Integration of Deep Learning and Extended Reality Technologies in Construction Engineering and Management: A Mixed Review Method

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Construction Innovation: Information, Process, Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>CI-04-2022-0075.R1</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Research Article</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Deep Learning, Extended Reality, Virtual Reality, Augmented Reality, Construction Engineering, Digital Transformation</td>
</tr>
</tbody>
</table>
Integration of Deep Learning and Extended Reality Technologies in Construction Engineering and Management: A Mixed Review Method

Abstract

Purpose
In recent years, deep learning and extended reality (XR) technologies have gained popularity in the built environment, especially in construction engineering and management. A significant amount of research efforts has been thus dedicated to the automation of construction-related activities and the visualization of the construction process. The primary aim of this study is to investigate potential research opportunities in the integration of deep learning and XR technologies in construction engineering and management.

Design/methodology/approach
This study presents a literature review of 164 research articles published in Scopus from 2006 to 2021, based on strict data acquisition criteria. A mixed review method, consisting of a scientometric analysis and systematic review, is conducted in this study to identify research gaps and propose future research directions.

Findings
The proposed research directions can be categorized into four areas, including: 1) Realism of training simulations; 2) Integration of visual and audio-based classification; 3) Automated hazard detection in head-mounted displays (HMDs); and 4) Context-awareness in head-mounted displays (HMDs).

Originality/value
This study contributes to the body of knowledge by identifying the necessity of integrating deep learning and XR technologies in facilitating the construction engineering and management process.

Keywords: Deep Learning; Extended Reality; Virtual Reality; Augmented Reality; Construction Engineering; Digital Transformation; Autonomous Construction

1. Introduction
In recent years, deep learning and XR technologies have been gaining popularity in the architectural, engineering, and construction (AEC) industry due to their excellence in automating and visualizing the construction process. The advancement of the technologies mentioned earlier enabled the development of various practical applications in the AEC industry, such as automated progress monitoring and inspection tasks, increased degree of visualization in building planning, virtual construction safety-related training, operation of...
heavy machinery through virtual simulations, and so on. The reliance on deep learning and XR technologies stems from various problems faced in the industry, including high labour practical training costs (Su et al., 2013; Tichon and Diver, 2012), construction design issues resulting from poor construction planning (Boton et al., 2013; Ivson et al., 2018), high accident and fatality rate due to inadequate safety management (Jiang et al., 2015; Man et al., 2017), low productivity in construction projects due to the lack of an effective monitoring system (Alwi et al., 2002; Roberts and Golparvar-Fard, 2019), high construction defects due to the lack of an effective building quality control system (Gordon et al., 2007; Hollis and Bright, 1999), and construction delays due to extensive resources required for progress monitoring (Navon and Sacks, 2007).

However, despite the implementation of deep learning and XR technologies, some challenges have yet to be tackled, hindering the practical applications of these technologies in the AEC industry. Some challenges include the flexibility limitations of using deep learning technology on-site as it often relies on visual input from stationary cameras installed on-site or other imaging techniques such as light detection and ranging (LiDAR) (Shirowzhan et al., 2019). In an enclosed or confined area, the flexibility of this technique is limited due to the occlusion problem. Meanwhile, the use of deep learning technology in minimizing construction safety risks is inadequate as most implementations are not developed to notify construction workers of safety hazards in real-time (Alwasel et al., 2017; Ding et al., 2018). Besides that, the approach of utilizing HMDs to warn users of safety hazards often relies solely on real-time data without any predicting or forecasting capabilities (Teizer et al., 2013). HMDs also heavily rely on manual input from humans to perform tasks, which is often not intuitive and time-consuming. Several review papers were conducted to study the applications of deep learning and XR technologies in the AEC industry to propose future research directions (Darko et al., 2020; Fang et al., 2020; Li et al., 2018; Wang et al., 2018) as a standalone technology. To date, there is a lack of research on the integration of both technologies to complement each other's limitations, which will be thoroughly explored in this review paper. This study aims to study the technological limitations of deep learning and XR technologies in an attempt to identify persisting issues that hinder their adoption in construction engineering and management, critically highlighting potential solutions towards the identified technological limitations through the integration of both technologies. A scientometric analysis and systematic review consisting of 164 carefully selected research articles from the past 15 years, 2006 to 2021, is conducted to guide the authors in achieving this aim. Based on the findings, several research directions that can potentially solve the technological limitations of deep learning and XR in construction engineering and management are then identified.

2. Research Questions

The authors hypothesized that integrating deep learning and XR technologies may provide solutions to various underlying technological limitations of each technology that limit its adoption in construction engineering and management. To be precise, the poor flexibility of on-site information retrieval process for classification and segmentation tasks, as well as the significant amount of manual work required to deploy a highly functional XR-HMD, the on-site information retrieval process for classification and segmentation tasks and the significant amount of manual work required to deploy a highly functional XR-HMD are hypothesized to be the main reasons for the limited adoption of these advanced technologies. As such, the following research questions were formulated to gain a better understanding of the application
of deep learning and XR technologies and explore potential research directions in this area to validate the hypothesis:

1. What are the applications of deep learning and extended reality technologies in construction engineering?
2. What are the technological limitations of deep learning and extended reality technologies in construction engineering?
3. What are the research opportunities in integrating deep learning and extended reality technologies in construction engineering?

The research methodology was carefully designed and elaborated based on the formulated research questions in the next section.

3. Research Methodology

This study critically reviewed existing literature and highlighted the technological limitations of deep learning and XR technologies in construction engineering and management. First, the scope of this study is heavily focused on the technological perspectives of these technologies. Although other factors such as workers’ perspectives and cost of deployment may impact the adoption of these technologies, this study aimed to study the potential integration of deep learning and XR technologies, whereby the aforementioned factors mentioned above are irrelevant in achieving the research aims. Second, this study excluded literature without direct relevance to on-site construction activities, such as building design, facilities management (FM), off-site manufacturing, and smart factories. Third, this study solely focused on deep learning techniques, excluding any other artificial intelligence (AI) techniques that do not fall under this subset of AI. Fourth, the type of technology that falls under the category of XR has changed dramatically as technology progresses. For instance, computer-aided design (CAD) models displayed on computers would fall under the category of virtual reality (VR) decades ago. In this study, the term XR refers to the virtualization or augmentation of virtual objects solely in HMDs. Last, this research topic often coincides with other topics of similar nature; hence the literature database searching process may yield irrelevant studies. Therefore, a set of exclusion criteria is defined by the author to provide a more concise and relevant literature review on this topic, which is as follows:

- Workers’ perspective on technological adoption
- Economic evaluations
- Building design and maintenance
- Facilities management
- Off-site manufacturing and smart factories
- Traditional machine learning
- Non-immersive extended reality technologies
- Construction robotics
- Environment and energy performance

As suggested by (Van Eck and Waltman, 2010; Moher et al., 2009; Nalimov and Mul’chenko, 1971), a set of data collection criteria were employed to guarantee the quality of data for this scientometric analysis. The criteria are as follows:

- Contemporary: The research articles selected were published within the past 15 years, between 2006 to 2021.
2. Relevance: The title, abstract, and keywords were manually reviewed to ensure their relevance to this research.

3. Quality assurance: Only research publications from leading journals were considered.

4. Inclusiveness: This review paper included conference papers to ensure that no critical research works were left out.

The authors selected Scopus as the literature database for this research, using the following search query: {
"Deep Learning" OR "Convolutional Neural Network" OR "CNN" OR "Recurrent Neural Network" OR "RNN" OR "Virtual Reality" OR "Augmented Reality" OR "Mixed Reality" OR "Extended Reality"} AND {
"Construction" OR "AEC" OR "Building" OR "Progress Monitoring" OR "Construction Safety" OR "Construction Management" OR "Project Management"}, as shown below in Figure 1. The conditional query was conducted within the articles’ title, abstract, and keywords.

Figure 1: Methodology for identifying relevant research literature.
The initial search in the database yielded 1007 results. After removing duplicating results, the search query yielded 991 articles, which were then filtered through a three-phase process. In the first phase, the title, abstract, and keyword of the articles were manually examined, resulting in a total of 641 filtered articles through the set of exclusion criteria defined previously by the authors. A complete paper-based screening process was conducted to further remove irrelevant articles, resulting in 187 articles. In the third phase, each manuscript was critically reviewed, removing any literature that did not directly contribute to the body of knowledge of this study, resulting in 164 articles. The selected papers formed the basis of this scientometric analysis.

4. Scientometric Analysis
This study selected 164 articles from selected sources for its analysis, as tabulated in Table 1. The majority of the published articles can be found among the top journals related to the AEC industry, including Automation in Construction, Advanced Engineering Informatics, and the Journal of Computing in Civil Engineering. These journals account for approximately 71% of the publications used in this study. Publications from several influential conferences were also considered in this study to ensure that no recent works are left out. This includes, including the Proceedings of the International Symposium on Automation and Robotics in Construction, the International Conference on Computing in Civil and Building Engineering, and the International Conference on Construction Applications of Virtual Reality. These conference proceedings account for approximately 12% of the identified publications.

Table 1: The list of sources and the number of articles from each journal used in this study.

<table>
<thead>
<tr>
<th>Source</th>
<th>No. of selected articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automation in Construction</td>
<td>71</td>
</tr>
<tr>
<td>Advanced Engineering Informatics</td>
<td>23</td>
</tr>
<tr>
<td>Journal of Computing in Civil Engineering</td>
<td>23</td>
</tr>
<tr>
<td>Journal of Information Technology in Construction</td>
<td>12</td>
</tr>
<tr>
<td>Proceedings of the International Symposium on Automation and Robotics in Construction</td>
<td>10</td>
</tr>
<tr>
<td>Journal of Construction Engineering and Management</td>
<td>6</td>
</tr>
<tr>
<td>Construction Innovation</td>
<td>5</td>
</tr>
<tr>
<td>International Conference on Computing in Civil and Building Engineering</td>
<td>5</td>
</tr>
<tr>
<td>International Conference on Construction Applications of Virtual Reality</td>
<td>5</td>
</tr>
<tr>
<td>Canadian Journal of Civil Engineering</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>164</td>
</tr>
</tbody>
</table>

The distribution of the selected articles for this study is shown in Figure 2. From 2006 to 2016, the number of yearly publications related to the use of XR technologies in construction management remained consistent at approximately three. With the introduction of Oculus Rift and High Tech Computer (HTC) Vive to the consumer market in 2016 (Greg, 2015; Hayden, 2021), the number of yearly publications increased in the following years. As the XR market started gaining popularity, the development of XR technologies rapidly grew, and several well-
known HMDs were released. This includes the release of Oculus Quest by Meta, formerly known as Facebook, and Valve Index by Valve in 2019 (Oculus VR, 2018; Stefan, 2019). As a result, the number of yearly publications averaged around 14 for the past three years. Although popular deep learning networks were developed much earlier, such as ResNet in 2015 (He et al., 2016), research in this area started gaining popularity in 2018. From 2018 to 2021, the number of yearly publications related to the use of deep learning in construction management remained high at approximately 20. The publication growth associated with the use of deep learning on in construction management in these past few years indicates a high probability of future research possibilities, with the integration of deep learning and XR potentially being one of the likely research areas.

![Figure 2: Number of published studies in deep learning and XR technologies in construction engineering and management.](image)

The bibliographic information of the filtered articles is obtained from Scopus and used as an input in VOSviewer to generate a bibliographic coupling network, as shown in Figure 3. The information obtained is accurate as of December 2021 and may vary as time passes. The bibliographic coupling is used to assess the skewness of research conducted and understand the link between the contribution of knowledge in each area (Youtie et al., 2013). A few criteria are applied to generate a more refined bibliographic coupling network, including documents that are most significantly connected to each other and must contain more than 30 citations. A total of 40 research articles fulfilled both criteria. From Figure 3, it is seen that the clusters are concentrated on each end, indicating that deep learning and XR-related research are not often related together.
The bibliographic coupling network is categorized into five clusters, whereby each cluster is represented by a unique colour, as shown in Figure 3. Table 2 shows the distribution of citations and total link strength of the research articles across the clusters, highlighting the top three most cited research in each cluster. A research theme can be generalized for each identified cluster used in the bibliographic coupling network as follows: 1) Cluster 1: Augmented Reality; 2) Cluster 2: Construction Safety; 3) Cluster 3: Sewer Defects; 4) Cluster 4: Crack Detection, and 5) Cluster 5: Virtual Reality. The articles are automatically categorized into the clusters during the bibliographic coupling network analysis in VOSviewer, and a research theme is assigned to each cluster manually by the authors after identifying the literature present within each cluster.

Table 2: Research articles in identified clusters for bibliographic coupling network

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Author and publication year</th>
<th>Article title</th>
<th>Number of citations</th>
<th>Total link strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(Wang et al., 2013)</td>
<td>A conceptual framework for integrating building information modeling with augmented reality</td>
<td>146</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>(Wang et al., 2014)</td>
<td>Integrating augmented reality with building information modeling: onsite construction process controlling for liquefied natural gas industry</td>
<td>145</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>(Koch et al., 2014)</td>
<td>Natural markers for augmented reality-based indoor navigation and facility maintenance</td>
<td>95</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>(Ding et al., 2018)</td>
<td>A deep hybrid learning model to detect unsafe behavior: integrating convolutional neural networks and long short-term memory</td>
<td>173</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>(Fang, Li, Luo, Ding, Luo, Rose, et al., 2018)</td>
<td>Detecting non-hardhat-use by a deep learning method from far-field surveillance videos</td>
<td>167</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>(Fang, Ding, et al., 2018)</td>
<td>Automated detection of workers and heavy equipment on construction sites: a convolutional neural network approach</td>
<td>114</td>
<td>52</td>
</tr>
<tr>
<td>3</td>
<td>(Cheng and Wang, 2018)</td>
<td>Automated detection of sewer pipe defects in closed-circuit television images using deep learning techniques</td>
<td>101</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>(Kumar et al., 2018)</td>
<td>Automated defect classification in sewer closed circuit television inspections using deep convolutional neural networks</td>
<td>91</td>
<td>41</td>
</tr>
</tbody>
</table>
In this study, author keywords are used to generate a co-occurrence network to illustrate the research topics and the interrelationship between the research interests (Lee and Su, 2010) that illustrates the research topics and their interrelationships, as shown in Figure 4. A thesaurus file was generated to merge similar keywords, such as "convolution neural network", “convolutional neural network”, “convolutional neural network (cnn)”, and “convolutional neural networks (cnns)”; as well as “virtual reality”, “virtual reality (vr)”, “vr”, and so on. A threshold of three keyword occurrences is also applied to refine the co-occurrence network further. The resulting network can be classified into six clusters, consisting of 32 keywords in total, and the distribution of the keywords can be found in Table 3. Similarly, the keywords were categorized into clusters automatically during the generation of the keyword co-occurrence network. A research theme was then assigned manually to each cluster after studying the relevance of the keywords in each cluster.
Figure 4: Author keyword co-occurrence network.

A unique colour represents each cluster in the author keyword co-occurrence network. Meanwhile, the size of the nodes represents the frequency of the keyword occurrence. The top five keywords that appear in the literature include “deep learning”, “augmented reality”, “virtual reality”, “building information modelling”, and “convolutional neural network”. Each cluster can be categorized into different research themes, as shown below.

Table 3: Top keywords in deep learning and XR literature.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Keywords</th>
<th>Links</th>
<th>Total link strength</th>
<th>Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Computer vision</td>
<td>9</td>
<td>20</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Construction safety</td>
<td>8</td>
<td>21</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Construction site</td>
<td>5</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Convolutional neural network</td>
<td>18</td>
<td>52</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>Deep learning</td>
<td>22</td>
<td>88</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>Long short-term memory</td>
<td>3</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Object detection</td>
<td>7</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Transfer learning</td>
<td>5</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>Augmented reality</td>
<td>17</td>
<td>34</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Building information modelling</td>
<td>17</td>
<td>37</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Cloud computing</td>
<td>4</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Cognitive load</td>
<td>2</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Construction</td>
<td>8</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Industrialized construction</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Virtual reality</td>
<td>13</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>3</td>
<td>3D modelling</td>
<td>9</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Automation</td>
<td>10</td>
<td>22</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Construction management</td>
<td>10</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>22</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------------------------</td>
<td>----</td>
<td>----</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Image processing</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Inspection</td>
<td>7</td>
<td>7</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Machine learning</td>
<td>10</td>
<td>12</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Progress monitoring</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Concrete defect</td>
<td>7</td>
<td>11</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Crack detection</td>
<td>13</td>
<td>21</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Defect detection</td>
<td>6</td>
<td>9</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Semantic segmentation</td>
<td>13</td>
<td>21</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Bridge inspection</td>
<td>5</td>
<td>13</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Ground penetrating radar</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Hololens</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Mixed reality</td>
<td>5</td>
<td>16</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Closed circuit television</td>
<td>6</td>
<td>10</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Condition assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Cluster 1: Construction safety**

In this research cluster, there are similar research efforts in both the use of using both deep learning and XR technologies in construction safety, as shown in Figure 4. However, the research focus in both applications differs from each other. Deep learning is heavily utilized to improve site safety by monitoring the behaviour and pose of construction workers, as well as the action of heavy machinery. Meanwhile, XR technologies are generally used in construction safety training to improve the safety awareness of construction workers.

- **Cluster 2: Virtual construction**

Among the top five keywords co-occurrence, three of them are located in this cluster, including “augmented reality”, “virtual reality”, and “building information modelling”. Virtualization techniques and building information modelling (BIM) are popularly utilised to provide a higher degree of visualization of three-dimensional (3D) models for construction planning and decision-making tasks. The integration of BIM and XR is commonly found in this cluster to improve collaboration among stakeholders in construction projects.

- **Cluster 3: Progress monitoring and inspection**

Most research attempts to facilitate progress monitoring and inspection tasks through automation techniques in this research cluster. Reconstruction of job sites through spatial mapping techniques and the monitoring of construction workers and heavy machinery through various tracking methods can be found in this cluster. Besides that, there is an emerging trend of identifying constructed objects on-site for progress monitoring purposes through using deep learning techniques. Some publications are also related to the superimposition of 3D models on-site in this research cluster.

- **Cluster 4: Construction defects**

In this research cluster, there is significant attention towards the automation of crack detection in concrete and masonry structures through deep learning techniques in this research cluster. Research efforts in this area cover a wide range array of structures, including buildings, bridges,
Besides, most research studies in this cluster focus on quantifying the degree of defects, such as spalling, efflorescence, and rusts.

- **Cluster 5: Non-destructive tests**

In addition to quantifying construction defects, the automated interpretation of non-destructive tests also received considerable attention from researchers. This research cluster mainly discusses the use of ground-penetrating radar (GPR) data and deep learning techniques in detecting reinforcement damages within structural components.

- **Cluster 6: Sewer condition assessment**

In this research cluster, it mainly discusses on the assessment of sewer conditions for maintenance purposes. Deep learning techniques are heavily applied in this research area to interpret video images obtained from closed-circuit television (CCTV) for sewer inspection. Most research in this area focuses on the automated detection of sewer damages such as deposit settlements, cracks, joint offset, and so on.

### 5. Systematic Review

Based on the classification of research themes from bibliographic coupling and keyword co-occurrence analysis, the review is classified into four different topics to discuss the findings systematically. The four topics include: 1) Construction safety management; 2) Progress monitoring; 3) Inspection; 4) Construction productivity. Figure 5 illustrates the relationship between the research clusters and the categorized topics of discussion. Each identified cluster falls under one or multiple topics discussed in-depth in other sections.

![Figure 5: The relationships between the research clusters and topics of discussion.](image-url)
The identified research articles are then categorized based on the topic of discussion and the technological application in each respective area, as shown in Table 4. Most deep learning-related research is heavily focused on inspection tasks, accounting for 44% of the total identified research directly related to the application of deep learning. Meanwhile, construction productivity is the major research area in applying XR technologies in construction projects, accounting for 46% of this study's total identified XR-related research. There are similar research efforts in both major research areas. Overall, each the distribution of research articles on the topic of inspection and construction productivity accounting accounted for 30% and 33% of the identified articles, respectively.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Deep learning-related research</th>
<th>Extended reality-related research</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construction safety management</td>
<td>20 (26%)</td>
<td>14 (16%)</td>
<td>34 (21%)</td>
</tr>
<tr>
<td>Progress monitoring</td>
<td>9 (12%)</td>
<td>17 (20%)</td>
<td>26 (16%)</td>
</tr>
<tr>
<td>Inspection</td>
<td>34 (44%)</td>
<td>16 (18%)</td>
<td>50 (30%)</td>
</tr>
<tr>
<td>Construction productivity</td>
<td>14 (18%)</td>
<td>40 (46%)</td>
<td>54 (33%)</td>
</tr>
</tbody>
</table>

5.1 Construction safety management

Current methods of maintaining safety in construction sites are still inadequate, as it remains one of the most hazardous workplaces, contributing to high fatality rates among different occupations (Le et al., 2014). In construction safety management, workers’ safety is improved through three different approaches. Workers’ safety is improved through three different approaches in construction safety management. The first approach is to provide a safer working environment by minimizing construction hazards exposed to the workers. The second approach ensures workers practice safe working behaviours and educate them on countermeasures to construction hazards. Lastly, the third approach provides real-time safety information on-site in real-time to alert construction workers of incoming construction hazards. The following subsections discuss different construction safety management methods developed based on one or more of the aforementioned approaches.

5.1.1 Worker, machinery, and worker-machinery interaction

Unsafe working behaviour may result in non-fatal injuries or fatalities in severe cases. A study showed that approximately 88% of on-site accidents are related to the unsafe working behaviour of construction workers (Heinrich, 1980). Countermeasures have been deployed to prevent such cases, but traditional monitoring methods are often conducted through manual observations, which are inefficient and inaccurate (Straker et al., 2010). As deep learning started gaining popularity in the AEC industry, numerous deep learning-related research studies were conducted to automatically monitor the actions of construction workers and machinery and the interaction between them. Early works utilized machine learning methods, such as support vector machine (SVM), to estimate worker pose, proving that artificial intelligence-AI has the potential in identifying unsafe working behaviour (Zhang et al., 2018). Then, uni-directional long short-term memory (LSTM) and bi-LSTM started gaining popularity in unsafe behaviour detection as prediction accuracy greatly improved by analyzing sequential data from video inputs (Kim and Cho, 2020; Kong et al., 2021; Yang et al., 2020). Besides that, some researchers proposed a hybrid deep learning model that consists...
of a convolutional neural network (CNN) and LSTM to automate this process (Ding et al., 2018). Several network structure variations of CNNs, such as cascaded pyramid network and stacked hourglass network, were also utilized in unsafe behaviour detection to address existing keypoint issues (Luo, Wang, et al., 2020; Yu et al., 2019). Alternatively, some research used CNN and recurrent neural networks (RNN) such as multitask cascaded convolutional networks (MTCNN) (Fang, Li, Luo, Ding, Rose, et al., 2018) and attention spatial-temporal pooling (ASTPN) (Wei, Love, et al., 2019) to identify workers’ identity. The tracking of machinery movements is also essential in reducing construction hazards as construction workers and machinery share an overlapping spatial location, which may lead to collisions. Hence, some researchers utilized RNN, gated recurrent to forecast equipment to address this issue (Luo et al., 2021). Some researchers also made further advancements by simultaneously predicting worker trajectory and equipment location to minimize unsafe worker-machinery interaction (Luo, Liu, et al., 2020; Wang, Wong, et al., 2019).

5.1.2 Personal protective equipment (PPE)

Brain injuries are among the most common cause of fatalities in the construction industry, accounting for roughly 24% of the total construction-related fatalities in the United States (Colantonio et al., 2009; Konda et al., 2016). Hence, the use of PPE is mandatory in construction job sites to minimize the risk of fatalities. Deep learning techniques are prevalent in detecting hardhats enforcing the proper use of PPE (Chen et al., 2021). Different CNNs and regions with convolutional networks (R-CNNs) architecture are deployed to automatically detect the use of hardhats, such as VGG16 and You Only Look Once (YOLO) v3 (Fang, Li, Luo, Ding, Luo, Rose, et al., 2018; Nath et al., 2020; Wu et al., 2019). Some researchers also expanded research in this area by covering a wider variety of PPEs, such as harness and anchorage, to improve the usability of this technology (Fang, Li, Luo, Ding, Luo and Li, 2018).

5.1.3 Site safety

Proper project planning during the design phase greatly influences the safe deliverance of construction projects (Dawood et al., 2005; Huang et al., 2007). BIM technology plays a pivotal role in the spatial-temporal analysis of construction workspace. Several XR-related research incorporated the use of BIM to optimize site safety planning, generating simulations under various scenarios to minimize construction hazards (Getuli et al., 2020; Motamedi et al., 2017; Regina and Xinming, 2021). However, construction hazards cannot be entirely avoided in this process due to the dynamic nature of construction sites. Hence, monitoring site safety during construction phases is needed, accounting for on-site continuous resource changes. The use of XR technologies, such as augmented reality (AR), and mixed reality (MR), are indispensable in delivering safety-related information to construction workers. Several researchers incorporated AR/MR into construction sites to improve workers’ awareness of construction hazards through alerts and prompts (Kim et al., 2017). Deep learning networks such as YOLOv4 were also deployed in conjunction with AR/MR devices, tracking machinery movements and alerting construction workers of incoming hazards (Sabeti et al., 2021). Some researchers also utilized XR technology to assist in excavation tasks, minimizing the risk of striking underground utility lines during machinery operation (Talmaki et al., 2013). Besides that, some image processing techniques were also deployed to monitor site safety by inspecting the proper installation of safety countermeasures such as safety guardrails (Kolar et al., 2018).
5.1.4 Safety training

Construction safety training has been extensively studied and has proven effective in improving construction workers' safety knowledge and performance (Burke et al., 2006, 2011; Wilkins, 2011). However, conventional training simulations lack realism, impacting the performance outcomes of the training. As virtual reality (VR) technology became more accessible, researchers developed various VR-based simulations as an alternative to conventional solutions to tackle this issue. Some researchers developed simulations to educate proper safety guardrail installation (Harichandran et al., 2021). Several hazard identification and management training were also designed to improve the safety knowledge of construction workers (Eiris et al., 2020; Getuli et al., 2021; Jeelani et al., 2020; Jeon and Cai, 2021). Personalized training simulations have also recently gained popularity to enhance the usability of VR-based training (Jacobsen et al., 2021; Solberg et al., 2020).

5.2 Progress monitoring

A proper progress monitoring system is crucial to ensure that construction projects are delivered on time, within construction budgets, and meet building standards. Most budget overruns and construction errors are directly associated with poor monitoring systems, as construction projects are highly susceptible to human errors (Akinci, Boukamp, et al., 2006; Arditi and Gunaydin, 1997; Nahangi and Haas, 2014). Over the years, automated progress monitoring systems have become more common and have attracted researchers' attention (Akinci, Kiziltas, et al., 2006; Bhatla et al., 2012; Fathi and Brilakis, 2013). Deep learning and XR technologies have received significant attention in this area due to their excellence in visualizing and automating progress monitoring tasks.

The primary responsibility of tracking construction activities' progress is to quantify the as-built changes over time and compare them against as-planned schedules (Alsafouri and Ayer, 2018). Before 3D spatial mapping became accessible, numerous research was conducted to track construction progress through two-dimensional (2D) images such as Scale Invariant Feature Transform (SIFT) and Speeded Up Robust Feature (SURF) method (Bay et al., 2006; Golparvar-Fard et al., 2009). As range imaging and laser scanning techniques became more popular, there was a need for automated analysis of 3D point cloud data (Zhang and Arditi, 2013; Zhu and Brilakis, 2009). Several neural network architectures were utilized to segment and classify either point cloud or depth data obtained on-site, such as SegNet and FuseNet (Lei, Zhou, et al., 2019; Pour Rahimian et al., 2020; Yuhan and Yong, 2020). Single Shot MultiBox Detector (SSD) method is also gaining popularity due to its capabilities in classifying multiple objects in a single forward pass (Liu et al., 2021; Xueliang et al., 2020). Generally, deep learning networks are used to identify structural objects and site information, often used in conjunction with BIM to extract and update information for progress monitoring purposes (Nath et al., 2019; Yeritza et al., 2021). Some innovative solutions were also proposed in recent years, such as using audio data to identify machinery actions for monitoring purposes (Xie et al., 2019). Besides that, the potential of AR/MR technology in tracking construction progress has been extensively studied since the 2000s (H. and R., 2007; Shin et al., 2008). As AR/MR technology advances, it started to play a significant role in aiding building professionals in tracking construction progress (Behzadan and Kamat, 2011). Some research focused on the usability of AR/MR technology on-site, addressing any user interface, information display, or data flow inefficiencies (Akbar et al., 2021; Kai-Chen et al., 2012; Zaher et al., 2018). Meanwhile, some research focused on the superimposition issues of BIM models on-site such
as occlusion problems (Marianna and Ioannis, 2020). The advancement of 3D spatial mapping and tracking technology also facilitate the process of virtual reconstruction of construction sites enabled reconstructing construction sites as virtual models (Tezel and Aziz, 2017). The application of scene reconstruction is mainly seen in generating simulations for off-site monitoring of construction projects through VR technology (Rafael et al., 2015; Xie et al., 2011; Zhang and Pan, 2021). With the aid of computer vision techniques, building professionals can even navigate on-site seamlessly in an augmented environment while performing progress monitoring tasks (Ali, Lee, et al., 2020; Tayeh et al., 2021; Yang et al., 2013a).

5.3 Inspection
Modern building designs are becoming increasingly complex, raising the difficulties of inspection tasks due to the multi-disciplinary nature of construction projects (Chan et al., 2014; Ismail et al., 2016). As such, the reliance on conventional 2D methods such as paper-based construction drawings in inspection tasks is becoming insufficient (Dadi et al., 2014). Nevertheless, a building inspection is crucial in ensuring the structural integrity of the building and minimizing the risks of budget overruns (Feng et al., 2016; Little et al., 2018). The following subsections elaborate on how deep learning and XR technologies aids building professionals in identifying building defects, offset, and discrepancies.

5.3.1 Defect detection
In recent years, deep learning techniques received significant attention and have been extensively applied in detecting construction defects in bridges (Deng et al., 2020; Evan et al., 2020; Hüthwohl et al., 2019), buildings (Rony et al., 2020), pavements (Allen et al., 2018; Dhakal et al., 2021; Nhat-Duc et al., 2018; Tanzim et al., 2021), dams (Ren et al., 2021), tunnels (Li et al., 2021), foundations (Zhang, Li, et al., 2020), and sewerage system (Cheng and Wang, 2018). A huge proportion of deep learning-related research efforts were made to identify surface cracks in various structural materials such as masonry (Wang, Zhao, et al., 2019), concrete (Yiqing et al., 2020), and steel (Dung et al., 2019). Different CNNs and R-CNNs were utilized to analyze still images for crack detection, such as VGG16, InceptionV3, ResNet, ZFNet, and DenseNet (Beckman et al., 2019; Dung and Anh, 2019; Kouzehgar et al., 2019; Mei et al., 2020). Aside from crack defects, other defects were also considered, such as spalling, bughole, corrosion, and so on (Atiqur et al., 2021; Chow et al., 2020; Huynh et al., 2019; Pan et al., 2021; Wei, Yao, et al., 2019). The use of GPR data is also widely used to analyze subsurface defects in different types of structures such as bridges and sewerage systems (Asadi et al., 2020; Dinh et al., 2018; Lei, Hou, et al., 2019; Zhang, Yang, et al., 2020). Meanwhile, video input obtained from CCTV is popularly used to analyze damage in sewerage systems, such as cracks, protrusion, joint faulty, holes, and so on (Cheng and Wang, 2018; Hassan et al., 2019; Kumar et al., 2018; Li et al., 2019; Shiv et al., 2020; Yin et al., 2020). Several VR-based training simulations were also developed to educate learners on the inspection process, improving inspectors' quality in identifying defects (Albeaino et al., 2021; Beh et al., 2021; Shi, Du and Ragan, 2020).

5.3.2 Positional accuracy
One of the most essential tasks in construction monitoring is to ensure that the constructed elements correspond to the design plans. The positional accuracy of the constructed elements is critical as any deviation or error may affect the structural integrity of the building (Construction, 2000). Shin et al. identified the potential of AR and successfully developed an
AR-based inspection tools became widespread, conventional tracking methods were commonly used, such as radio-frequency identification (RFID), barcoding, quick response (QR), and so on (Koch et al., 2014; Kuo et al., 2013; Kwon et al., 2014; Suyang et al., 2013; Zhou et al., 2017). With the introduction of Microsoft HoloLens, the development of AR/MR-based inspection tools rapidly advanced as researchers realized that marker-based solutions were no longer necessary (Ali, JoonOh, et al., 2020; Loporcaro et al., 2019). Integrating BIM and AR/MR is also prominent in this area to facilitate information flow during the inspection process (Donghai et al., 2021; Kim and Olsen, 2021; Nguyen et al., 2021, 2020). Besides that, some researchers also demonstrated the benefits of deep learning techniques in inspection tasks (Saovana et al., 2020).

5.4 Construction productivity

Technological innovation is vital for enhancing on-site construction productivity and lowering overall construction costs (Greco et al., 2021). The optimization of construction activities through proper planning and management, as well as automation techniques, are proven to have a strong positive relationship with construction productivity (Ballesteros-Pérez et al., 2017; Ghodrati et al., 2018). In the subsections, the application of deep learning and XR technologies in assisting building professionals in assembly and decision-making tasks will be explored.

5.4.1 Task assistance and automation

Conventional construction assembly processes tend to rely on the interpretation of 2D paper-based drawings, which is time-consuming and prone to miscommunication (Neumann and Majoros, 1998). Studies show that misinterpretation of drawings is common, attributing to 60% of procedural errors (Lattanzio et al., 2008; Veinott et al., 1995). Hence, this is necessary to develop a solution that can provide a more natural interpretation method of assembly instructions (Wang and Dunston, 2006). Several marker-based AR solutions were developed in the early years, demonstrating the benefits of XR in improving construction productivity (Byungil et al., 2012; Chu et al., 2018; Fazel and Izadi, 2018; Hou and Wang, 2013; Lei et al., 2013, 2015; Wang et al., 2014). The successful implementation of AR technology in assembly tasks reduced assembly time required by over 80%, positively impacting construction productivity (Chalhoub and Ayer, 2019; Kwiatek et al., 2019). The release of Microsoft HoloLens also enabled the omission of physical markers (Dallasega et al., 2021), further improving the usability of this AR/MR technology in assembly tasks (Bademosi et al., 2019; Chalhoub et al., 2021; Chalhoub and Ayer, 2018; Deshpande and Kim, 2018; Qin et al., 2021; Wang et al., 2021). Besides that, offsite production time also decreased with the integration of AR/MR technology and robots, decreasing the overall project completion time (Amtsberg et al., 2021; Tavares et al., 2019; Xiang et al., 2021).

5.4.2 Productivity planning and training

Construction productivity increases as construction workers are equipped with sufficient technical knowledge (Wang et al., 2013). Hence, a training simulation is an effective method of delivering such information to construction workers. Several XR-based training simulations were developed over the years, educating construction workers on machinery operation, assembly tasks, and so on (Juang et al., 2013; Lucas, 2018; Meža et al., 2014; Paes et al., 2017; Toraño et al., 2008; Xiangyu and Dunston, 2007). Aside from that, simulations can also be used for path planning purposes, optimizing worker and machinery operation during
construction stages (Krishna Lakshmanan et al., 2020; Rashidi et al., 2021; Shi, Du and Worthy, 2020; Shiva et al., 2021; Wang et al., 2020; Wei, Justin, et al., 2019). The use of XR also enabled stakeholders to collaborate more efficiently and make well-informed decisions in productivity planning (Amin et al., 2009; Boton, 2018; Czerniawski and Leite, 2020; Du et al., 2018; Goulding et al., 2014; Jiao et al., 2013; Lee et al., 2020; Sangiorgio et al., 2021; Wang and Dunston, 2008). Some researchers also utilized CNN to capture information on-site, which can be used for simulation purposes to further improve construction productivity (Park et al., 2021). Meanwhile, various deep learning techniques are also applied to track workers (Fang, Ding, et al., 2018; Son et al., 2019) and machinery (Bo and Shih-Chung, 2021a, 2021b; Dong et al., 2022; Rashid and Louis, 2019; Roberts and Golparvar-Fard, 2019; Sim et al., 2020), used for productivity analysis purposes.

6. Discussion and future research directions

The limitations of deep learning and XR technologies in construction engineering are summarized in this section based on the scientometric analysis and systematic review. The potential future research directions are proposed in this section and are categorized into four subsections, including 1) Realism of training simulations; 2) Integration of visual and audio-based classification; 3) Automated hazard detection in head-mounted displays (HMDs), and 4) Context-awareness in head-mounted displays (HMDs).

6.1 Realism of training simulations

The learning effectiveness of training simulations is directly related to the implementation method as presented by Edgar’s pyramid of learning (Masters, 2013). As such, the realism of XR-based training simulations has been a priority in most studies as it directly affects the memory retention capabilities of learners (R. et al., 2014; Turkan et al., 2017; Zhao and Lucas, 2015). A common method to provide realism is through the utilization of actual construction site information in generating training simulations (Kamat and Martinez, 2003; R. et al., 2014; Rashidi and Rahinah, 2012). However, it is time-consuming to fully capture the dynamic environment of construction sites, and therefore, most training simulations are conducted in an idealized condition (Retik and Shapira, 1999; Wang et al., 2008). Nevertheless, deep learning methods have proven to be efficient in tracking workers (Son et al., 2019), machinery (Bo and Shih-Chung, 2021a), and structural elements (Evan et al., 2020). Therefore, future research should be directed towards using deep learning methods in capturing on-site information, which is then used to generate a more realistic XR-based training simulation. The use of deep learning methods can replace traditional on-site information retrieval methods in scene reconstruction tasks, such as the use of physical markers (Yang et al., 2013b). The scalability issues present in a marker-based method that prevents the generation of a large-scale realistic training simulation can be solved through this method.

6.2 Integration of visual and audio-based classification

The automated tracking of machinery actions using visual data has been extensively researched upon in recent years. In recent years, the automated tracking of machinery actions using visual data has been extensively researched (Bo and Shih-Chung, 2021a; Dong et al., 2022). However, occlusion issues remain a huge challenge and have yet to be fully resolved (Bo and Shih-Chung, 2021b). Recently, audio-based classification methods such as continuous wavelet transform (CWT) have been popularly used, which can be used to classify machinery actions as well (Cheng et al., 2017; Zhang, Li, et al., 2020). Although audio-based classification methods are
not subjected to occlusion issues, there is a lack of positional awareness of the audio source. With the aid of MR-HMDs, this issue would not persist as MR-HMDs such as Microsoft HoloLens comes with spatial mapping capabilities. Therefore, MR-HMDs equipped with **Red Green Blue Depth** (RGB-D) cameras and audio recorders can integrate both visual and audio-based classification methods. Hence, future research should be directed towards integrating visual-audio classifiers and MR-HMDs in tracking machinery actions. **The integration of both in MR-HMDs can omit the need for external localization devices, such as ultra-wideband (UWB) and RFID** (Van Herbruggen *et al.*, 2019; Ko, 2013), **effectively lowering the cost of deployment.** Aside from that, the applicability of visual-based neural networks in motion recognition or object detection tasks will no longer be hindered by semi or fully obstructed visual data.

### 6.3 Automated hazard detection in head-mounted displays (HMDs)

Automated hazard detection has been one of the research interests in construction safety management (Wang, Wong, *et al.*, 2019). However, most AR-based tools depend on positional data to inform workers about safety hazards, which may not be suitable for sudden hazards which require fast responses (Kim *et al.*, 2017). Introducing RNNs such as LSTM into HMDs may further reduce construction risks as workers can anticipate construction hazards before occurring. Hence, the integration of RNNs into HMDs in detecting construction hazards should be further studied in the future. **The feedback nature of RNNs enables the process of data in a temporal sequence through an internal state.** Therefore, the trajectory of workers and machinery can be predicted through visual recordings. As such, the usability of HMDs in improving workers’ safety can be expanded through the active prevention of collision between workers and machinery.

### 6.4 Context-awareness in head-mounted displays (HMDs)

Context-awareness in HMDs is essential to facilitate information flow, providing relevant and concise information to users (Bae *et al.*, 2013). Numerous AR/MR-based tools were developed to guide users on assembly tasks, yet most advanced tools can only superimpose assembly instructions without considering the relevance of information (Chalhoub *et al.*, 2021; Chalhoub and Ayer, 2018; Deshpande and Kim, 2018). The complexity of information display may be overwhelming, leading to a decrease in assembly performance. As deep learning techniques proved to help identify as-built information (Xueliang *et al.*, 2020), the relevancy of construction information can be improved by introducing context-aware assembly instructions through the same techniques. Therefore, the gain of context-awareness in HMDs through the integration of deep learning techniques should be further studied to facilitate the assembly process. **The introduction of context-awareness in HMDs can improve the usability of this technology through the reduction of its technology’s usability by reducing cognitive load among users, which will dramatically improve user experience and potentially solvedramatically improving user experience and potentially solving one of the many issues that hinders the technological adoption of HMDs on-site.** Aside from that, context-awareness in HMDs can greatly aid in the smarter real estate market, specifically in improving the user experience during virtual showcases, positively influencing the emotion and purchase intention of homebuyers (Azmi *et al.*, 2021).
7. Conclusion

The emergence of deep learning and XR technologies transformed the AEC industry, providing innovative solutions to many current construction monitoring issues. The primary aim of this study is to investigate potential research opportunities in the integration of deep learning and XR technologies to facilitate the construction engineering and management process. A mixed review method incorporating scientometric analysis and systematic review is conducted in this study to achieve this goal. A total of 164 research articles from 2006 to 2021 were critically reviewed and discussed in this paper. Bibliographic coupling and keyword co-occurrence analysis were also conducted, which forms the categorization structure of this review paper. The topic of discussion of this review paper is categorized into four main areas, namely: 1) Construction safety management; 2) Progress monitoring; 3) Inspection; and 4) Construction productivity. The discussion of topics leads to identifying research gaps in the implementation of deep learning and XR technologies in construction engineering and management. Several research directions are proposed after the scientometric analysis and systematic review, including 1) Realism of training simulations; 2) Integration of visual and audio-based classification; 3) Automated hazard detection in head-mounted displays (HMDs), and 4) Context-awareness in head-mounted displays (HMDs). The limitations of this study are summarized in two points. There are some limitations in this study, which are First, the proposed research directions are strictly directed towards focusing on integrating the integration of deep learning and XR technologies. However, there might be alternative solutions that do not require such the integration of these two techniques. Secondly, future research directions are proposed based on the literature obtained solely from Scopus and are limited to the authors’ knowledge. In future works, the study scope should be expanded to cover more literature databases; besides, the study should be shared amongst a larger group of experts for a more comprehensive review.

References


(2006), “Relative Effectiveness of Worker Safety and Health Training Methods”,
American Journal of Public Health, American Public Health Association, Vol. 96 No. 2,
p. 315–324.

Construction Equipment Operation Using Augmented Reality”, Journal of Computing in


aided point layout”, Journal of Information Technology in Construction (ITcon), Vol. 24
No. August 2018, pp. 95–111.

under-trained individuals to complete specialty construction tasks”, Journal of

development and implementation of the mandatory building inspection scheme (MBIS)
in Hong Kong”, Built Environment Project and Asset Management, Emerald Group
Publishing Limited.

Ensuring Proper Use of Personal Protective Equipment (PPE) in Construction Site”, in
Toledo Santos, E. and Scheer, S. (Eds.), International Conference on Computing in Civil
and Building Engineering, Vol. 98, Springer International Publishing, Cham, pp. 1115–
1126.

construction equipment using audio signals and support vector machines”, Automation in

circuit television images using deep learning techniques”, Automation in Construction,
Vol. 95, pp. 155–171.

defects on concrete structures with the convolutional autoencoder”, Advanced

Modelling and Augmented Reality systems: An experimental study”, Automation in


American Institute of Steel Construction.

comprehensive set of building components using deep learning”, Advanced Engineering


No. 7, pp. 1–16.


object detection in modular construction based on a comparative evaluation of deep
learning algorithms”, Construction Innovation, Emerald Publishing Limited, Vol. ahead-
HoloLens Augmented Reality Technology as a Construction Checking Tool”,
3 Lucas, J. (2018), “Immersive VR in the construction classroom to increase student
understanding of sequence, assembly, and space of wood frame construction”, Journal of
Information Technology in Construction, Department of Computer Science, Vol. 23, pp.
179–194.
surveillance to manage safety: A case study of a transport mega-project”, Advanced
construction equipment using computer vision and deep learning techniques”, Automation
Forecasting of Construction Equipment for Monitoring Construction Site Safety BT -
Proceedings of the 18th International Conference on Computing in Civil and Building
Engineering”, in Toledo Santos, E. and Scheer, S. (Eds.), International Conference on
Computing in Civil and Building Engineering, Springer International Publishing, Cham,
pp. 1127–1138.
Reality–Based Progress Monitoring”, Journal of Computing in Civil Engineering,
American Society of Civil Engineers, Vol. 34 No. 4, p. 4020016.
review”, Medical Teacher, Taylor & Francis, Vol. 35 No. 11, pp. e1584–e1593.
110, p. 103018.
items for systematic reviews and meta-analyses: the PRISMA statement”, PLoS Medicine,
Public Library of Science San Francisco, USA, Vol. 6 No. 7, p. e1000097.
analysis and optimization system using BIM-enabled virtual reality (VR) environments”,


Stefan, E. (2019), “Valve’s Index VR headset will ship this June, with preorders starting May 1st”, *The Verge*.


1 International Association for Automation and Robotics in Construction (IAARC).

detection of historic masonry buildings based on mobile deep learning”, \textit{Automation in Construction}, Vol. 103, pp. 53–66.

personalized operation training using value stream mapping”, \textit{Automation in Construction},
Vol. 119, p. 103355.

4 Wang, W., Brambley, M.R., Kim, W., Somasundaram, S. and Stevens, A.J. (2018),
“Automated point mapping for building control systems: Recent advances and future
research needs”, \textit{Automation in Construction}, Vol. 85, pp. 107–123.

AEC: An experimental prototype study”, \textit{Automation in Construction}, Vol. 15 No. 3, pp.
314–326.

for face-to-face collaborative design review”, \textit{Automation in Construction}, Elsevier, Vol.
17 No. 4, pp. 399–412.

Augmented Reality within an Architectural Firm”, \textit{Electronic Journal of Information

framework for integrating building information modeling with augmented reality”,
\textit{Automation in Construction}, Vol. 34, pp. 37–44.

Reality with Building Information Modeling: Onsite construction process controlling for

role of user-centered AR instruction in improving novice spatial cognition in a high-
precision procedural task”, \textit{Advanced Engineering Informatics}, Vol. 47, p. 101250.

for concrete surface bughole based on deep learning”, \textit{Automation in Construction}, Vol.
107, p. 102920.

construction sites with computer vision: A spatial and temporal attention pooling
network”, \textit{Advanced Engineering Informatics}, Vol. 42, p. 100981.

Assessment in Virtual and Mixed Reality Environments: Comparison of Novices and
Experts”, \textit{Journal of Construction Engineering and Management}, American Society of
Civil Engineers, Vol. 145 No. 9, p. 4019049.

10, pp. 1017–1026.


Reply to the Reviewers' Comments

The authors would like to appreciate the constructive comments given by the respected reviewers. The reply to the comments, corresponding revisions and their respected locations are listed in Tables 1 and 2.

This document presents our detailed responses to the Reviewers' Comments. We believe that the revised manuscript improved significantly in terms of content, discussions, and presentation. We are grateful for the detailed comments of the reviewers, which guided us in preparing the revised manuscript.

As guided by the editorial decision letter, this response letter consists of (a) the comments made by the Reviewers and (b) our point-to-point responses to the Reviewers. In our revision, we significantly edited the manuscript by adding newly published papers as references to support the study's practical implications, providing precise explanations of the research methodology, revising the introduction section, and ensuring all abbreviation used for the first time is explained in full words. In addition, we checked the writing style of the whole manuscript. The edited/modified sections in this revision are highlighted through the "Track Changes" feature of MS Word.

Once again, we thank the reviewers for their valuable comments and suggestions.

Thank You!

Responses to Reviewer #1:

After mentioning that the paper is interesting, Reviewer 1 listed some comments and suggestions. We tried to address all of these valuable comments in our revision. Thanks a lot for these comments. Below, the detailed responses to the comments are presented in Table 1.

Table 1. Response to the Reviewer #1.

<table>
<thead>
<tr>
<th>No.</th>
<th>Comment</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Please explain the full word for the first time in text, prior to using abbreviation.</td>
<td>All of the abbreviations used have been explained in full words now, including: - pg 3, line 20 - pg 3, line 21 - pg 3, line 24 - pg 3, line 25 - pg 5, line 25 - pg 10, line 13 - pg 13, line 21 - pg 14, line 26 - pg 18, line 4 The abbreviations used for the first time in the following sections are not explained in full words as it is either part of the literature search results or keyword filtering method: - Research methodology: the search query used for this study - Scientometric analysis: Table 2, the article title of the literature found</td>
</tr>
<tr>
<td>No.</td>
<td>Comment</td>
<td>Responses</td>
</tr>
<tr>
<td>-----</td>
<td>-------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>2.</td>
<td>In research question section, line 39-41: please specify the limitations the authors aiming to address.</td>
<td>The limitations that the authors have aimed to address are added in the research question section, on pg 2, lines 39-44.</td>
</tr>
<tr>
<td>3.</td>
<td>The methodology section needs to be expanded and elaborated on.</td>
<td>The methodology section has been revised, with added definition scope and exclusion criteria, on pg 3, lines 12-38. Figure 1 has also been further elaborated to explain the research methodology more clearly, on pg 5, lines 1-7.</td>
</tr>
<tr>
<td>4.</td>
<td>In research methodology section, please add the elimination criteria for the eliminated papers. More specific information is needed in this section to clearly describe the process of searching and selecting the required papers.</td>
<td>The elimination criteria and the research methodology process have been embedded within comment 3 to address both comments at the same time.</td>
</tr>
<tr>
<td>5.</td>
<td>Please identify the databases searched.</td>
<td>The database searched for this study is Scopus, highlighted in pg 4, line 7, and Figure 1.</td>
</tr>
<tr>
<td>6.</td>
<td>Figure 2 should be a bar chart this type of graph is not meaningful for such information.</td>
<td>Figure 2 is revised from a line graph to a bar chart.</td>
</tr>
<tr>
<td>7.</td>
<td>Practical implications of the outcome of this study can be discussed further.</td>
<td>The practical implications of this study have been further elaborated in Sections 6.1, 6.2, 6.3, and 6.4.</td>
</tr>
</tbody>
</table>

**Additional comments**

1. Research Methodology: this section must be improved prior to publication. The research methodology is improved based on the given comments 3 and 4.

<table>
<thead>
<tr>
<th>No.</th>
<th>Comment</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Introduction: You need to highlight research aim and methods in the introduction</td>
<td>The research aim and methods have been highlighted in the introduction, on pg 2, lines 24-32.</td>
</tr>
<tr>
<td>2.</td>
<td>The motivation of research should be clearly highlighted</td>
<td>The research motivation has been highlighted more clearly now, explicitly targeting the issues this study aims to solve, on pg 2, lines 39-44.</td>
</tr>
<tr>
<td>3.</td>
<td>Methodology: Try to restructure the methodology to highlight and justify</td>
<td>The research methodology has been improved, highlighting the defined scope of the study and</td>
</tr>
</tbody>
</table>

**Responses to Reviewer #2:**

Reviewer 2 lists several comments/suggestions. The authors tried to address all of those valuable comments in our revision. Thanks a lot for these comments. The detailed responses to the comments are presented in Table 2.

**Table 2. Response to the Reviewer #2.**

<table>
<thead>
<tr>
<th>No.</th>
<th>Comment</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Introduction: You need to highlight research aim and methods in the introduction</td>
<td>The research aim and methods have been highlighted in the introduction, on pg 2, lines 24-32.</td>
</tr>
<tr>
<td>2.</td>
<td>The motivation of research should be clearly highlighted</td>
<td>The research motivation has been highlighted more clearly now, explicitly targeting the issues this study aims to solve, on pg 2, lines 39-44.</td>
</tr>
<tr>
<td>3.</td>
<td>Methodology: Try to restructure the methodology to highlight and justify</td>
<td>The research methodology has been improved, highlighting the defined scope of the study and</td>
</tr>
<tr>
<td>Additional comments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>1. Research Methodology: The research methodology is robust but improvement is</td>
<td>The research methodology is improved based on comment 3.</td>
<td></td>
</tr>
<tr>
<td>required.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Implications for research, practice and/or society: should be more highlighted</td>
<td>The implications of the study have been elaborated based on comment 4.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Quality of Communication: The quality is good but figures are a bit hazy</td>
<td>Figure 2 has been changed from a line graph to a bar chart to deliver better information.</td>
<td></td>
</tr>
</tbody>
</table>

4. The practical implications should be more highlighted

The practical implications have been further highlighted in Sections 6.1, 6.2, 6.3, and 6.4.