Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 19 No. 9 (2025) | OPEN ACCESS

https://doi.org/10.3991/ijim.v19i09.55579

PAPER

Real-Time Defect Detection and Carbon Footprint Visualization in Green Construction Using Mobile Augmented Reality and Building Information Modeling

Yi Liao, Huan Luo(✉)

College of Civil Engineering, Guangxi Vocational Normal University, Nanning, China

luohuan802020@163.com

ABSTRACT

The integration of mobile augmented reality (AR) technology with building information modeling (BIM) has introduced novel solutions for construction management, particularly in real-time defect detection and carbon footprint monitoring. AR technology enables the real-time provision of three-dimensional visual information at construction sites, which, when combined with BIM, facilitates accurate defect identification and feedback. Additionally, BIM provides a scientific basis for planning carbon emission pathways in construction projects. However, existing defect detection and carbon footprint management systems face challenges such as limited accuracy and insufficient real-time capabilities. Current study on green construction primarily focuses on defect detection and carbon footprint calculations, yet most approaches continue to rely on traditional two-dimensional drawings or manual inspection, which fail to meet the real-time demands of construction sites. The absence of an integrated solution leveraging both AR and BIM technologies has constrained their practical application in construction. To address these limitations, this study proposes a real-time defect detection and carbon footprint visualization and path planning system for green construction, based on mobile AR technology and BIM. The system employs AR-based stereo matching for real-time defect identification and utilizes BIM for carbon footprint visualization path planning. This study aims to provide an efficient and accurate approach to defect detection while enhancing the environmental protection level during the construction process through effective carbon footprint management.

KEYWORDS

green construction, mobile augmented reality (AR), building information modeling (BIM), defect detection, carbon footprint, real-time monitoring

1 INTRODUCTION

With the rapid development of the construction industry, environmental protection and sustainability have become global focal points [1, 2]. During construction

Liao, Y., Luo, H. (2025). Real-Time Defect Detection and Carbon Footprint Visualization in Green Construction Using Mobile Augmented Reality and Building Information Modeling. *International Journal of Interactive Mobile Technologies (iJIM)*, 19(9), pp. 92–106. https://doi.org/10.3991/ijim.v19i09.55579

Article submitted 2025-01-12. Revision uploaded 2025-03-07. Final acceptance 2025-03-18.

processes, issues such as energy consumption, material waste, and construction defects exert significant environmental impacts [3–5]. To mitigate these challenges, the principles of green building and construction have increasingly shaped industry development [6, 7]. The adoption of advanced technological approaches for green construction management has emerged as a critical strategy for enhancing construction quality and reducing carbon emissions [8]. In recent years, the integration of mobile augmented reality (AR) technology with building information modeling (BIM) [9–11] has provided innovative solutions for the construction sector, demonstrating substantial potential, particularly in real-time monitoring and defect detection during construction processes.

Although previous studies have attempted to apply AR and BIM technologies in green construction and construction management, several limitations remain [12–15]. The defect detection method proposed by Cheng and Shen [16] relies on traditional two-dimensional drawings and manual inspection, resulting in low efficiency and a high likelihood of overlooking minor defects. Moreover, this approach often fails to provide real-time feedback during actual construction processes. While Luo and Lin [17] made efforts to improve carbon footprint calculation and management, the absence of a dynamic, real-time visualization system hindered the precision of carbon footprint monitoring and control. Urban et al. [18] focused on static data analysis but overlooked the potential of mobile AR technology for real-time feedback in construction environments. To address these shortcomings, this study aims to integrate mobile AR with BIM technology to develop a stereo matching-based real-time defect detection system for green construction and a BIM-integrated carbon footprint path planning system, thereby enhancing the efficiency of defect detection at construction sites while improving carbon emission management.

The primary research objectives are divided into two key components. The first component involves the development of a real-time defect detection system based on mobile AR stereo matching. This system leverages AR technology to enable real-time defect detection and feedback at construction sites, incorporating BIM data to provide a more precise and efficient construction quality control method. The second component focuses on BIM-based carbon footprint visualization path planning. By utilizing architectural information and construction data within BIM, a scientifically grounded carbon emission pathway was established, and visualization techniques were employed to facilitate real-time monitoring and control of carbon footprints by construction managers. Through this study, an intelligent and environmentally friendly construction management mode is expected to be established, which not only enhances construction quality but also contributes to carbon emission reduction, promoting the green and sustainable development of the construction industry.

2 REAL-TIME DEFECT DETECTION IN GREEN CONSTRUCTION BASED ON MOBILE AR STEREO MATCHING

The proposed mobile AR stereo matching algorithm for real-time defect detection in green construction consists of five stages, which are elaborated upon in the following sections.

2.1 Real-time defect feature extraction in green construction

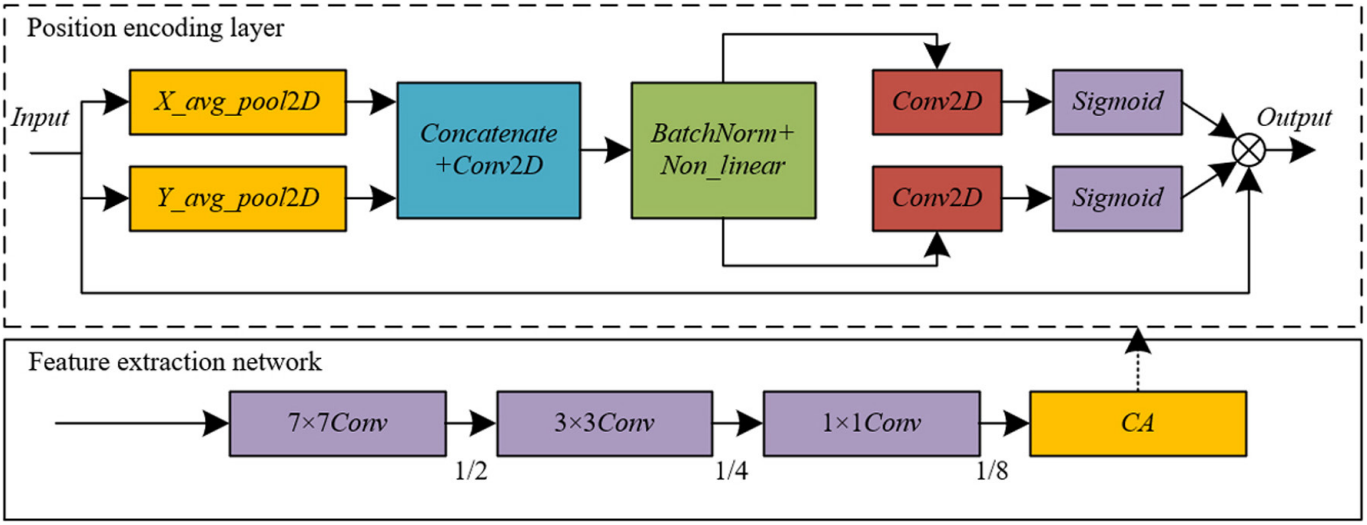

Fig. 1. Network architecture for feature extraction and structure of the position encoding layer

Due to the complex environment of construction sites, where weakly textured or texture less regions are often present, traditional feature extraction methods may fail to accurately capture defect information. In such regions, the AR camera captures images with low pixel similarity, leading to potential recognition ambiguities. To address this issue, a position encoding-based feature extraction method was employed in this study. Through a global average pooling operation, the real-time defect feature map $F_{1/8}$ for green construction was aggregated along both horizontal and vertical directions, allowing potential defect information within the region to be effectively extracted. A CA layer of the attention mechanism further combines feature maps from both directions and enhances feature representation through the 1×1 convolution and batch normalization operations, enabling the feature map to better reflect the fine-grained defect features such as cracks and misalignments. By introducing a coordinate attention mechanism, the CA layer assigns spatial position information to the feature map, significantly enhancing the recognition accuracy of defects in weakly textured regions. In this approach, the real-time defect feature map for green construction was first decomposed along the horizontal and vertical directions. Subsequently, two 1×1 convolution operations were applied to transform the features into the same channel dimension as the input. The processed features were then activated using the Sigmoid() function to generate two directional attention weights, h_g^z and h_q^z. Figure 1 visually illustrates the network architecture for feature extraction and the structure of the position encoding layer. Given an input of size $q \times g$, the feature map $D_{1/8}$ of the z-th channel is represented by $D_{1/8}^z$. The attention weights along the horizontal and vertical directions at point (u, k) are denoted as $h_g^z(u)$ and $h_q^z(k)$, respectively. The enhanced feature representation generated after position encoding is represented by $D^z{}_{1/8}'$, and the encoded features obtained through weighted summation with the input can be mathematically expressed as follows:

$$D_{\frac{1}{8}}^{a\prime} = D_{\frac{1}{8}}^{a} \times h_g^z(u) \times h_q^z(k) \tag{1}$$

2.2 Description of real-time defect features in green construction

In conventional transformer models, when the standard multi-head self-attention mechanism deals with a large number of tokens with the computational complexity of $P(j^2)$, the batch matrix multiplication and the Softmax function are used for computing the attention matrix, leading to inefficiencies when processing complex image data from construction sites. To address this issue, a separable self-attention mechanism with linear complexity was introduced in this study. By simplifying the computational process, this mechanism substantially reduces computational and memory consumption. Figure 2 illustrates the structure of the developed lightweight transformer model. Figure 3 presents the adopted multi-head self-attention layer architecture.

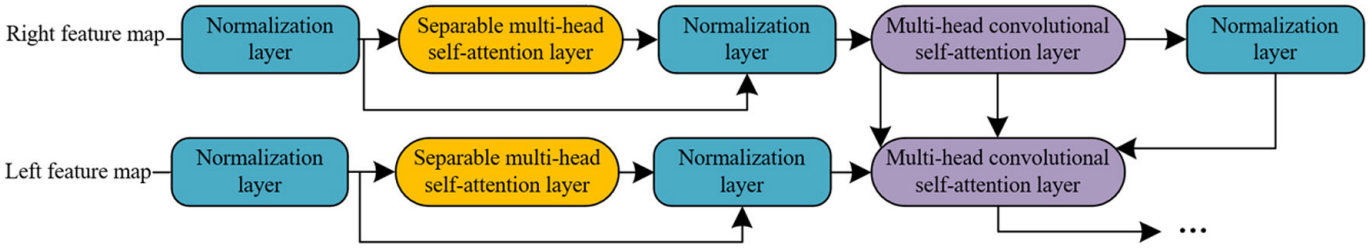

Fig. 2. Structure of the lightweight transformer model

Unlike conventional multi-head self-attention mechanisms, the proposed mechanism computes only the contextual scores of latent tokens, reducing computational complexity from $P(j^2)$ to $P(j)$. This enhancement significantly improves efficiency when processing large-scale component image data captured by AR cameras. In the separable self-attention mechanism, the input a is mapped into three branches representing inputs, keys, and values. Through a linear mapping of each token in a, latent tokens were generated, and the distance between each latent token and a was computed, producing a j-dimensional vector. The softmax function was then applied to normalize the vector, yielding the contextual scores z_t, which were subsequently used to compute the contextual vector z_n. The mathematical formulation of z_t is given by:

$$z_t = \delta(aQ_U), z_t \in E^j \tag{2}$$

The contextual vector z_n represents a vector of encoded defect feature information, effectively capturing defect characteristics such as cracks and misalignments in component images obtained by AR cameras. During the feature description process, z_n was generated using a weighted summation approach. Information from all tokens in input a is encoded into the contextual vector z_n, and the contextual information within z_n is shared across all tokens. This method enables not only the centralized processing of defect features at specific locations but also the capturing of information variations in the construction site on a global scale, further enhancing the accuracy of feature description. The mathematical expression of z_n is given by:

$$z_n = \sum_{u=1}^{j} z_t(u) a_{J(u)} \tag{3}$$

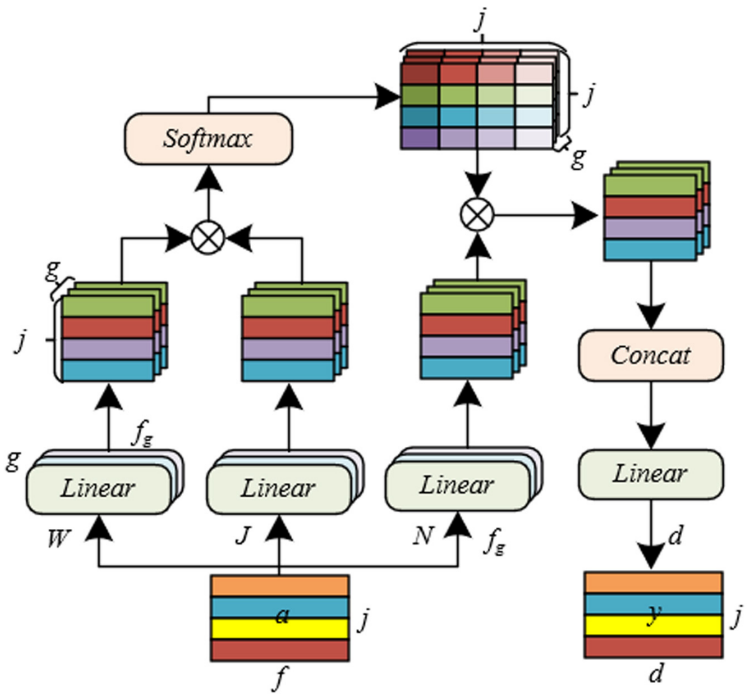

Fig. 3. Structure of the multi-head self-attention layer

To further enhance the performance of the real-time defect feature description in green construction, a value branch was incorporated in this study. This branch linearly projects the input a into an f-dimensional space and processes it using the ReLU activation function to generate a new representation, denoted as a_N. The contextual information in z_n is then subjected to element-wise multiplication with a_N to produce a more expressive real-time defect feature map for green construction. Finally, these refined contextual features are passed through a linear layer with weights denoted as Q_p to generate the final output b. This output contains real-time defect information extracted from the construction site, including features such as cracks and misalignments. Assuming that element-wise multiplication and summation operations are represented by * and Σ, respectively, the computation can be expressed as follows:

$$b = \left(\underbrace{\sum (z_t * aQ_J)}_{z_n} * \mathrm{ReLU}(aQ_N) \right) Q_P \tag{4}$$

2.3 Real-time defect feature matching for green construction

Traditional matching algorithms typically impose a uniqueness constraint on each matched point, meaning that each pixel can have only a single corresponding match. While this constraint simplifies the problem, it also introduces several limitations. First, such methods do not allow gradient propagation, making them unsuitable for learning-based stereo matching algorithms. Second, traditional approaches often lack sufficient flexibility when performing stereo matching between mobile AR and construction site environments, leading to incorrect matches in complex construction scenarios. To overcome these challenges, a dual-softmax operator was introduced in this study for matching, providing a soft constraint approach that enhances the flexibility and accuracy of defect matching.

The dual-softmax operator is a differentiable matching strategy that differs from traditional uniqueness constraints by applying soft constraints to each matching point. This allows each pixel in the component images captured by AR cameras to have multiple potential matching points. In the real-time defect feature matching process for green construction, the feature map $F_{1/8}$ was first transformed to obtain two feature maps, $\tilde{D}_{se}^{X}$ and $\tilde{D}_{se}^{Y}$. A score matrix T between these two feature maps was then computed, representing the similarity between each pair of feature points. The computation of the score matrix T serves as the foundation for the subsequent matching process. The softmax function was applied to the score matrix T, independently processing it along the horizontal and vertical directions to obtain the matching probability for each feature point. The formulation of the score matrix T is given by:

$$T(u,k) = \frac{1}{S} \cdot \left\langle \tilde{D}_{se}^{X}(u), \tilde{D}_{se}^{Y}(k) \right\rangle \tag{5}$$

To enhance the accuracy and robustness of matching, the probability of each real-time defect feature point in green construction corresponding to multiple matching points was further weighted. The computation formula for the obtained matching probability matrix O_z is given by:

$$O_z(u,k) = \text{softmax}(T(u,\cdot))_k \cdot \text{softmax}(T(\cdot,k))_i \tag{6}$$

To further improve the accuracy of stereo matching between mobile AR and construction site stereo matching, the mutual nearest neighbor (MNN) criterion was applied to filter the matching results in this study. When the confidence score of certain feature matching points falls below a predefined threshold φ_z, the MNN criterion filters these low-confidence matching points, eliminating potential mismatches. The newly obtained matching probability matrix L_z is expressed as:

$$L_z = \left\{ (\tilde{u},\tilde{k}) \mid \forall (\tilde{u},\tilde{k}) \in MNN(O_z), O_z(\tilde{u},\tilde{k}) \geq \varphi_z \right\} \tag{7}$$

2.4 Matching refinement using an adaptive window

In real-time defect detection for green construction, the accuracy of feature matching is directly related to defect detection performance, particularly for defects such as cracks and misalignments. While the initial matching step, based on the Dual-Softmax operator, provides a coarse matching probability matrix, this preliminary result often fails to meet the precision required for high-accuracy detection. In mobile AR environments, the complexity of construction sites and image resolution may cause defect features, such as cracks and misalignments, to become blurred or distorted. Consequently, a more flexible refinement approach is necessary to refine the matching and mitigate the errors introduced by traditional matching methods. To address this issue, an adaptive window-based matching refinement method was proposed in this study.

This method refines real-time defect features in green construction at multiple scales, significantly improving matching accuracy. First, each matched point pair $(\tilde{u},\tilde{k})$ from the coarse matching probability matrix was mapped onto a finer-scale real-time defect feature map for green construction. At this scale, local windows of size $q \times q$ were cropped around the matched points $\hat{u}$ and $\hat{k}$. These local windows

were then transformed to generate localized feature maps $\hat{D}_{se}^{X}(\hat{u})$ and $\hat{D}_{se}^{Y}(\hat{k})$. The correlation between each point within the local window was then computed, resulting in a local matching probability distribution. Finally, an expectation calculation was performed over all probability distributions, yielding sub-pixel-level matched points for the component images captured by AR cameras.

To further enhance matching accuracy, the concept of adaptive windows was employed in this study. Deformable convolution was utilized to learn additional offsets *fa* and *fb*, thereby adjusting the shape and position of the local windows in the component images captured by AR cameras to better accommodate different defect features. This approach allows local windows to dynamically adapt based on image content, effectively addressing the issue of matching accuracy caused by regional differences in the feature map. Based on these refined matching probability distributions, a more accurate matching matrix L_d was generated. This matrix not only improves the accuracy of defect feature matching but also provides more reliable input data for subsequent disparity regression, thereby enhancing the overall quality of disparity estimation. As a result, the real-time defect detection task can be performed with greater efficiency and accuracy. Assuming that the points on the left and right feature maps are represented as $u(a, b)$ and $k(a, b)$, the horizontal coordinates of points u and k are denoted as a_u and a_k, while their vertical coordinates are represented as b_u and b_k. After incorporating position offsets, the new expressions for points $u(a, b)$ and $k(a, b)$ are denoted as u' and k'. The number of input feature map channels is represented by z, and the fixed offsets in the horizontal and vertical directions of the current pixel are denoted as $d(f)$ and $h(f)$, respectively. The new matching probability matrix can be computed as follows:

$$O_z(u,k) = \text{softmax}(T(,\cdot))_k \cdot \text{softmax}(T(\cdot,k'))_u \tag{8}$$

$$u' = D_u^E(a_u + d(f) + fa, b_u + h(f) + fb), (a_u, b_u) \in \Re^{G\times Q} \tag{9}$$

$$k' = D_u^E(a_k + d(f) + fa, b_u + h(f) + fb), (a_k, b_k) \in \Re^{G\times Q} \tag{10}$$

2.5 Disparity regression

Following the preceding steps of feature extraction, feature matching, and matching refinement, a high-precision refined matching matrix was obtained in this study. However, the matching points in the matrix only provide the relative positional relationships between feature points. To achieve precise defect detection, these relative positional relationships must be converted into specific disparity values, which necessitates disparity regression.

First, a 3 × 3 pixel window was constructed around each matching point pair in the matching matrix. This window enables a local analysis of the pixels surrounding the matching points to capture more detailed information. The matching probabilities within the window were then re-normalized to obtain the weights for the disparity of each corresponding pixel. This process helps eliminate matching errors caused by local feature variations, thereby ensuring the accuracy of disparity estimation. Finally, a weighted computation was performed on the candidate disparities within the window to obtain the final regressed disparity value. Assuming that the

matching probability of a matching point pair is represented as s, the computation can be expressed as follows:

$$\tilde{s} = \frac{s}{\sum_{m \in N_3(j)} s} \tag{11}$$

$$DISP(s) = \sum_{m \in V_3(j)} f_s \tilde{s} \tag{12}$$

2.6 Loss function

To ensure efficient optimization of the algorithm at different stages, a multi-level loss function was proposed in this study, including coarse-level loss, refined loss, and disparity loss. First, the coarse-level loss was computed using the negative log-likelihood function on the confidence matrix O_z. This approach employs the MNN criterion to compute the ground truth label L_z^{hs} for the confidence matrix. By minimizing the negative log-likelihood loss, the model is effectively guided to learn more reliable matching relationships during the coarse matching stage. The computation is formulated as follows:

$$loss_z = -\frac{1}{\left|L_z^{hs}\right|} \sum_{(\tilde{u},\tilde{k}) \in L_z^{hs}} LN[O_z(\tilde{u},\tilde{k})] \tag{13}$$

After completing the coarse matching, the algorithm proceeds to the refined matching stage. Defect features such as cracks and misalignments often exhibit high levels of detail and complexity, necessitating more refined matching strategies. By minimizing the total variance, the model is able to capture these intricate features with greater precision, thereby enhancing detection accuracy and reliability. Assuming that point $\hat{u}$ in the feature map $\hat{D}_{se}^{X}$ undergoes a warping operation and is represented as $\hat{k}_{hs}$, and the corresponding mapped point in the feature map is denoted as $\hat{D}_{se}^{Y}$, the following can be computed in the refined matching stage as follows:

$$loss_d = \frac{1}{\left|L_d\right|} \sum_{(\hat{u},\hat{k}') \in L_d} \frac{1}{\delta^2(\hat{u})} \left\| \hat{k} - \hat{k}_{hs} \right\|_2 \tag{14}$$

$$loss'_d = \frac{1}{\left|L'_d\right|} \sum_{(\hat{u},\hat{k}') \in L'_d} \frac{1}{\delta^2(\hat{u}')} \left\| \hat{k}' - \hat{k}'_{hs} \right\|_2 \tag{15}$$

Given the ground truth disparity, the disparity loss for the final output layer is computed as follows:

$$loss_{DISP} = \sum_{u \in U^X} \left\| f_{hs}(u) - f(u) \right\| \tag{16}$$

The final loss function is formulated as the weighted sum of these four loss components:

$$loss = loss_z + loss_d + loss'_d + loss_{DISP} \tag{17}$$

This design considers multiple levels of the matching process, ranging from global coarse matching to local fine-grained matching, and ultimately to disparity regression. Therefore, the model is ensured to be effectively optimized at every stage.

3 CARBON FOOTPRINT VISUALIZATION PATH PLANNING ASSOCIATED WITH BIM

In the approach for associating BIM with a carbon emission accounting model to achieve carbon footprint visualization pathways, BIM must first be integrated as the foundational data platform. BIM provides a detailed three-dimensional model of the entire construction project along with various engineering data, including geometric information of building components, material properties, construction schedules, and associated costs. By linking BIM with the carbon emission accounting model, an accurate mapping of carbon emission data across all construction phases can be achieved. Specifically, the construction process and engineering quantity information within BIM were first utilized to determine each construction step along with the corresponding equipment, materials, and reusable materials. This information was then input into the carbon emission accounting model, where carbon emission factors related to construction machinery, building material consumption, and reusable materials were incorporated to calculate the carbon emissions of each construction step. By associating these calculated results with various construction phases in BIM, a comprehensive carbon emission database was established.

The introduction of the carbon emission accounting model serves the primary purpose of quantifying carbon emissions during the construction process accurately, thereby facilitating more efficient carbon management and green building objectives. Specifically, based on the analysis of the working procedure list and the decomposition of the construction project during the construction process, carbon emission sources were further categorized by phase and type to quantify carbon emissions at each stage.

In the carbon emission accounting model, carbon emissions from construction machinery and equipment are critical considerations. Construction equipment such as bulldozers, excavators, and slurry pumps accounts for a significant proportion of carbon emissions during the construction process. By establishing a unit procedure carbon emission quantification model based on the working procedure list analysis and using the input-output analysis method, the carbon emissions of various construction equipment were associated with factors such as engineering quantities, operating time, and equipment efficiency, enabling precise calculations of carbon emissions from different construction machinery. Assuming that the carbon emission equivalent generated by construction machinery and equipment for a unit procedure is denoted as R_m, the engineering quantity for unit procedure u is represented by L_u, the quota for machinery and equipment working hours for category u is denoted as ω_u, and the greenhouse gas emission factor associated with machinery and equipment of type u is given by RD_u, the calculation formula is expressed as:

$$R_1 = \sum_u L_u \omega_u RD_u \tag{18}$$

Construction material loss is another indirect source of carbon emission accounting. Although construction materials themselves are not a direct source of emissions from construction activities, material loss is closely related to construction quality, management efficiency, and construction techniques. To ensure that this factor is fully considered in carbon emission calculations, the carbon emissions associated with material loss were quantified in this study. Specifically, the carbon emissions from material loss were calculated based on material consumption and the loss rate during construction, thereby obtaining more accurate carbon emission values.

Assuming that the carbon emission equivalent from major construction material losses is represented as R_l, the consumption of material type u is denoted as LQ_u, the loss rate of material type u is represented as e_u, and the carbon emission factor of material type u is given by RD_u, the computation is formulated as follows:

$$R_l = \sum_u LQ_u \times e_u \times ED_u \tag{19}$$

For the quantification of carbon emissions from reusable materials, an amortization-based approach was proposed in this study. Unlike single-use construction materials, reusable materials can be reused across multiple construction projects and phases. Therefore, their carbon emissions should be allocated proportionally based on the number of usage cycles. During the accounting process, the initial carbon emissions generated when the reusable material was first used were considered, and the total lifetime usage cycles were incorporated to calculate the carbon emissions per use. These emissions were then distributed across each instance of usage. Assuming that the carbon emission equivalent of reusable materials allocated to a specific construction process is denoted as R_p, the usage quantity of material type u is represented as LP_u, the carbon emission factor of material type u is given by RD_u, and the rated usage cycles of material type u is denoted as Z_u, the computation formula is expressed as follows:

$$R_p = \sum_u LP_u \times RD_u / z_u \tag{20}$$

Once the carbon emission database was established, the visualization of carbon footprint pathways was achieved by leveraging the three-dimensional visualization capabilities of BIM. First, the timeline functionality within BIM was utilized to decompose the construction process into sequential time phases, and the carbon emissions at each time point were annotated. Color coding and heatmap visualization techniques were then employed to intuitively display carbon emission intensity across different construction stages. For instance, a gradient color scheme ranging from green to red was used to represent varying levels of carbon emissions across different construction phases. Additionally, interactive functionalities can be embedded within BIM, allowing project managers to click on specific building components or construction stages to see detailed carbon emission data and analytical results.

4 EXPERIMENTAL RESULTS AND ANALYSIS

Table 1. Results of the ablation study

Configuration				Training Set		Testing Set		Time Consumption
Lightweight Transformer Module	Coordinate Attention Mechanism	Adaptive Refinement Matching Window	Cascade Layers	Noc	All	Noc	All	
			2	1.78	1.75	1.45	1.58	0.178
√			2	1.66	1.74	1.42	1.42	0.062
√	√		2	1.74	1.65	1.36	1.44	0.064
√	√	√	2	1.62	1.69	1.24	1.44	0.068
√	√	√	3	1.48	1.62	1.26	1.48	0.077
√	√	√	4	1.49	1.65	1.28	1.46	0.098

Table 2. Comprehensive performance comparison

Algorithm	Number of Parameters ($\times 10^6$)	Floating-Point Operations ($\times 10^9$)	Memory Consumption ($\times 10^9$)	Processing Time (s)	Average Pixel Disparity Error
MobileViT	5.13	6521.87	4.12	0.51	1.12
EfficientViT	3.87	215.85	1.36	0.08	0.88
Next-ViT	6.45	568.62	3.58	0.22	0.76
LeViT	2.48	524.31	3.78	0.48	0.72
Shuffle Transformer	4.25	378.69	2.36	0.37	0.85
Proposed algorithm	2.19	224.26	1.15	0.09	0.48

The results of the ablation study, as shown in Table 1, indicate that different modules and configurations significantly affect the performance on both the training set and the testing set. In the table, "all" refers to the evaluation of all pixels in the component images captured by the AR camera, whereas "noc" refers to the testing of pixels in non-occluded regions within the AR-captured images. Experimental results demonstrate variations in performance under different configurations of the lightweight Transformer module, coordinate attention mechanism, and adaptive refinement matching window, as well as when adjusting the number of cascade layers. Specifically, across all experiments, an increase in the number of cascade layers resulted in an overall improvement in performance; however, the increased number of layers did not bring significant performance gains. For instance, with two cascade layers, the "noc" and "all" scores in the testing set were 1.45 and 1.58, respectively, while in the training set, the corresponding scores were 1.78 and 1.75, demonstrating favorable performance. When both the coordinate attention mechanism and adaptive refinement matching window were incorporated, further improvements were observed. Notably, when three cascade layers were used, the "noc" and "all" scores in the testing set improved to 1.48 and 1.62, respectively, indicating a significant enhancement in accuracy. However, at four cascade layers, while a slight improvement in "noc" and "all" scores was observed, the computational cost also increased. This suggests that although precision improves with additional cascade layers, the computational overhead rises significantly.

As shown in Table 2, the proposed algorithm demonstrates strong performance across multiple key metrics. Specifically, in terms of the number of parameters, the proposed algorithm contains 2.19 million parameters, which is significantly lower than other methods such as MobileViT (5.13 million) and Next-ViT (6.45 million), indicating a lower computational complexity. Regarding floating-point operations, the proposed algorithm requires 224.26×10^9, demonstrating a more efficient computational demand compared to MobileViT (6521.87×10^9) and EfficientViT (215.85×10^9). In terms of memory consumption, the proposed algorithm consumes only 1.15×10^9, which is relatively low and outperforms LeViT and Shuffle Transformer. For processing time, the proposed algorithm achieves the shortest processing time of 0.09 seconds, which is significantly faster than MobileViT and Next-ViT. Finally, in average pixel disparity error, the proposed algorithm achieves an error of 0.48, outperforming MobileViT, EfficientViT, and Shuffle Transformer, demonstrating superior accuracy in disparity estimation.

The summary of green construction case studies, as presented in Table 3, provides an overview of key parameters related to pile foundation construction, including pile foundation area, construction cost, number of piles, and piling method.

First, in terms of pile foundation area, Project ID 3 has the largest area, reaching 75,485 m^2, with a pile foundation cost of $3,125.24 \times 10^4$ CNY. Compared to other projects, Project ID 3 exhibits both the largest scale and the highest cost. Regarding the number of piles, Project ID 2 has the highest count, with 3,215 piles, suggesting that this project may involve a more complex construction environment. The choice of piling method also reflects different construction techniques. Static pressure piling is applied in most projects, while rail-mounted diesel piling is used in Project ID 3 and Project ID 4, which may be attributed to specific geological conditions and construction requirements.

Table 3. Summary of basic information in green construction case studies

Project ID	Pile Foundation Area (m^2)	Pile Foundation Cost ($\times 10^4$ CNY)	Number of Piles	Piling Method	Pile Diameter (mm)
1	15628	812.3512	726	Static pressure piling	511
2	42356.51	2156.1245	3215	Static pressure piling	511
3	75485	3125.2356	2548	Rail-mounted diesel piling	511
4	24526.32	234.1152	489	Rail-mounted diesel piling	511
5	2231.26	74.2568	94	Static pressure piling	511
6	1235.95	17.2653	179	Static pressure piling	511

Table 4. Summary of carbon footprint calculations for green construction case studies

Project ID	Total Carbon Emissions ($kgCO_2$)	Construction Equipment Emissions ($kgCO_2$)	Material Transportation Emissions ($kgCO_2$)	Office Area Emissions ($kgCO_2$)
1	121256.23	76524.21	14562.35	81256.32
2	412356.1245	356254.625	12658.23	21547.568
3	512458.3125	278562.32	189325.5489	28654.254
4	71256.2315	56412.32	468.521	13256.32
5	24526.25146	17568.225	3125.63	4215.36
6	18958.55215	11452.36	4356.236	2895.64

The carbon footprint sources for each project, as summarized in Table 4, can be categorized into four primary components: construction equipment emissions, material transportation emissions, office area emissions, and total carbon emissions. Among all projects, Project ID 2 exhibits the highest total carbon emissions, with construction equipment emissions accounting for the majority at 356,254.63 $kgCO_2$, indicating that the use of construction equipment is the dominant contributor to the project's carbon footprint. In contrast, material transportation and office area emissions are relatively lower, at 12,658.23 $kgCO_2$ and 21,547.57 $kgCO_2$, respectively, yet they still contribute to the overall carbon emissions. On the other hand, Project ID 5 records the lowest total carbon emissions, at 24,526.25 $kgCO_2$. In this project, construction equipment emissions amount to 17,568.23 $kgCO_2$, material transportation

emissions to 3,125.63 $kgCO_2$, and office area emissions to 4,215.36 $kgCO_2$. These values suggest that this project achieves greater carbon emission efficiency in construction equipment usage and material transportation.

5 CONCLUSION

This study primarily focuses on two key technologies: a real-time defect detection system for green construction based on mobile AR stereo matching and a BIM-based carbon footprint visualization path planning system. The core objective of these two components is to enhance quality control in green construction and carbon emission management. Through the integration of AR technology, defects occurring at construction sites can be detected and addressed in real time, ensuring that quality issues encountered during construction are resolved promptly. Additionally, by incorporating BIM, construction managers can obtain a more precise understanding of carbon emissions throughout the construction process and optimize carbon footprint pathways through visualization techniques, thereby reducing the overall carbon footprint of construction projects. The findings of this study indicate that the combination of AR and BIM technologies can significantly improve the efficiency of construction quality monitoring and provide a scientific basis for carbon emission control. This approach demonstrates strong adaptability and application potential, particularly in large-scale and complex construction projects.

Overall, this study holds significant value, particularly in its application to green construction. First, the real-time defect detection system based on mobile AR technology enables instant feedback at construction sites, thereby improving construction quality and efficiency while preventing delays and resource wastage caused by defects. Second, the carbon footprint path planning and visualization monitoring system implemented through BIM allows construction managers to monitor carbon emissions dynamically and in real time, facilitating effective carbon emission management and promoting the transition of the construction industry toward low-carbon and green development. However, certain limitations remain in this study. Although AR technology and BIM provide powerful technical support, their adaptability under varying conditions requires further validation, particularly in projects of different scales and geotechnical conditions, where technical implementation challenges may arise. Additionally, the accuracy and real-time performance of carbon footprint path planning and visualization monitoring rely heavily on the precision and update frequency of BIM data, which may be constrained by data acquisition and processing efficiency in practical construction scenarios. Future research can focus on enhancing the compatibility and scalability of AR technology and BIM, exploring more intelligent and automated carbon footprint monitoring systems, and considering more complex construction scenarios and multi-dimensional environmental factors to improve the universality and practicality of the system. Furthermore, future studies could explore the integration of interdisciplinary technologies, such as artificial intelligence and big data analytics, to further optimize quality control and carbon emission management in green construction.

6 ACKNOWLEDGEMENT

This paper was funded by the Guangxi Universities Fundamental Research Capability Enhancement Project for Early/Mid-Career Faculty Members (Grant No.2024KY0930): BIM-driven Digital Construction Technology for Empowering Green Construction in Engineering Projects.

7 REFERENCES

[1] N. Sasaki and K. Fukunaga, "Global environmental protection in the construction field," *NTT Rev*, vol. 12, no. 5, pp. 29–35, 2000.

[2] H. W. Chang, Y. L. Lin, and W. M. Wey, "Advancing sustainability: Development of an ESG evaluation framework for Taiwan's science parks," *Challenges in Sustainability*, vol. 12, no. 4, pp. 255–272, 2024. https://doi.org/10.56578/cis120402

[3] H. R. Kim, D. Y. Kim, and H. S. Jang, "Analyzing barrier factors due to overseas construction market environment changes: Focused on Korean construction companies," *KSCE Journal of Civil Engineering*, vol. 23, no. 2, pp. 491–498, 2019. https://doi.org/10.1007/s12205-018-0933-6

[4] J. Underwood and M. Alshawi, "Forecasting building element maintenance within an integrated construction environment," *Automation in Construction*, vol. 9, no. 2, pp. 169–184, 2000. https://doi.org/10.1016/S0926-5805(99)00003-5

[5] J. Q. Wei, Z. H. Zhou, Q. Yu, J. W. Xie, X. Cheng, and Y. X. Hong, "Analyzing technical barriers to green farmhouse construction in China: A DEMATEL-ISM-MICMAC approach," *Journal of Green Economy and Low-Carbon Development*, vol. 3, no. 2, pp. 106–121, 2024. https://doi.org/10.56578/jgelcd030204

[6] O. Alshboul, A. Shehadeh, G. Almasabha, R. E. A. Mamlook, and A. S. Almuflih, "Evaluating the impact of external support on green building construction cost: A hybrid mathematical and machine learning prediction approach," *Buildings*, vol. 12, no. 8, p. 1256, 2022. https://doi.org/10.3390/buildings12081256

[7] J. Shurrab, M. Hussain, and M. Khan, "Green and sustainable practices in the construction industry: A confirmatory factor analysis approach," *Engineering, Construction and Architectural Management*, vol. 26, no. 6, pp. 1063–1086, 2019. https://doi.org/10.1108/ECAM-02-2018-0056

[8] Z. Yang, M. Li, E. D. Chen, H. Li, S. C. Cheng, and F. Zhao, "Research on the application of BIM-based green construction management in the whole life cycle of hydraulic engineering," *Water Supply*, vol. 23, no. 8, pp. 3309–3322, 2023. https://doi.org/10.2166/ws.2023.190

[9] J. Garbett, T. Hartley, and D. Heesom, "A multi-user collaborative BIM-AR system to support design and construction," *Automation in Construction*, vol. 122, p. 103487, 2021. https://doi.org/10.1016/j.autcon.2020.103487

[10] X. Wang, M. Truijens, L. Hou, Y. Wang, and Y. Zhou, "Integrating augmented reality with building information modeling: Onsite construction process controlling for liquefied natural gas industry," *Automation in Construction*, vol. 40, pp. 96–105, 2014. https://doi.org/10.1016/j.autcon.2013.12.003

[11] S. Meža, Ž. Turk, and M. Dolenc, "Component based engineering of a mobile BIM-based augmented reality system," *Automation in Construction*, vol. 42, pp. 1–12, 2014. https://doi.org/10.1016/j.autcon.2014.02.011

[12] J. Sermarini, R. A. Michlowitz, J. J. LaViola, L. C. Walters, R. Azevedo, and J. T. Kider, "Investigating the impact of augmented reality and BIM on retrofitting training for non-experts," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 11, pp. 4655–4665, 2023. https://doi.org/10.1109/TVCG.2023.3320223

[13] F. A. Qutaiba, S. M. Omaran, and R. S. Abd Ali, "Integrate Building Information Modeling (BIM) and occupant characteristics simulator to assess the effectiveness of emergency requirements," *International Journal of Safety and Security Engineering*, vol. 14, no. 1, pp. 37–46, 2024. https://doi.org/10.18280/ijsse.140104

[14] W. Gottmann, V. Djakona, and A. Stankevics, "Investment in housing construction: Current trends and digital technologies," *International Journal of Sustainable Development and Planning*, vol. 19, no. 6, pp. 2125–2133, 2024. https://doi.org/10.18280/ijsdp.190612

[15] B. Schiavi, V. Havard, K. Beddiar, and D. Baudry, "BIM data flow architecture with AR/VR technologies: Use cases in architecture, engineering and construction," *Automation in Construction*, vol. 134, p. 104054, 2022. https://doi.org/10.1016/j.autcon.2021.104054

[16] C. Cheng and Z. Shen, "Semi real-time detection of subsurface consolidation defects during concrete curing stage," *Construction and Building Materials*, vol. 270, p. 121489, 2021. https://doi.org/10.1016/j.conbuildmat.2020.121489

[17] H. Luo and X. Lin, "Dynamic analysis of industrial carbon footprint and carbon-carrying capacity of Zhejiang province in China," *Sustainability*, vol. 14, no. 24, p. 16824, 2022. https://doi.org/10.3390/su142416824

[18] H. Urban, N. Breitschopf, and C. Schranz, "Development and validation of an AR acceptance tool for local building supervision using the example of technical building equipment," *Bauingenieur*, vol. 97, no. 11, pp. 353–361, 2022. https://doi.org/10.37544/0005-6650-2022-11-35

8 AUTHORS

Yi Liao is a Lecturer. His research primarily focuses on BIM-based digital technologies and emerging construction materials. He currently serves as the Director of the Civil Engineering Experimental Center and leads the BIM Innovation Laboratory at the institution (E-mail: proace11@163.com; ORCID: https://orcid.org/0009-0001-1724-4381).

Huan Luo is an Associate Professor, graduated in Civil engineering. He is mainly engaged in the research of new building materials and engineering drawings, and is currently the deputy dean of the School of Civil and Architectural Engineering, mainly responsible for the management of teaching and talent training projects (E-mail: luohuan802020@163.com; ORCID: https://orcid.org/0009-0001-7311-7029).