)
Sw = physical wall area obtained from the construction models (in m^2)

It can be discerned from Equation (1) that the pixel ratio ($\frac{Pd}{Pw}$) equals the work-in-progress completion ratio $\left(\frac{Sd}{S\,w}\right)$. To calculate the as-built area completion of drywall installed (Pd) in pixels, the area of the segmented mask of the drywall (D) is obtained in pixels. Hence, Pd = D. To calculate the pixel mask area of the entire wall, the area of the segmented mask of the drywall (D) and the area of the segmented mask of the insulation (I) are calculated. Therefore, Pw = D + I. This

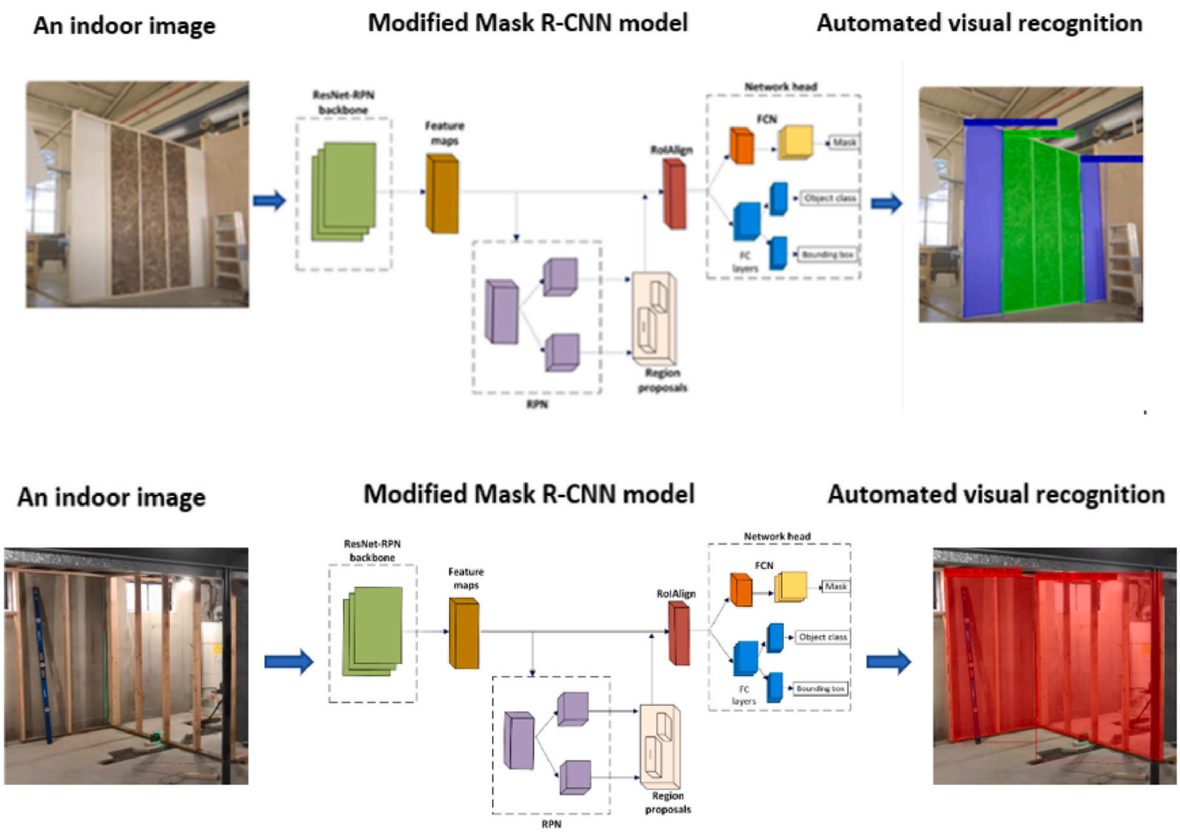

Fig. 8. Examples of indoor as-built state recognition using Mask R-CNN.

relationship can be expressed in Equation (3) for further clarification.

$$\frac{\textit{Area of the as}-\textit{built region (drywall)}}{\textit{Area of the entire wall}}=\frac{Pd}{Pw}=\frac{D}{D+I} \qquad \text{Equation 3}$$

As per Equation (2) and Equation (3), the actual as-built area of the wall (Sd) on a given date can be calculated by multiplying the pixel ratio by the as-planned physical wall area (Sw). This data can be captured from the 3D construction drawings and models obtained from the building contractor. Equation (4) can be used for the physical area calculation.

$$\text{Physical area of the as}-\text{built region of the wall}=\frac{Pd}{Pw} * \text{Sw}(\text{m}^2) \qquad \text{Equation 4}$$

Based on the equations and the explanations above, the wall completion percentage as a ratio of the area of drywall installed to the area of the entire wall is shown in Fig. 9.

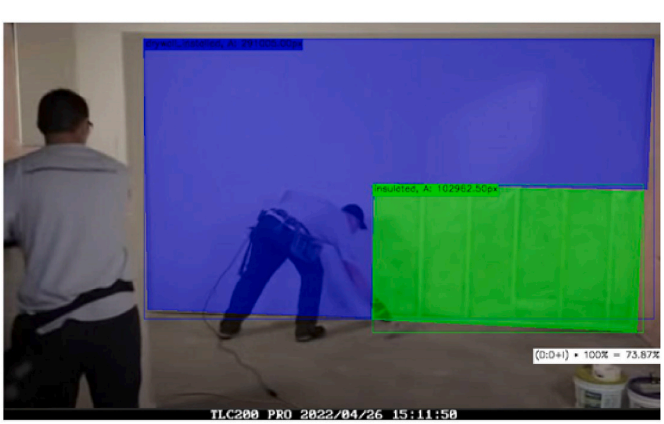

Fig. 9. Work completion percentage.

As per the image, the wall completion percentage as a ratio of the area of drywall installed to the area of the entire wall is **73.87%** on April 26, 2022. When deployed on Streamlit, project managers see a similar output image like the one in Fig. 9.

4.2. Performance evaluation of the proposed model

The base model (SGD + MultiStepLR + 0.02) achieved a mean average precision (mAP) of 86.72%. Further optimization with a learning rate of 0.00025 and the Adam optimizer (Adam + CosineLR + 0.00025) improved the mAP to 88.02%. This indicates that the optimized Mask R-CNN model effectively learns and recognizes features despite occlusions.

The training and validation losses were evaluated to determine the model's object recognition accuracy. The base model (SGD + MultiStepLR + 0.02) showed training and validation losses of 0.21. The model with SGD + CosineLR +0.00025 had a training loss of 0.21 and a validation loss of 0.22. The best results were achieved with Adam + CosineLR +0.00025, showing a training loss of 0.19 and a validation loss of 0.20, along with the highest mAP, indicating it as the optimum Mask R-CNN model for the dataset. The comparison of these learning rates on mAP and loss are demonstrated in Fig. 10.

4.3. Deploying details of the mask R-CNN model

To deploy the deep learning model, initially the corresponding GitHub repository was created uploading the site images and the underlying Python code. Currently, this GitHub repository is not publicly available to be deployed to the Streamlit cloud. Therefore, the user interface of the deployed deep learning model was set up on the local machine. In terms of the user interface, the initial user interface, when someone opens the application for the first time is shown in Fig. 11a.

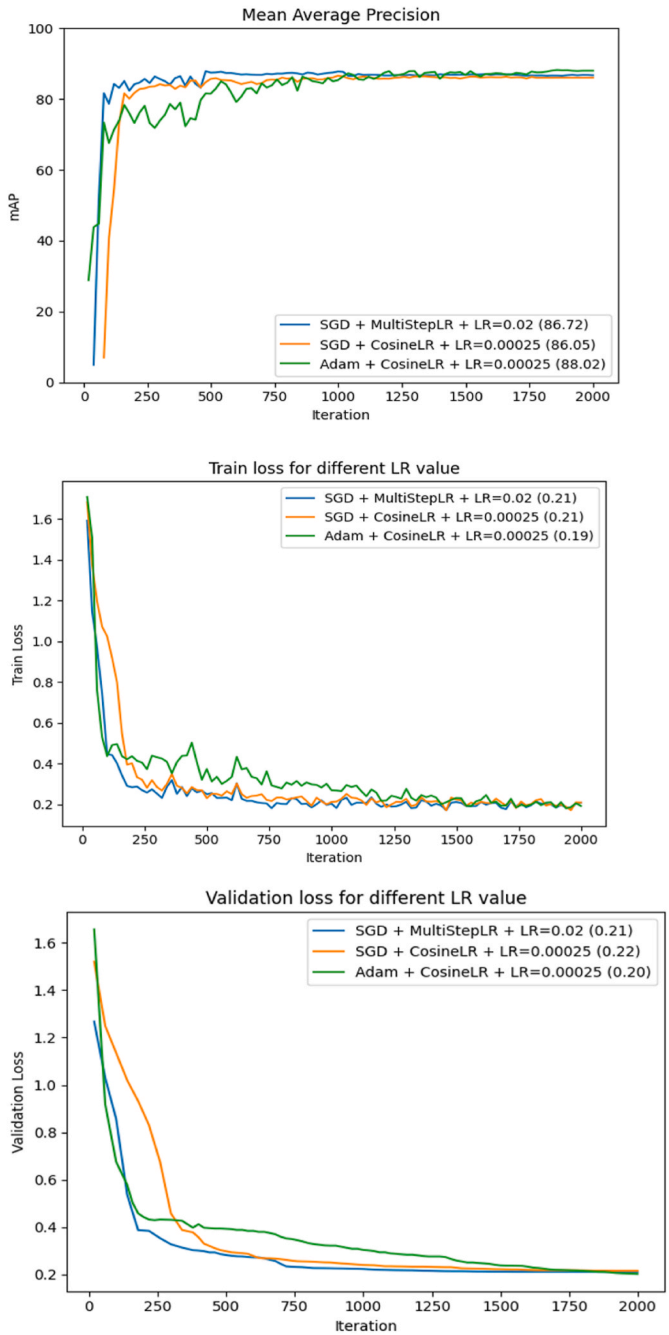

Fig. 10. Comparison of mAP and loss with different learning rates.

The final user interface after the uploading of a site image and segmentation is completed is shown in Fig. 11b. The results are generated with the as-built state of the construction element and the work completion percentage. The application can upload the test image by the front-end code. The image is sent back to the server, and the model can make an inference and send back the inferred result to the front-end. Both the input and the output are in image format. For this model, JPEG images are preferred.

5. Discussion

The study presents an optimized Mask R-CNN deep learning model for recognizing the as-built states and calculating the work-in-progress of indoor construction elements. For the proof-of-concept purposes, the current study focuses on the progress monitoring scenario of

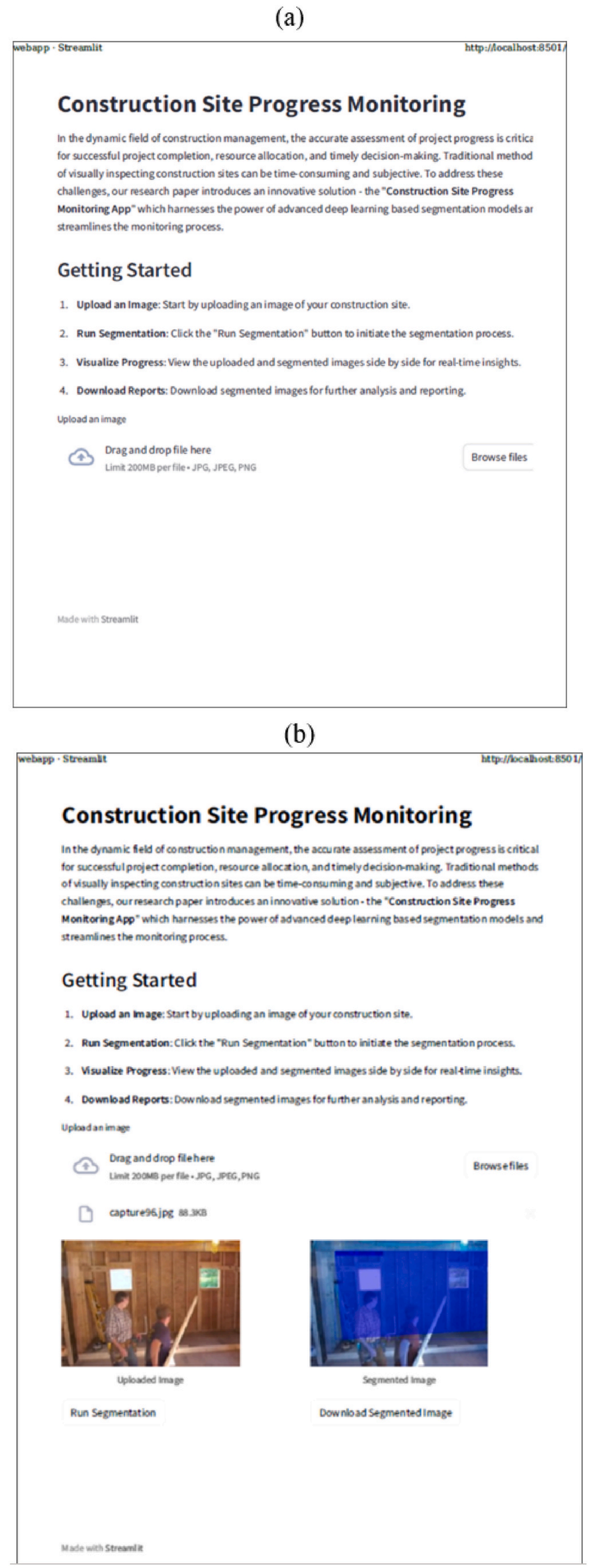

Fig. 11. a) Initial interface; b) Final interface.

framing, insulation, and installation of drywalls of indoor walls to train the Mask R-CNN model. This study also provides an understanding on the deployment process of a trained deep learning model, such as Mask R-CNN, on the Streamlit platform.

The proposed method for indoor construction progress monitoring leverages computer vision and deep learning to address the specific challenges identified in traditional and state-of-the-art monitoring approaches. Traditional progress monitoring relies heavily on manual data collection, which is prone to subjectivity and errors (Wei et al., 2022; Golparvar-Fard et al., 2015). The proposed method automates the data collection process using time-lapse cameras to capture a comprehensive set of images, significantly reducing human intervention and the associated biases. Indoor environments pose specific challenges for computer vision, including occlusions, variable lighting, and clutter (Ekanayake et al., 2021b; Hamledari et al., 2017). The proposed method employs advanced preprocessing techniques such as perspective transformation, low illumination enhancement and photometric and geometric data augmentation, to enhance the robustness of the deep learning model against these issues. By using multiple cameras positioned to capture different lighting conditions and angles, a comprehensive coverage and accurate feature extraction was ensured.

On the training image dataset, the Mask R-CNN model (He et al., 2017) was optimized to a mAP of 88.02%. through hyperparameters such as learning rate (0.00025) and CosineLR scheduler with Adam optimizer. This indicates that the optimized Mask R-CNN model effectively learns and recognizes features despite occlusions. In a previous indoor construction progress monitoring study, Wei et al. (2022) explored the segmentation capabilities of deep learning for indoor work-in-progress calculation. However, they did not extend their methodology to demonstrate the calculation of the area of the segmented masks derived from the model. The present study addresses this gap. Wei et al. (2022), concentrated on brick laying and plastering with a model trained on a limited dataset of 500 images captured from the same site for both training and testing the model. Their approach increases susceptibility to overfitting. Comparatively, the present study has used a different progress monitoring scenario, which is the framing, insulation and drywall installation captured from two different case projects to train the Mask R-CNN model, which is then to be tested on a third case project. Training deep learning models requires significant computational resources (O'Mahony et al., 2019; Wang et al., 2021). For example, Wei et al. (2022) trained their Mask R-CNN model on a physical computer. The current study leverages the freely available Google Colab platform to provide the necessary computational power for model training, eliminating the need for extensive on-site hardware. This approach offers a cost-effective and efficient solution for virtualization and model training (Carneiro et al., 2018; Ohkawara et al., 2021). While such a deep learning model can be easily executed on Google Colab by someone with a programming background, without the deployment, project managers, with non-programming background cannot access and employ such models (Khorasani et al., 2022). Based on the previous successful deployment of deep learning models on freely available Streamlit platform (Prapas et al., 2021; Shukla et al., 2021), the current study presents how the Mask R-CNN based progress monitoring model could be deployed to the cloud or the local machine when the code and the dataset is uploaded to a GitHub Repository.

The deployment process of the model on the Streamlit platform offers an approach to a reliable and stable environment for executing and interacting with the deep learning or machine learning models as highlighted by Kreuzberger et al. (2023). Their study on machine learning operations provides a detailed framework that includes continuous integration, continuous deployment, and model monitoring for successful deployment of machine learning or deep learning models. It must be noted that the successful deployment of machine learning or deep learning models in construction project management significantly depends on user acceptance. Several factors influence this acceptance, including perceived usefulness, ease of use, and trust in the technology (Venkatesh et al., 2003). The Technology Acceptance Model suggests that for project managers to adopt these tools, they must perceive them as enhancing their efficiency and effectiveness without requiring extensive technical expertise (Venkatesh and Bala, 2008). Integrating machine learning or deep learning models with existing project management workflows can be challenging, especially in industries like construction that have long relied on traditional methods (Hartmann et al., 2012). Ensuring compatibility and offering customizable solutions that can be tailored to specific project requirements can also enhance integration (Chauhan et al., 2020).

5.1. Theoretical contributions and practical implications

The research presented in this study makes several theoretical contributions to the field of project management and construction progress monitoring. By integrating an optimized Mask R-CNN model deployed on the Streamlit platform, this study advances the theoretical understanding of digital tools, specifically deep learning and computer vision and how such advanced algorithms can be translated into practical applications for the benefit of project management. The use of advanced deep learning techniques such as Mask R-CNN for recognizing as-built states and calculating work-in-progress in indoor construction environments extends and advances existing literature on computer vision applications in indoor construction (Hamledari et al., 2017; Kropp et al., 2018; Deng et al., 2020; Wei et al., 2022; Wong et al., 2024). The methodology adopted in this study demonstrates how the principles of the DSR framework are applied throughout the study. By iterating through the relevance, rigor, design, and change and impact cycles (Drechsler and Hevner, 2016), the research contributes to the theoretical discourse on developing a technology-based artefact for addressing a construction project management related problem.

This study also provides an understanding on the deployment process of a deep learning model for the construction project managers aiming to improve efficiency and accuracy in progress monitoring. By demonstrating the automation potential of the indoor construction progress monitoring process with the Mask R-CNN model, that can be deployed on the Streamlit platform, this study offers a promising tool that can significantly reduce the reliance on manual data collection and subjective assessments. This type of automation underscores the importance of real-time data, enabling project managers to make informed decisions and minimize human error in progress assessments, which are vital for identifying and mitigating delays and cost overruns (Golparvar-Fard et al., 2015; Zhang et al., 2009). However, it is worth noting that this study does not claim to address all practical challenges in construction project progress monitoring comprehensively.

6. Conclusions and future directions

This study applied computer vision and deep learning for automated visual recognition and work-in-progress calculation of as-built construction elements, utilizing an optimized Mask R-CNN model with a deployment process on the Streamlit platform for indoor construction progress monitoring. The research presented makes significant contributions to both the theoretical framework and practical applications within project management, particularly through the lens of digital innovation in construction progress monitoring. With digital tools like the one proposed in this study, project managers can shift focus to leadership and strategic responsibilities rather than labour-intensive tasks. While transitioning tasks such as progress monitoring to digital platforms offers long-term cost and time savings, it involves an initial learning curve and upfront investment. The societal and leadership challenges introduced as a result of digital transition must also be addressed. From a societal perspective, upskilling the digital literacy of the workforce is crucial. Protecting sensitive information of the project with the increased use of digital cameras for data collection, algorithms for data analysis, and cloud platforms for data storage is also of paramount importance. From a leadership perspective, project managers must be adaptive to change, ensure ethical use of digital tools, and facilitate fair access to these tools through a culture of innovation and collaboration.

Despite the contributions of this study, certain limitations need to be acknowledged for improvements in the future studies. The Mask R-CNN model was trained on data collected from two indoor sites and tested on a third site. This methodological choice ensures that the model's performance is tested in a completely new environment. However, the generalizability could be improved by testing on more indoor sites and different progress monitoring scenarios in future studies. For the continuous improvement of the Mask R-CNN model and the Streamlit-based platform, future studies should incorporate qualitative research that gathers project managers' feedback on the usability of this approach. Understanding the factors that influence successful useability through qualitative feedback can provide valuable insights for enhancing the model and platform to better meet the needs of end users such as construction project managers.

CRediT authorship contribution statement

Biyanka Ekanayake: Writing – original draft, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Johnny Kwok Wai Wong:** Writing – review & editing, Supervision, Project administration. **Alireza Ahmadian Fard Fini:** Writing – review & editing, Visualization, Supervision. **Peter Smith:** Writing – review & editing, Supervision. **Vishal Thengane:** Software, Data curation.

Declaration of competing interest

The authors have no competing interests to declare.

Data availability

The data that has been used is confidential.

Acknowledgements

The authors wish to acknowledge the funding support received through the University of Technology -School Built Environment Research Day Prize 2023.

References

Agarwal, R., Chandrasekaran, S., Sridhar, M., 2016. Imagining Construction's Digital Future. McKinsey & Company. https://www.mckinsey.com/capabilities/operations/our-insights/imagining-constructions-digital-future.

Ahmadian Fard Fini, A., Maghrebi, M., Forsythe, P.J., Waller, T.S., 2022. Using existing site surveillance cameras to automatically measure the installation speed in prefabricated timber construction. Eng. Construct. Architect. Manag. 29 (2), 573–600. https://doi.org/10.1108/ECAM-04-2020-0281.

Armeni, I., Sener, O., Zamir, A.R., Jiang, H., Brilakis, I., Fischer, M., Savarese, S., 2016. 3D semantic parsing of large-scale indoor spaces, 1534-1543. https://openaccess.thecvf.com/content_cvpr_2016/html/Armeni_3D_Semantic_Parsing_CVPR_2016_paper.html.

Bednar, P.M., Welch, C., 2020. Socio-technical perspectives on smart working: creating meaningful and sustainable systems. Inf. Syst. Front 22 (2), 281–298. https://doi.org/10.1007/s10796-019-09921-1.

Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer.

Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y.M., 2020. Yolov4: optimal speed and accuracy of object detection. arXiv. https://doi.org/10.48550/arXiv.2004.10934.

Bohn, J.S., Teizer, J., 2010. Benefits and barriers of construction project monitoring using high-resolution automated cameras. J. Construct. Eng. Manag. 136 (6), 632–640. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000164.

Bosché, F., Ahmed, M., Turkan, Y., Haas, C.T., Haas, R., 2015. The value of integrating Scan-to-BIM and Scan-vs-BIM techniques for construction monitoring using laser scanning and BIM: the case of cylindrical MEP components. Autom. ConStruct. 49, 201–213. https://doi.org/10.1016/j.autcon.2014.05.014.

Canesche, M., Bragança, L., Neto, O.P.V., Nacif, J.A., Ferreira, R., 2021. *Google Colab CAD4U: hands-on cloud Laboratories for digital design* [paper presentation]. 2021 IEEE International Symposium on Circuits and Systems (ISCAS), Daegu, South Korea. https://doi.org/10.1109/ISCAS51556.2021.

Carneiro, T., Da Nóbrega, R.V.M., Nepomuceno, T., Bian, G.-B., De Albuquerque, V.H.C., Reboucas Filho, P.P., 2018. Performance analysis of Google Colaboratory as a tool for accelerating deep learning applications. IEEE Access 6, 61677–61685. https://doi.org/10.1109/ACCESS.2018.2874767.

Chauhan, S., Agarwal, N., Kar, A.K., 2020. Addressing big data challenges in smart cities: a systematic literature review. Info 22 (2), 145–170. https://doi.org/10.1108/info-03-2020-0030.

Chilukuri, D.M., Yi, S., Seong, Y., 2022. A robust object detection system with occlusion handling for mobile devices. Comput. Intell. 38 (4), 1338–1364. https://doi.org/10.1111/coin.12401.

Chollet, F., 2017. Deep Learning with Python. Simon and Schuster.

Deng, H., Hong, H., Luo, D., Deng, Y., Su, C., 2020. Automatic indoor construction process monitoring for tiles based on BIM and computer vision. J. Construct. Eng. Manag. 146 (1) https://doi.org/10.1061/(asce)co.1943-7862.0001744.

Drechsler, A., Hevner, A., 2016. A four-cycle model of IS design science research: capturing the dynamic nature of IS artifact design. DESRIST, 2016. https://hdl.handle.net/10468/2560.

Ekanayake, B., Wong, J.K., Fini, A.a.F., Smith, P., 2021a. Computer vision-based interior construction progress monitoring: a literature review and future research directions. Autom. ConStruct. 127, 103705 https://doi.org/10.1016/j.autcon.2021.103705.

Ekanayake, B., Ahmadian Fard Fini, A., Wong, J.K.W., 2021b. Technical challenges for automated indoor construction progress monitoring. In: 44th AUBEA Conference 2021. https://doi.org/10.1109/AUBEA.2021.9654723.

Ekanayake, B., Fini, A.a.F., Wong, J.K.W., Smith, P., 2024. A deep learning-based approach to facilitate the as-built state recognition of indoor construction works. Construct. Innovat. 24 (4), 933–949. https://doi.org/10.1108/ci-05-2022-0121.

El-Omari, S., Moselhi, O., 2011. Integrating automated data acquisition technologies for progress reporting of construction projects. Autom. ConStruct. 20 (6), 699–705. https://doi.org/10.1016/j.autcon.2011.06.009.

Fang, W., Ding, L., Zhong, B., Love, P.E.D., Luo, H., 2018. Automated detection of workers and heavy equipment on construction sites: a convolutional neural network approach. Adv. Eng. Inf. 37, 139–149. https://doi.org/10.1016/j.aei.2018.05.003.

Fellows, R.F., Liu, A.M., 2015. Research Methods for Construction, fourth ed. John Wiley & Sons.

Forget, P., Lamouri, S., Pellerin, R., 2022. Enhancing the decision-making process through Industry 4.0 technologies. Sustainability 14 (1), 461. https://doi.org/10.3390/su14010461.

Google Research, 2024. Colaboratory. Retrieved from. https://research.google.com/colaboratory/faq.html.

Golparvar-Fard, M., Peña-Mora, F., Savarese, S., 2015. Automated progress monitoring using unordered daily construction photographs and IFC based Building Information Models. J. Comput. Civ. Eng. 29 (1) https://doi.org/10.1061/(asce)cp.1943-5487.0000205.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. The MIT Press.

Gregor, S., Zwikael, O., 2024. Design science research and the co-creation of project management knowledge. Int. J. Proj. Manag. 102584 https://doi.org/10.1016/j.ijproman.2024.102584.

Ham, Y., Han, K., Lin, J.J., Golparvar-Fard, M., 2016. Visual monitoring of civil infrastructure systems via camera-equipped Unmanned Aerial Vehicles (UAVs): a review of related works. Visualization in Engineering 4 (1). https://doi.org/10.1186/s40327-015-0029-z.

Hamledari, H., McCabe, B., 2016. Automated visual recognition of indoor project-related objects: challenges and solutions. In: Perdomo-Rivera, J.L., Gonzáles-Quevedo, A., López del Puerto, C., Maldonado-Fortunet, F., Molina-Bas, O.I. (Eds.), Construction Research Congress 2016: Old and New Construction Technologies Converge in Historic San Juan. American Society of Civil Engineers, pp. 2573–2582. https://doi.org/10.1061/9780784479827.256.

Hamledari, H., McCabe, B., Davari, S., 2017. Automated computer vision based detection of components of under-construction indoor partitions. Autom. ConStruct. 74, 78–94.

Han, K., Cline, D., Golparvar-Fard, M., 2015. Formalized knowledge of construction sequencing for visual monitoring of work-in-progress via incomplete point clouds and low-LoD 4D BIMs. Adv. Eng. Inf. 29 (4), 889–901. https://doi.org/10.1016/j.aei.2015.10.006.

Hartmann, T., Van Meerveld, H., Vossebeld, N., Adriaanse, A., 2012. Aligning building information model tools and construction management methods. Autom. ConStruct. 22, 605–613. https://doi.org/10.1016/j.autcon.2011.12.011.

He, D., Qiu, Y., Miao, J., Zou, Z., Li, K., Ren, C., Shen, G., 2022. Improved Mask R-CNN for obstacle detection of rail transit. Measurement 190, 110728. https://doi.org/10.1016/j.measurement.2022.110728.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. *Mask R-CNN* [paper presentation]. October 22-29. In: 2017 IEEE International Conference on Computer Vision. https://doi.org/10.1109/ICCV.2017.322. Venice, Italy.

Hevner, A.R., March, S.T., Park, J., Ram, S., 2004. Design science in information systems research. MIS Q. 75–105. https://doi.org/10.2307/25148625.

Iivari, J., Venable, J.R., 2009. Action research and design science research-Seemingly similar but decisively dissimilar. In: 17th European Conference on Information Systems (ECIS), Verona, Italy. Imgaug. (2022). *Overview Of Augmenters*. [Paper presentation]. https://imgaug.readthedocs.io/en/latest/index.html.

Kardovskyi, Y., Moon, S., 2021. Artificial intelligence quality inspection of steel bars installation by integrating Mask R-CNN and stereo vision. Autom. ConStruct. 130, 103850 https://doi.org/10.1016/j.autcon.2021.103850.

Kebao, D., Jinling, W., 2021. Adoption of advanced technologies in project management: a systematic review. Int. J. Proj. Manag. 39 (2), 89–101. https://doi.org/10.1016/j.ijproman.2020.10.001.

Khorasani, M., Abdou, M., Hernández Fernández, J., 2022. Streamlit basics. In: *Web Application Development With Streamlit*. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-8111-6_2.

Kim, J., Chi, S., Seo, J., 2018. Interaction analysis for vision based activity identification of earthmoving excavators and dump trucks. Autom. ConStruct. 87, 297–308. https://doi.org/10.1016/j.autcon.2017.12.016.

Kingma, D.P., Ba, J., 2014. Adam: a method for stochastic optimization. arXiv. https://doi.org/10.48550/arXiv.1412.6980.

Kopsida, M., Brilakis, I., Vela, P.A., 2015. *A review of automated construction progress monitoring and inspection methods* [Paper presentation]. 32nd CIB W78 Conference 2015, Eindhoven, Netherlands.

Kropp, C., Koch, C., König, M., 2018. Interior construction state recognition with 4D BIM registered image sequences. Autom. ConStruct. 86, 11–32. https://doi.org/10.1016/j.autcon.2017.10.027.

Kreuzberger, D., Kühl, N., Hirschl, S., 2023. Machine learning operations (MLOps): overview, definition, and architecture. IEEE Access. https://doi.org/10.1109/ACCESS.2023.1234567.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444. https://doi.org/10.1038/nature14539.

Li, Y., Han, Z., Zhang, Q., Li, Z., Tan, H., 2020a. Automating cloud deployment for deep learning inference of real-time online services. IEEE INFOCOM 2020 - IEEE Conference on Computer Communications 1668–1677. https://doi.org/10.1109/INFOCOM41043.2020.9155267.

Li, Y., Cao, H., Zhao, X., Yu, H., 2020b. Cloud computing and big data technologies in construction management: a review. Autom. ConStruct. 110, 103003 https://doi.org/10.1016/j.autcon.2019.103003.

Li, Y., Lu, Y., Chen, J., 2021. A deep learning approach for real-time rebar counting on the construction site based on YOLOv3 detector. Autom. ConStruct. 124, 103602 https://doi.org/10.1016/j.autcon.2021.103602.

Loshchilov, I., Hutter, F., 2016. SGDR: stochastic gradient descent with warm restarts. In: *International Conference on Learning Representations* [Paper Presentation]. ICLR 2016. San Juan, Puerto Rico.

Moragane, H., Perera, B.A.K.S., Palihakkara, A.D., Ekanayake, B., 2024. Application of computer vision for construction progress monitoring: a qualitative investigation. Construct. Innovat. 24 (2), 446–469. https://doi.org/10.1108/CI-05-2022-0130.

Navon, R., 2007. Research in automated measurement of project performance indicators. Autom. ConStruct. 16 (2), 176–188. https://doi.org/10.1016/j.autcon.2006.03.003.

O'Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G.V., Krpalkova, L., Riordan, D., Walsh, J., 2019. Deep learning vs. traditional computer vision. In: Arai, K., Kapoor, S. (Eds.), Advances in Computer Vision: Proceedings of 2019 Computer Vision Conference. Springer, pp. 28–144. https://doi.org/10.1007/978-3-030-17795-9.

Ohkawara, M., Saito, H., Fujishiro, I., 2021. Experiencing GPU path tracing in online courses. Graphics and Visual Computing 4, 200022. https://doi.org/10.1016/j.gvc.2021.200022.

Omar, H., Mahdjoubi, L., Kheder, G., 2018. Towards an automated photogrammetry based approach for monitoring and controlling construction site activities. Comput. Ind. 98, 172–182. https://doi.org/10.1016/j.compind.2018.03.012.

OpenCV, 2022. Geometric Image Transformations. OpenCV. https://docs.opencv.org/4.x/da/d6e/tutorial_py_geometric_transformations.html.

Pal, A., Lin, J. J., & Hsieh, S. (2022). A framework for automated daily construction progress monitoring leveraging unordered site photographs. *Computing in Civil Engineering 2021*. https://doi.org/10.1061/9780784483893.067.

Papadonikolaki, E., Galera-Zarco, C., 2024. Data analytics in managing projects. In: The Handbook of Project Management. Routledge, pp. 187–202.

Prapas, I., Derakhshan, B., Mahdiraji, A.R., Markl, V., 2021. Continuous training and deployment of deep learning models. Datenbank Spektrum 21, 203–212. https://doi.org/10.1007/s13222-021-00386-8.

Qureshi, A.H., Alaloul, W.S., Manzoor, B., Musarat, M.A., Saad, S., Ammad, S., 2020. Implications of machine learning integrated technologies for construction progress detection under industry 4.0 (IR 4.0). In: 2020 Second International Sustainability and Resilience Conference: Technology and Innovation in Building Designs, 51154. https://doi.org/10.1109/IEEECONF51154.2020.9319974, 1-6.

Rao, A.S., Radanovic, M., Liu, Y., Hu, S., Fang, Y., Khoshelham, K., Palaniswami, M., Ngo, T., 2022. Real-time monitoring of construction sites: sensors, methods, and applications. Autom. ConStruct. 136, 104099 https://doi.org/10.1016/j.autcon.2021.104099.

Raymond, L., Bergeron, F., 2008. Project management information systems: an empirical study of their impact on project managers and project success. Int. J. Proj. Manag. 26 (2), 213–220. https://doi.org/10.1016/j.ijproman.2007.06.002.

Reconstruct, 2024. Remote Quality Control & Progress Monitoring. Reconstruct Inc. Retrieved March 1, 2024, from. https://reconstructinc.com/.

Reja, V.K., Varghese, K., Ha, Q.P., 2022. Computer vision-based construction progress monitoring. Autom. ConStruct. 138, 104245 https://doi.org/10.1016/j.autcon.2022.104245.

Rice, L., Wong, E., Kolter, Z., 2020. Overfitting in adversarially robust deep learning. In: Daumé III, H., Singh, A. (Eds.), 37th International Conference on Machine Learning. PMLR, pp. 8093–8104.

Sacks, R., Brilakis, I., Pikas, E., Xie, H.S., Girolami, M., 2020. Construction with digital twin information systems. Data-Centric Engineering 1, e14. https://doi.org/10.1017/dce.2020.16.

Shi, Z., Zhu, M.m., Guo, B., Zhao, M., Zhang, C., 2018. Nighttime low illumination image enhancement with single image using bright/dark channel prior. EURASIP Journal on Image and Video Processing 2018 (1), 13. https://doi.org/10.1186/s13640-018-0251-4.

Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. Journal of Big Data 6 (1), 1–48. https://doi.org/10.1186/s40537-019-0197-0.

Shukla, S., Maheshwari, A., Johri, P., 2021. Comparative analysis of ML algorithms & stream lit web application [Paper presentation]. In: 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N). Greater Noida, India, pp. 175–180.

Soman, R.K., Raphael, B., Varghese, K., 2017. A System Identification Methodology to monitor construction activities using structural responses. Autom. ConStruct. 75, 79–90. https://doi.org/10.1016/j.autcon.2016.12.006.

Smith, L.N., 2018. A disciplined approach to neural network hyper-parameters: Part 1–learning rate, batch size, momentum, and weight decay. arXiv. https://doi.org/10.48550/arXiv.1803.09820.

Stake, R.E., 2013. Multiple Case Study Analysis. Guilford press.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15 (1), 1929–1958. https://www.jmlr.org/papers/volume15/srivastava14a/srivastava14a.pdf.

Streamlit, 2024. Community cloud. Streamlit. Retrieved March 1, 2024, from. https://streamlit.io/cloud.

Szeliski, R., 2010. Computer Vision*: Algorithms and Applications*. Springer Science & Business Media.

Valero, E., Adán, A., Cerrada, C., 2015. Evolution of RFID applications in construction: a literature review. Sensors 15 (7), 15988–16008. https://doi.org/10.3390/s150715988.

Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D., 2003. User acceptance of information technology: toward a unified view. MIS Q. 27 (3), 425–478. https://doi.org/10.2307/30036540.

Venkatesh, V., Bala, H., 2008. Technology acceptance model 3 and a research agenda on interventions. Decis. Sci. J. 39 (2), 273–315. https://doi.org/10.1111/j.1540-5915.2008.00192.x.

Wang, Z., Zhang, Q., Yang, B., Wu, T., Lei, K., Zhang, B., Fang, T., 2021. Vision based framework for automatic progress monitoring of precast walls by using surveillance videos during the construction phase. J. Comput. Civ. Eng. 35 (1) https://doi.org/10.1061/(ASCE)CP.1943-5487.0000933.

Wei, W., Lu, Y., Zhong, T., Li, P., Liu, B., 2022. Integrated vision based automated progress monitoring of indoor construction using mask region based convolutional neural networks and BIM. Autom. ConStruct. 140, 104327 https://doi.org/10.1016/j.autcon.2022.104327.

Wong, J.K.W., Bameri, F., Ahmadian Fard Fini, A., Maghrebi, M., 2023. Tracking indoor construction progress by deep-learning-based analysis of site surveillance video. Construct. Innovat. https://doi.org/10.1108/CI-10-2022-0275.

Wong, J.K.W., Maghrebi, M., Ahmadian Fard Fini, A., Alizadeh Golestani, M.A., Ahmadnia, M., Er, M., 2024. Development of a refined illumination and reflectance approach for optimal construction site interior image enhancement. Construct. Innovat. 24 (2), 470–491. https://doi.org/10.1108/CI-03-2023-0056.

Xi, D., Qin, Y., Wang, Y., 2020. Vision measurement of gear pitting under different scenes by deep Mask R-CNN. Sensors 20 (15), 4298. https://doi.org/10.3390/s20154298.

Yang, J., Park, M.-W., Vela, P.A., Golparvar-Fard, M., 2015. Construction performance monitoring via still images, time-lapse photos, and video streams: Now, tomorrow, and the future. Adv. Eng. Inf. 29 (2), 211–224. https://doi.org/10.1016/j.aei.2015.01.011.

Ying, H., Lee, S., 2019. A Mask R-CNN based approach to automatically construct as-is IFC BIM objects from digital images. In: Al-Hussein, M. (Ed.), 36th International Symposium on Automation and Robotics in Construction (ISARC 2019), The International Association for Automation and Robotics in Construction, pp. 764–771. https://doi.org/10.22260/ISARC2019/0103.

Zhang, X., Bakis, N., Lukins, T.C., Ibrahim, Y.M., Wu, S., Kagioglou, M., Aouad, G., Kaka, A.P., Trucco, E., 2009. Automating progress measurement of construction projects. Autom. ConStruct. 18 (3), 294–301. https://doi.org/10.1016/j.autcon.2008.09.004.

Zhao, X., Jin, Y., Selvaraj, N.M., Ilyas, M., Cheah, C.C., 2023. Platform-independent visual installation progress monitoring for construction automation. Autom. ConStruct. 154, 104996 https://doi.org/10.1016/j.autcon.2023.104996.

Zhu, L., Huang, S., Schindler, K., Armeni, I., 2023. Living scenes: multi-object relocalization and reconstruction in changing 3D environments. arXiv:2312.09138. https://doi.org/10.48550/arXiv.2312.09138 arXiv.