

SEMI-AUTOMATED VISUAL QUALITY CONTROL INSPECTION DURING CONSTRUCTION OR RENOVATION OF RAILWAYS USING DEEP LEARNING TECHNIQUES AND AUGMENTED REALITY VISUALIZATION

Apostolia Gounaridou, Evangelia Pantraki, Vasileios Dimitriadis, Athanasios Tsakiris, Dimosthenis Ioannidis & Dimitrios Tzovaras

Centre for Research & Technology Hellas, Greece

ABSTRACT: *The construction industry stands to greatly benefit from the technological advancements in deep learning and computer vision, which can automate time-consuming tasks such as quality control. In this paper, we introduce a framework that incorporates two advanced tools - the Visual Quality Control (VQC) tool and the Digital Twin visualization with Augmented Reality (DigiTAR) tool - to perform semi-automated visual quality control in the construction site during the execution phase of the project. The VQC tool is a backend service that detects potential defects on images captured on-site using the Mask R-CNN algorithm trained on annotated images of concrete and railway defects. The surveyor, aided by the Augmented Reality (AR) technology through the DigiTAR tool, can in-situ confirm/reject the detected defects and propose remedial actions. All the quality control results are recorded in the relevant BIM model and can be viewed on-site overlaid on the physical construction elements. This solution offers a semi-automated visual inspection that can speed up and simplify the quality control process, especially in case of large linear infrastructures, illustrating the added value of AR-based applications in Digital Twins.*

KEYWORDS: *BIM, Augmented Reality, AR in Construction, Deep Learning, Computer Vision, Visual Inspection, Digital Twins*

1. INTRODUCTION

A prominent challenge in the construction industry is the ability to swiftly and seamlessly adapt to changes. To address this issue, an effective approach involves harnessing the power of computer-aided tools that can replace time-consuming activities. By integrating such tools into construction processes, valuable time and effort are saved, leading to significant cost reductions. By introducing digitalized processes to handle repetitive and labor-intensive tasks, construction projects can enhance their adaptability and responsiveness to changes. This allows teams to allocate their resources more efficiently, enabling them to focus on more critical aspects of the project. Furthermore, the digitalization of manual processes and the use of machine learning algorithms facilitate faster decision-making and reduce the likelihood of errors, as they can process vast amounts of data accurately and consistently. The increased accuracy and efficiency provided by these tools contribute to improved project outcomes and overall productivity.

This study introduces a semi-automated approach for visual quality control during the execution phase of construction projects. The proposed framework leverages recent technological advancements in deep learning and computer vision. It is designed to incorporate two essential components: the Visual Quality Control (VQC) tool and the Digital Twin visualization with Augmented Reality (DigiTAR) tool. The VQC tool serves as a backend service; it incorporates a deep learning network trained to detect concrete and railway defects in construction site images. The DigiTAR tool harnesses the power of AR technology to provide a unique visualization experience of the BIM model. Through DigiTAR, users can immerse themselves in the construction site and witness the 3D Building Information Modeling (BIM) model in real-time, where the digital BIM model components are overlaid onto the physical components. DigiTAR is responsible for visualizing the VQC results on-site. This means that key stakeholders, such as the project manager and quality manager of the construction project, can conveniently review and confirm these results firsthand. Their confirmation of these VQC results is pivotal, as it determines whether additional remedial works are assigned to the components identified in the VQC data. By having access to such crucial data on-site, decision-making processes can be expedited, and effective collaboration among stakeholders is further enhanced.

The novelty of the proposed approach in comparison to other existing solutions is that it simultaneously allows: (1) a collaborative inspection of construction sites (different inspectors, both in-situ and asynchronously); (2) different types of annotations (texts, strokes, images, 3D models); (3) geolocated annotations (related to specific

elements of the virtual BIM model); (4) the monitoring and editing of registered annotations; and (5) the in-situ visualization of both the designed and the actual state of the building by means of the AR technology. By incorporating these novel features, the proposed approach significantly improves the inspection process, fosters collaboration among stakeholders, and ensures higher-quality construction outcomes. This comprehensive and innovative approach addresses critical challenges faced in the construction industry, promoting efficiency and excellence throughout the project lifecycle.

The rest of the paper is structured as follows. In Section 2, related work is surveyed, focusing on: i) visual inspection methods of railways using deep learning techniques and ii) AR approaches for on-site construction inspection. In Section 3 we present the bundle of the quality control tools in detail, addressing its design and technological and implementation issues. Section 4 provides the results of the evaluation process and a case study demonstration example of utilizing the proposed framework in a real environment. Finally, the conclusion section summarizes the main findings.

2. RELATED WORK

In this section, the relevant literature review is presented. Firstly, we focus on the automated inspection of construction sites using mainly deep learning techniques. Secondly, research approaches concerning the construction sites inspection with the use of AR technology are briefly presented.

2.1 Visual inspection using deep learning techniques

In recent years, deep-learning algorithms have shown remarkable performance in image object recognition and Convolutional Neural Networks (CNNs) have attracted wide attention as an effective recognition method. CNNs have been applied successfully to detect structural damages. Many studies have been conducted focusing on binary classification issues, such as crack detection (Brien et al., 2023), including additional estimations regarding the depth of the crack (Laxman et al., 2023) or the width of the crack (Meng et al., 2023). In addition, multiple surveys have been focused on crack detection and segmentation (Attard et al., 2019; X. Xu et al., 2022), corrosion detection (Atha & Jahanshahi, 2018; Papamarkou et al., 2021), bughole detection (F. Wei et al., 2019), and multi-damage detection (Cha et al., 2018; Kumar et al., 2021).

Focusing on railways inspection, most of the studies examine the defects on the railway track lines due to the long-term pressure from train operations and direct exposure to the natural environment, which have a direct impact on the safety of train operations (Cao et al., 2020; Guo et al., 2021; Liang et al., 2019; Zhang, Liang, et al., 2021). In (Gan et al., 2017), an automatic inspection system for rail surface discrete defects due to fatigue was created and tested, extending the literature review with the Rail Surface Discrete Defects (RSDD) dataset. The Rail-5k dataset (Zhang, Yu, et al., 2021) includes the thirteen most common types of rail defects and is considered a benchmark dataset for rail surface and fastener defects. In (Zheng et al., 2021), a multi object detection method based on deep CNN is proposed, achieving a non-destructive detection of rail surface and fastener defects. In this method, rails and fasteners on the railway track images are firstly localized by YOLOv5. Then, surface defects of the rail are detected and segmented based on Mask R-CNN (He et al., 2017), while a ResNet framework is used to classify the state of the fasteners. In (X. Wei et al., 2019), the authors compare different methods for fastener defect detection and recognition, concluding that with the Faster R-CNN the fastener positioning and recognition can be carried out simultaneously. (Y. Xu et al., 2021) proposed a novel method for tunnel defect inspection (such as leakage and spalling) based on the Mask R-CNN. The network was modified appropriately (extra feature pyramid network and edge detection branch) to achieve a higher accuracy in tunnel defect detection and segmentation. (Xue & Li, 2018) proposed a fully convolutional network (FCN) model for automatic classification and detection of tunnel lining defects (such as leakage, crack, and segment joint). The authors compare their proposed method with traditional convolutional networks (such as VGG) and Faster R-CNN, concluding that the proposed model is very fast and efficient. In (Xue et al., 2020), a deep learning-based model for automatic calculation of the water leakage areas of a shield tunnel surface is proposed. Optimization measurements, such as data augmentation, transfer learning, and cascade strategy, were adopted to improve the performance of the original model.

In conclusion, many of the existing studies focus on concrete surfaces and tackle the issue of binary classification (e.g., crack/non-crack). To the best of our knowledge, studies that concern multiclass classification and detection focus mostly on long-term concrete defects. In addition, they mainly refer to bridge or rail track deterioration and defect detection.

2.2 AR for construction inspection

In (García-Pereira et al., 2020) an AR-based tool is developed for the inspection of prefabricated buildings. The tool has been evaluated positively, as it allows collaborative inspection, supports multi-type, geolocated annotations, and in-situ augmented reality visualizations. (Chi et al., 2022) present a method, which combines AR and laser-scanning technologies to provide intuitive and accurate rebar inspection. The as-built (point clouds) and as-planned data are compared to provide discrepancy information for the inspectors. With the AR, the user is able to visualize the rebar inspection outputs and provide rework instructions. (Zhou et al., 2017) propose an AR-based method to rapidly inspect segment displacement during tunneling construction. The quality inspector is able to overlay the baseline model, which is established according to the quality standard, onto the real structure and measure the differences between them. In (Kwon et al., 2014), a defect management system for reinforced concrete work is presented, utilizing BIM, image-matching, and AR. The authors developed two separate applications: an image-matching system for quality inspection without visiting the construction site (by comparing the 2D images from the BIM model with the real on-site images) and a mobile AR application for workers and managers to detect dimension errors/omissions on-site, in order to save time and reduce rework costs.

The proposed framework combines the automated image-based visual inspection, powered by advanced deep learning techniques, with AR on-site visualization and confirmation of the QC outputs. The scope is to provide an efficient solution that not only saves time, but also prevents chained construction error and reduces the need for costly reworks during the construction phase.

3. MATERIALS AND METHODS

The work presented in this paper is developed as part of the COstruction phase diGital Twin mOdel (COGITO) project (*COGITO Project*, n.d.). The COGITO project offers, among others, a bundle of tools for conducting a semi-automated visual quality control during the construction phase of large linear infrastructures (especially railways) aiming at minimizing the effort and the time usually needed for on-site visual inspection.

Within COGITO, an image-based inspection system is developed, complemented with AR visualization and interaction. Firstly, as-built data (2D images) are acquired on-site using various capturing devices, such as smartphones, cameras, and AR devices. Secondly, the acquired images are processed (e.g., cropping, resizing) by a dedicated Visual Data Pre-processing tool. At this point, each processed image is linked to a specific QC task and to the respective BIM elements depicted in the image. In the third step, the data are forwarded for the automated visual quality control. Since each image is linked to a specific element of the BIM model, the quality control results and the detected defects are also linked to elements of the BIM model (fourth step). Therefore, the inspector is able to visualize and confirm the QC results on-site using AR with each QC result pinned on the corresponding BIM element (fifth step). The inspector can either confirm or reject each detected defect and propose a rework or a mitigation work, if needed (sixth step). Finally, workers perform the proposed remedial works (seventh step). Since the defects, as well as the proposed reworks, are recorded to the BIM model, the defect management is facilitated, resulting in cost and time savings during the construction phase. The overview of the COGITO Visual QC framework is presented in Figure 1.

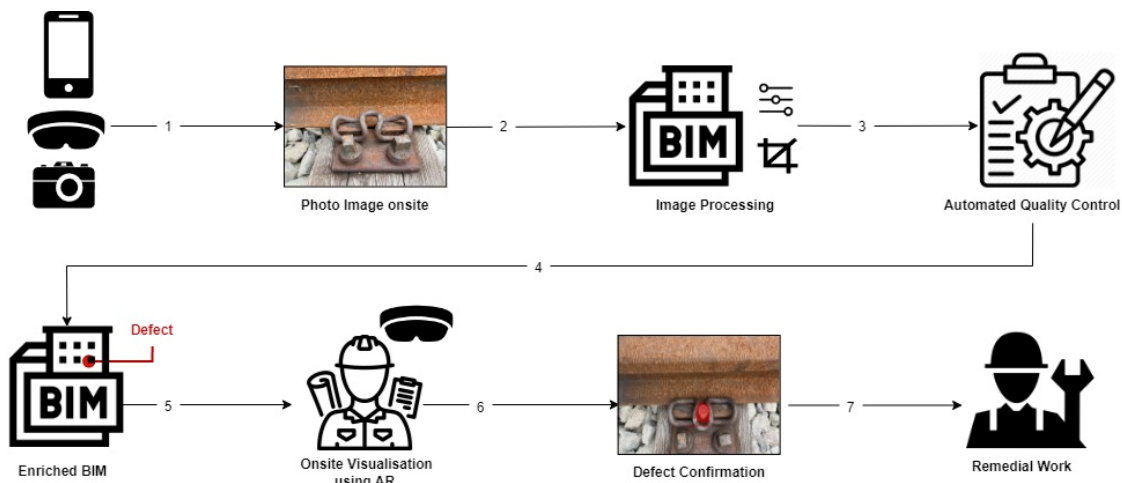


Figure 1: COGITO QC workflow

3.1 On-site Data Acquisition

Since it is necessary to capture images or videos of specific new as-built elements, various means of data acquisition can be utilized, such as cameras, mobile phones, drones and/or AR glasses (*Microsoft HoloLens2*, n.d.). Regardless of the means used for this purpose, some generic guidelines should be followed during this procedure, in order to achieve successful automated quality control and optimize the quality of the obtained results. More specifically, the images should be approximately 1000 x 1000 pixels, without spray markers or other signs that may affect negatively the QC results. They also need to be close shot and clear (not too generic or blurry) and the lighting conditions should be appropriate to ensure that the desired element is visible in the image. In case of video, the captured video will be automatically converted into a panorama image during the data processing phase. However, the video duration should be approximately five seconds (less than eight seconds) in order to generate an appropriate panorama image. In addition, a straightforward path should be followed while capturing. It is recommended to avoid rotating, shifting, sliding back or maneuvering. Finally, all the conditions for image capturing (close shot and clear, sufficient lighting conditions, without spray markers) should be also applied in case of video capturing.

3.2 Visual Data Pre-processing

After the on-site data capturing, the images need to be prepared and uploaded for the automated quality control. Within the COGITO project, this can be achieved both via a Pre-Processing Desktop application or the DigiTAR application in-situ, if the images are captured with a mobile phone or with HoloLens 2, respectively. The images should be linked to a specific QC task and BIM element before processing. The image processing includes filter application, such as modifying the contrast or the brightness of the photo and resizing or cropping it to focus on the region of interest. The aim of preprocessing is to prepare the image for the automated quality control. In case of uploading a video, a respective panorama image is generated automatically and the user is able to process it in a similar way to the normal images. Once all the desired data (images and videos) have been processed and related to a QC task, they are forwarded for the automated quality control. In Figure 2, the COGITO visual data pre-processing workflow is depicted.

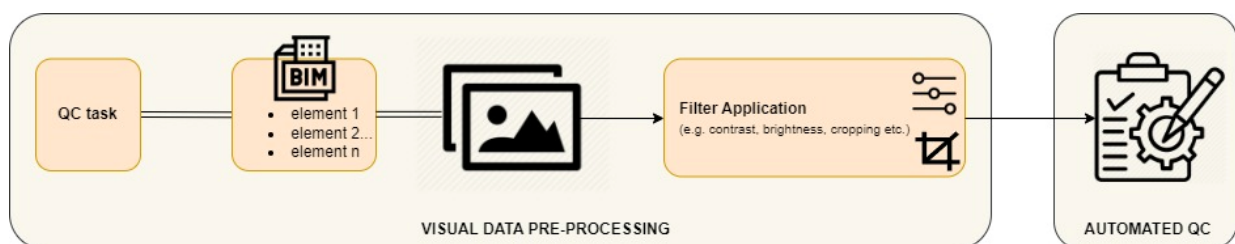


Figure 2: COGITO Visual Data Pre-Processing workflow

3.3 Automated Quality Control

The preprocessed visual data are forwarded for the automated quality control. Since the scope of the COGITO solution is to perform an automated quality control during the construction phase of new large linear infrastructures (especially railways), the VQC tool has been specifically designed to serve this purpose and the chosen defect classes for the algorithms are tailored to address construction-related issues, rather than covering defects attributed to aging of materials. Based on the deep learning algorithms employed, the VQC tool is able to detect defects, which are likely to occur during the railway construction, on concrete and steel elements. More specifically, in case of concrete surfaces, the system is able to detect cracks or honeycomb defects, while in case of railway steel elements, it detects missing clamps, missing screws, and missing screw nuts. The defect detection includes both the object detection and semantic segmentation. The goal of object detection is to classify individual defects and localize them using a bounding box and the goal of semantic segmentation is to distinguish the defects at the pixel level.

3.3.1 Dataset Preparation

For the concrete case, a dataset with concrete cracks and honeycomb images was built. The images have been combined from (*Crack Segmentation Dataset*, n.d.) and (*Concrete Crack Segmentation Dataset*, n.d.). Furthermore, additional data (with high resolution and image size) captured by Unmanned Aerial Vehicles (UAVs) were used. The original large UAVs images were divided into several smaller images using a Python script. For the steel case, a dataset was built using images collected from an above ground railway construction site in Munich. The images

depict three types of defects that can occur during the railway placement: missing clamp, missing screw, and missing screw nut. The data were resized to be consistent and have fixed dimensions (i.e., 1024 x 1024 pixels). Since the scope of the project is the automated quality control in railways, underground areas are likely to exist. Therefore, an offline data augmentation was performed to the images, in order to reduce the brightness and simulate the tunnel lighting conditions. For the concrete dataset, 1970 images were used in total to train the model for detecting the two aforementioned types of concrete defects. The ratio of the training and validation sets was almost 4:1; the training and validation sets comprise 1544 and 426 images, respectively. For the steel dataset, 2195 images were used in total to train the model for detecting the railway joint defects. The ratio of the training and validation sets was almost 5:1; the training and validation sets comprise 1720 and 493 images, respectively. The annotation of the dataset is an important and fundamental step. The image label tool LabelMe (Russell et al., 2008) was used to label the masks of the objects in both cases (concrete and steel elements).

3.3.2 Transfer Learning Implementation

The VQC tool is designed to detect defects on concrete surfaces and in steel railway elements. For this purpose, two different models (for concrete and steel case, respectively) have been trained using the Matterport's implementation of Mask R-CNN for TensorFlow2.0 (Abdulla, 2017). Mask R-CNN is an extension to the original Faster R-CNN, by adding a branch for predicting segmentation masks on each Region of Interest (RoI) using an FCN, in parallel with the existing branch for classification and bounding box regression (He et al., 2017). Therefore, Mask R-CNN not only outputs a class label and a bounding box, but also a binary mask for each detected object. The network was trained with a learning rate of 0.001, momentum of 0.90, and weight decay of 0.0001. ResNet50 was used as a backbone architecture. The IMG_SIZE and the TRAIN_ROIS_PER_IMAGE parameters were set to 512 and to 80, respectively. The RPN_ANCHOR_SCALES parameter was set to (16, 32, 64, 128, 256). The value of MAX_GT_INSTANCES and DETECTION_MAX_INSTANCES parameters were set in both cases to 5. Since a transfer learning technique was applied, the COCO dataset was used to pre-train the network and initialize its weights. Finally, only the head layers were re-trained and fine-tuned on the appropriate datasets.

The configuration of system environment was Python 3.8, Keras 2.4.3, TensorFlow 2.4.1, CUDA 11.0, and CUDNN 8.0.5 on a computer with a NVIDIA GeForce RTX 3080 GPU and a Core i7-10700 @2.9GHz CPU, with 32 GB RAM memory.

3.4 AR Visualization

The QC results obtained by the automatic visual quality control are visualized on-site with the DigiTAR tool, in order to be confirmed by the relevant stakeholders, such as the project manager and the quality manager of the construction project. Based on their decision, additional remedial works can be assigned to the components included in the VQC results. In addition, the DigiTAR tool enables the AR visualization of the BIM models. The user is able to view the 3D BIM model on-site, i.e., view the 3D BIM elements overlaying the physical elements. The workflow of the QC results confirmation process within DigiTAR is depicted in Figure 3. DigiTAR is developed using the Unity 3D Game Engine and is specifically optimized to operate on (*Microsoft HoloLens2*, n.d.) devices. In Section 3.4.1, the BIM model visualization functionality of the DigiTAR tool is described in detail, while the registration process of the BIM model is described in Section 3.4.2. Details for the visualization of the relevant QC results and the data acquisition functionality of DigiTAR are enclosed in Sections 3.4.3 and 3.4.4, respectively.

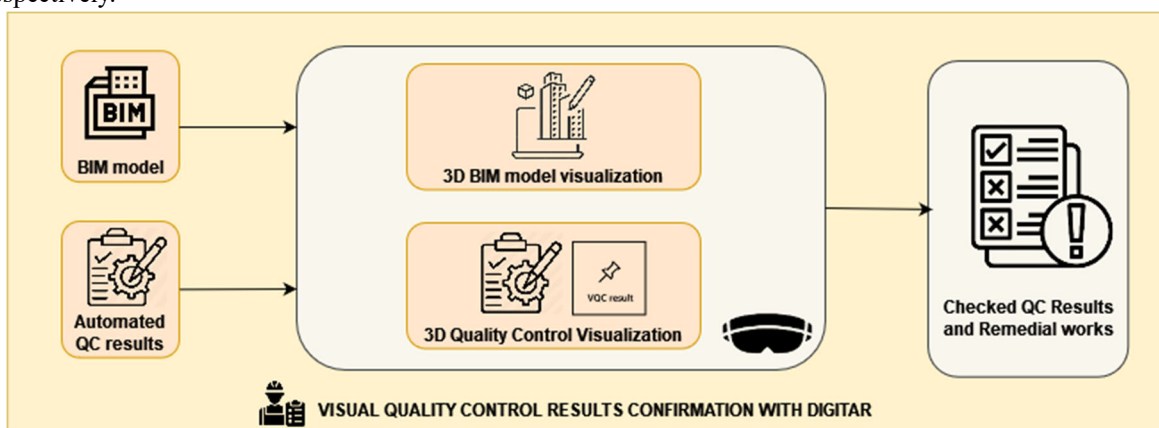


Figure 3: DigiTAR BIM model visualization and QC results confirmation workflow

3.4.1 BIM model visualization

BIM model visualization is a key functionality of the DigiTAR tool. To enable this functionality, DigiTAR requires as input the BIM model of the construction site in an Industry Foundation Classes (IFC) format. Additionally, the tool needs the 3D geometry representation of the BIM model. The geometry representation of the BIM model is achieved through the transformation of the IFC file to a file format supported by the Unity Game Engine, such as the OBJ file format.

The IFC parsing process in DigiTAR involves importing the IFC and OBJ files, extracting the IFC data, and mapping those data to the 3D model. This process is handled by custom C# classes based on the Xbim library (Lockley et al., 2017). The IFC parsing process is implemented by recursively querying and retrieving data from the IFC file for the elements of the IFC file using the IFC Schema. A GameObject is generated for each IFC element and parent/child relationships are established based on the hierarchical relationships of the elements in the IFC file. Upon completion of the IFC parsing, the result is a hierarchical structure, where each GameObject has its own IFC properties extracted in a dedicated C# class.

3.4.2 BIM model registration

After visualizing the BIM model, the next crucial step in the DigiTAR tool is registration, which involves aligning the 3D model of the construction site to the actual site. Within DigiTAR, registration relies on image targets using the Vuforia SDK (*Vuforia Engine*, n.d.). An image target is an image that the application running on HoloLens will detect and track. This image will be the link between the static 3D world (BIM model) and the real world.

The image target is printed and positioned at a location in the real world, ensuring that it is accessible to the person wearing the HoloLens. At the same time, an identical image is placed in exactly the same spot in the 3D BIM model. To enable the detection of the image target, the user uses speech command “Scan for marker”. This way, the data captured by the HoloLens sensors and cameras are utilized by DigiTAR for image target detection. More specifically, features are extracted from the HoloLens camera stream and are compared to the reference features already extracted from the image target. In the context of pattern recognition, the features that are extracted in advance from the image target constitute the pattern that the algorithm searches across the continuous flow of data streams. When the person wearing the HoloLens looks at the image target, the features extracted from the data stream of HoloLens are matched to the pattern of features belonging to the image target. Therefore, the image target is detected and registration is performed.

After successful registration of the 3D BIM model and in order to maintain it, the registered 3D BIM model is continuously tracked. In the DigiTAR application, the registration of the 3D BIM model is tracked using spatial anchors; spatial anchors represent important points in the world that the HoloLens coordinate system keeps track of over time. The registered 3D BIM model can be set as a spatial anchor using the dedicated in DigiTAR speech command “Anchor model”. This way, the next time the user opens the DigiTAR application, the 3D BIM model is loaded aligned to the real world without the need to repeat the registration process.

3.4.3 QC results visualization

The QC results are visualized using 3D QC tags that are pinned on the elements of the BIM model that are included in the QC result. The QC tags are displayed in Figure 4. To visually notify the user, the color of the tag is indicative of the relevant QC results: green if no defect has been detected, red if all QC results have detected defects, and orange if the QC results include both detected defects and no defects.

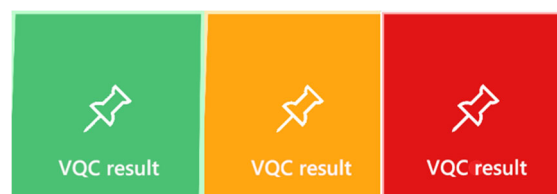


Figure 4: Visual Quality Control tags are pinned on the involved elements

Firstly, the QC tag is placed on the center of the element’s bound. When an element with a QC gets in the user’s field of view, the QC tag dynamically changes its position and rotation while staying on the surface of the 3D element. More specifically, the position of the QC tag is dynamically adjusted to the user’s height, while the rotation of the QC tag is dynamically adjusted so that the QC tag is displayed vertically in front of the user. A view

of a 3D BIM model with pinned QC tags is depicted in Figure 5. Moreover, if the user selects (using the Hand ray gesture on HoloLens¹) a 3D element that has a pinned QC tag, the QC tag follows the movement of the user's hand, while staying on the surface of the 3D element. An illustration of this feature is depicted in Figure 6. If the user performs the air tap gesture² on a QC tag, the related VQC results are displayed using dedicated AR menus, as described subsequently in Section 4.2.

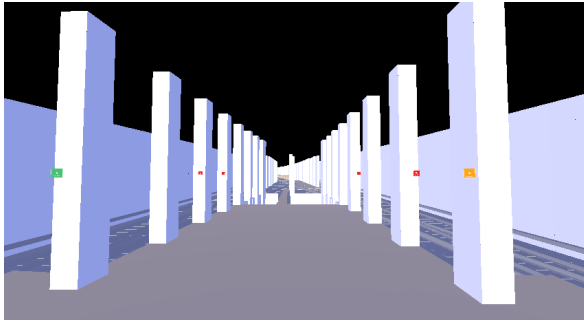


Figure 5: View of the 3D BIM model with QC tags

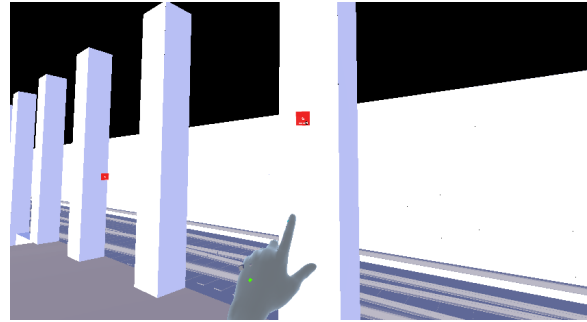


Figure 6: When selected, the QC tag follows the movement of the user's hand

3.4.4 Data acquisition and pre-processing

DigiTAR acts as a data acquisition tool for gathering images on-site to be used for automated quality control. This functionality is implemented in DigiTAR using a hand-attached menu³. When the user selects the dedicated "Capture Image" button, an asynchronous process is initiated to assess the HoloLens camera stream for photo capturing.

When the user looks at what they want to capture and say "Capture image", a photo is captured. The photo is saved in a folder of the HoloLens device. This folder, exclusively created by DigiTAR, stores only the images captured within the tool. This segregation is essential since these photos are accompanied by important metadata, including the capture time and the user's position and orientation at the time of capture. The alignment of the 3D BIM model with the real world, achieved through the registration process and spatial anchoring, enables precise association of the captured images with their corresponding locations in the BIM model.

After capturing the images, users can perform pre-processing on them before uploading them to be utilized by the automated quality control system. For this purpose, DigiTAR establishes direct communication with the backend of the Visual Data Pre-processing module. This seamless integration streamlines the process of preparing the captured images for subsequent quality control analysis, ultimately enhancing the efficiency and accuracy of the entire construction quality management process.

4. RESULTS AND DISCUSSION

The first subsection presents and analyses the evaluation process of the trained Mask R-CNN for defect detection. In the second subsection, a use case of the overall quality control process is presented, endowed with the in-situ results' visualization and confirmation via the DigiTAR application.

4.1 Automated Quality Control Evaluation

The performance of the two models (concrete and steel case) was evaluated using the mean Average Precision (mAP), since this metric is often used to evaluate object detection models. Precision is the percentage of correct positive predictions for overall predictions. Specifically, mAP is the mean value of average precision (AP) for each object class (Guo et al., 2021). The concrete model and the railway model were evaluated for 20 and 10 epochs respectively. The mAP for the concrete model reached the value of 0.87, while the mAP for the railway defects was calculated 0.95. Figure 7 shows the ground truth and the respective predictions of the proposed models for some typical examples. For each example, the generated images contain the label prediction, the confidence level, and the respective mask. The label prediction indicates the identified defect type detected by the model. The confidence level represents the model's level of certainty or confidence in its prediction. The mask displayed in

¹ <https://learn.microsoft.com/en-us/windows/mixed-reality/design/point-and-commit#hand-rays>

² <https://learn.microsoft.com/en-us/dynamics365/mixed-reality/guides/operator-gestures-hl2#air-tap>

³ <https://learn.microsoft.com/en-us/windows/mixed-reality/design/hand-menu>

the images highlights the specific region or area where the defect has been identified. This visual representation allows for a clear understanding of the location and extent of the detected defect within the image.

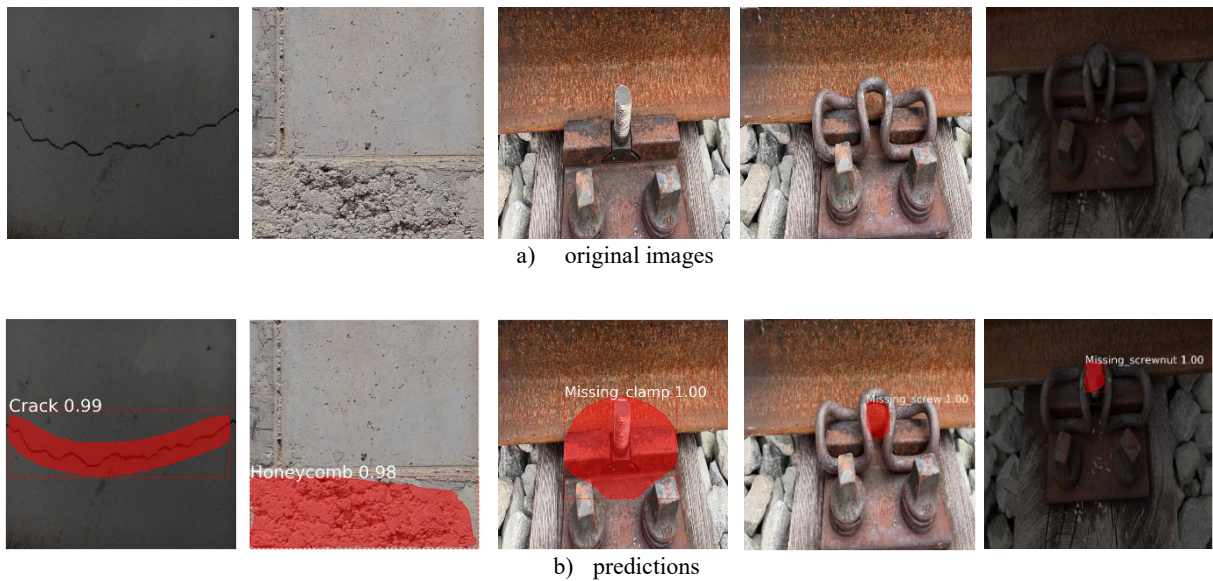


Figure 7: Original images (a) and predictions (b) for crack, honeycomb, missing clamp, missing screw, and missing screw nut.

4.2 Use Case Demonstration

The case study is focused on a railway line across Munich, Germany. The old line had to be replaced with a new one. During the reconstruction phase, the site was checked for cracks, honeycombs and rail defects, such as missing clamps, missing screws, and screw nuts.

Regarding the AR visualization, the IFC and the OBJ files for the railway site were parsed within DigiTAR using the BIM model visualization process, which is described in Section 3.4.1. The registration process, that is described in Section 3.4.2, was conducted using a strategically positioned image target within the construction zone. Precise measurements in meters, obtained from the IFC file, guided the accurate placement of the image target on-site. After the registration process was completed, the 3D BIM model became aligned with the actual construction site. This alignment allowed for accurate integration of the digital model with the real-world environment. An on-site 3D BIM model visualization using DigiTAR is illustrated in Figure 8. Figure 9 illustrates the successful visualization of the QC outcomes on HoloLens 2 using DigiTAR.



Figure 8: Screenshot of the 3D BIM model, as visualized on-site with DigiTAR



Figure 9: Screenshot of QC tags visualized on-site with DigiTAR

The surveyor captured images of the new elements using the DigiTAR tool following the procedure described in Section 3.4.4. Also, the surveyor processed the images on-site and uploaded them for automatic quality control. Utilizing the power of the VQC tool, the uploaded images underwent comprehensive assessment, generating valuable results. These results were then promoted to the DigiTAR tool for on-site inspection and confirmation. By performing the air tap gesture on a QC tag, an overview of the related Visual QC results was displayed to the surveyor, as depicted in Figure 10.



Figure 10: Overview visualization of the Visual QC results for a specific element

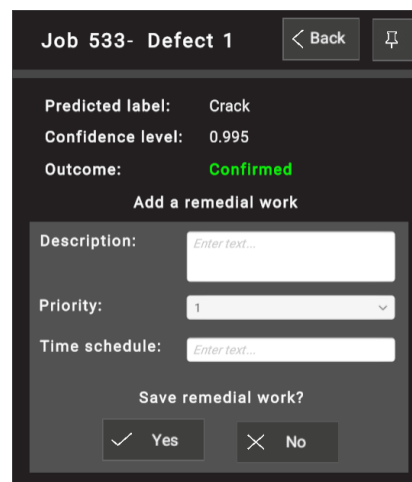


Figure 11: Menu to add remedial work to a QC result

By selecting the “Next” button in the menu in Figure 10, the surveyor could view details for the detected defect, as can be seen in Figure 12. The annotated image, the label of the detected defect and the confidence level were displayed (left figure in Figure 12). By selecting the “Original image” button, the surveyor could switch to viewing the original image that was sent for automatic visual quality control (right figure in Figure 12).

Upon confirming a detected defect, the surveyor was presented with the option to add a remedial work for the identified issue. The user-friendly menu to add a remedial work, as depicted in Figure 11, facilitated this process within the DigiTAR tool. To input the necessary information for the remedial work, the surveyor simply selected the relevant input fields on the menu. Upon selection, the HoloLens system keyboard was activated, allowing the user to type using hand gestures, making the data input intuitive and efficient.

The ability to process the remedial work in real-time within DigiTAR provided valuable advantages. It allowed for immediate consideration of mitigation measures and enabled rapid decision-making to address the identified defect effectively. This dynamic workflow streamlined the process of adding remedial works and contributed to enhanced project management and quality control.

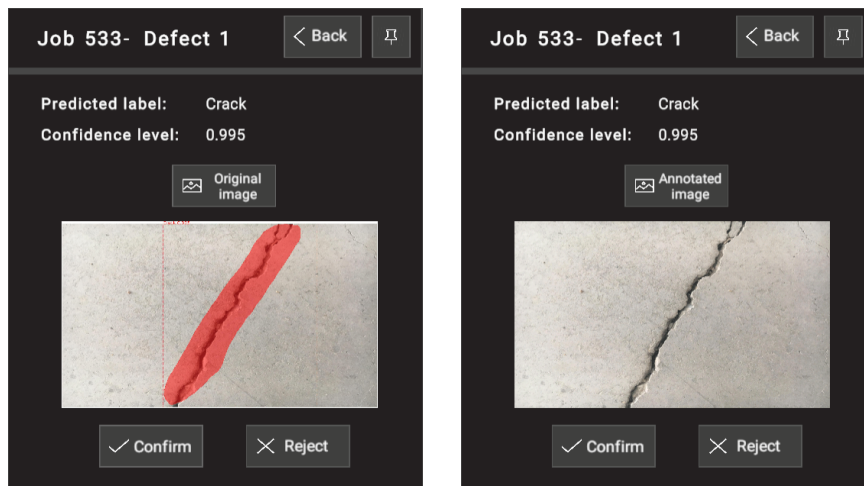


Figure 12: Visualization of a VQC result. The user can view the annotated (left) and the original image (right).

5. CONCLUSIONS AND FUTURE WORK

Embracing automation in the construction industry leads to improvements in the adaptability of the project and paves the way for greater innovation and advancement. As technology continues to evolve, leveraging automatic tools becomes a crucial aspect of staying competitive in the ever-changing construction landscape.

This study presents a framework for semi-automated visual quality control inspection in construction sites during the execution phase of the project. The framework incorporates two tools; the Visual Quality Control (VQC) tool and the Digital Twin visualization with Augmented Reality (DigiTAR) tool. The first tool incorporates a deep learning network trained to detect concrete and railway defects and serves as a backend service for automatic visual quality control on images captured at construction sites. The second tool leverages AR technology to display the visual quality control results on-site. The surveyors can inspect the detected defects in-situ and confirm or reject them. They are also prompted to add remedial works, if needed. DigiTAR displays the 3D BIM model of the construction site, i.e., the model is visualized to overlay the actual site, allowing construction professionals to interact with the BIM model in a dynamic and realistic manner using AR technology. This critical functionality enhances the overall understanding and visualization of the construction site, promoting better decision-making and coordination throughout the project lifecycle.

By combining automated quality control (performed by the VQC tool) with DigiTAR's intuitive interface and augmented reality capabilities, the surveyors gain real-time access to the quality control outcomes. This facilitates decision-making and enables prompt confirmation of the results, ensuring the construction project adheres to the highest quality standards. The seamless flow of data and information between the automatic quality control system and the DigiTAR tool enhances efficiency and accuracy, ultimately contributing to the successful execution of the construction project. The proposed framework aims to demonstrate how the synergy between cutting-edge technology and user-friendly interfaces can create a powerful asset for construction professionals in ensuring top-notch project outcomes. Future efforts will be dedicated to improving and expanding the model's training to encompass a wider range of defects. This endeavor aims to enhance the model's accuracy and efficiency in detecting various types of issues within the construction site. Additionally, the image acquisition procedure could be automatized and significantly improved utilizing drones and construction site inspection robots (such as Spot robots that are used for automated laser scanning), since our framework has been developed to support this functionality. Finally, there is a plan to equip the model with the capability to detect defects on video streams and empower DigiTAR to also display the video captures. This enhancement will enable real-time monitoring and analysis of ongoing construction activities, empowering construction professionals to address potential issues as they arise.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Union's Horizon 2020 Research and Innovation Programme through COGITO project under Grant Agreement No. 958310. Special thanks to Olympia Odos and to Rhomberg Sersa Rail Group for providing UAV images of concrete defects and images of railway steel defects, respectively. Special thanks to Rhomberg Sersa Rail Group, and especially Katharina Leer, for their valuable participation in the use case demonstration.

REFERENCES

- Abdulla. (2017). Mask R-CNN for object detection and segmentation using TensorFlow 2.0. In *GitHub repository*. <https://github.com/ahmedfgad/Mask-RCNN-TF2>
- Atha, D. J., & Jahanshahi, M. R. (2018). Evaluation of deep learning approaches based on convolutional neural networks for corrosion detection. *Structural Health Monitoring*, 17(5), 1110–1128. <https://doi.org/10.1177/1475921717737051>
- Attard, L., Debono, C. J., Valentino, G., Castro, M. Di, Masi, A., & Scibile, L. (2019). Automatic crack detection using Mask R-CNN. *2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA)*, 152–157. <https://doi.org/10.1109/ISPA.2019.8868619>
- Brien, D. O., Osborne, J. A., Perez-Duenas, E., Cunningham, R., & Li, Z. (2023). Automated crack classification for the CERN underground tunnel infrastructure using deep learning. *Tunnelling and Underground Space Technology*, 131. <https://doi.org/10.1016/j.tust.2022.104668>
- Cao, X., Xie, W., Ahmed, S. M., & Li, C. R. (2020). Defect detection method for rail surface based on line-structured light. *Measurement: Journal of the International Measurement Confederation*, 159. <https://doi.org/10.1016/j.measurement.2020.107771>
- Cha, Y. J., Choi, W., Suh, G., Mahmoudkhani, S., & Büyüköztürk, O. (2018). Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types. *Computer-Aided Civil and Infrastructure Engineering*, 33(9), 731–747. <https://doi.org/10.1111/mice.12334>
- Chi, H. L., Kim, M. K., Liu, K. Z., Thedja, J. P. P., Seo, J., & Lee, D. E. (2022). Rebar inspection integrating augmented reality and laser scanning. *Automation in Construction*, 136. <https://doi.org/10.1016/j.autcon.2022.104183>
- COGITO project*. (n.d.). Retrieved September 29, 2023, from <https://cogito-project.eu/>
- Concrete Crack Segmentation Dataset*. (n.d.). Retrieved September 29, 2023, from <https://www.kaggle.com/datasets/motono0223/concrete-crack-segmentation-dataset>
- Crack Segmentation Dataset*. (n.d.). Retrieved September 29, 2023, from https://www.kaggle.com/datasets/lakshaymiddha/crack-segmentation-dataset?select=crack_segmentation_dataset
- Gan, J., Li, Q., Wang, J., & Yu, H. (2017). A hierarchical extractor-based visual rail surface inspection system. *IEEE Sensors Journal*, 17(23), 7935–7944. <https://doi.org/10.1109/JSEN.2017.2761858>
- García-Pereira, I., Portalés, C., Gimeno, J., & Casas, S. (2020). A collaborative augmented reality annotation tool for the inspection of prefabricated buildings. *Multimedia Tools and Applications*, 79(9–10), 6483–6501. <https://doi.org/10.1007/s11042-019-08419-x>
- Guo, F., Qian, Y., Rizos, D., Suo, Z., & Chen, X. (2021). Automatic rail surface defects inspection based on mask r-cnn. In *Transportation Research Record* (Vol. 2675, Issue 11, pp. 655–668). SAGE Publications Ltd. <https://doi.org/10.1177/03611981211019034>
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). *Mask R-CNN*. <http://arxiv.org/abs/1703.06870>
- Kumar, P., Sharma, A., & Kota, S. R. (2021). Automatic multiclass instance segmentation of concrete damage using deep learning model. *IEEE Access*, 9, 90330–90345. <https://doi.org/10.1109/ACCESS.2021.3090961>
- Kwon, O. S., Park, C. S., & Lim, C. R. (2014). A defect management system for reinforced concrete work utilizing BIM, image-matching and augmented reality. *Automation in Construction*, 46, 74–81. <https://doi.org/10.1016/j.autcon.2014.05.005>
- Laxman, K. C., Tabassum, N., Ai, L., Cole, C., & Ziehl, P. (2023). Automated crack detection and crack depth prediction for reinforced concrete structures using deep learning. *Construction and Building Materials*, 370. <https://doi.org/10.1016/j.conbuildmat.2023.130709>

- Liang, Z., Zhang, H., Liu, L., He, Z., & Zheng, K. (2019). Defect detection of rail surface with deep convolutional neural networks. *Proceedings of the World Congress on Intelligent Control and Automation (WCICA), 2018-July*, 1317–1322. <https://doi.org/10.1109/WCICA.2018.8630525>
- Lockley, S., Benghi, C., & Černý, M. (2017). Xbim.Essentials: A library for interoperable building information applications. *The Journal of Open Source Software*, 2(20), 473. <https://doi.org/10.21105/joss.00473>
- Meng, S., Gao, Z., Zhou, Y., He, B., & Djerrad, A. (2023). Real-time automatic crack detection method based on drone. *Computer-Aided Civil and Infrastructure Engineering*, 38(7), 849–872. <https://doi.org/10.1111/mice.12918>
- Microsoft HoloLens2. (n.d.). Retrieved September 29, 2023, from <https://www.microsoft.com/en-us/hololens/>
- Papamarkou, T., Guy, H., Kroencke, B., Miller, J., Robinette, P., Schultz, D., Hinkle, J., Pullum, L., Schuman, C., Renshaw, J., & Chatzidakis, S. (2021). Automated detection of corrosion in used nuclear fuel dry storage canisters using residual neural networks. *Nuclear Engineering and Technology*, 53(2), 657–665. <https://doi.org/10.1016/j.net.2020.07.020>
- Russell, B. C., Torralba, A., Murphy, K. P., & Freeman, W. T. (2008). LabelMe: A database and web-based tool for image annotation. *International Journal of Computer Vision*, 77(1–3), 157–173. <https://doi.org/10.1007/s11263-007-0090-8>
- Vuforia Engine. (n.d.). Retrieved September 29, 2023, from <https://developer.vuforia.com/downloads/SDK>
- Wei, F., Yao, G., Yang, Y., & Sun, Y. (2019). Instance-level recognition and quantification for concrete surface bughole based on deep learning. *Automation in Construction*, 107. <https://doi.org/10.1016/j.autcon.2019.102920>
- Wei, X., Yang, Z., Liu, Y., Wei, D., Jia, L., & Li, Y. (2019). Railway track fastener defect detection based on image processing and deep learning techniques: A comparative study. *Engineering Applications of Artificial Intelligence*, 80, 66–81. <https://doi.org/10.1016/j.engappai.2019.01.008>
- Xu, X., Zhao, M., Shi, P., Ren, R., He, X., Wei, X., & Yang, H. (2022). Crack detection and comparison study based on faster R-CNN and Mask R-CNN. *Sensors*, 22(3). <https://doi.org/10.3390/s22031215>
- Xu, Y., Li, D., Xie, Q., Wu, Q., & Wang, J. (2021). Automatic defect detection and segmentation of tunnel surface using modified Mask R-CNN. *Measurement: Journal of the International Measurement Confederation*, 178. <https://doi.org/10.1016/j.measurement.2021.109316>
- Xue, Y., Cai, X., Shadabfar, M., Shao, H., & Zhang, S. (2020). *Deep learning-based automatic recognition of water leakage area in shield tunnel lining*. <https://doi.org/10.17632/xz2nykszbs.1>
- Xue, Y., & Li, Y. (2018). A fast detection method via region-based fully convolutional neural networks for shield tunnel lining defects. *Computer-Aided Civil and Infrastructure Engineering*, 33(8), 638–654. <https://doi.org/10.1111/mice.12367>
- Zhang, Z., Liang, M., & Wang, Z. (2021). A deep extractor for visual rail surface inspection. *IEEE Access*, 9, 21798–21809. <https://doi.org/10.1109/ACCESS.2021.3055512>
- Zhang, Z., Yu, S., Yang, S., Zhou, Y., & Zhao, B. (2021). *Rail-5k: A real-world dataset for rail surface defects detection*. <http://arxiv.org/abs/2106.14366>
- Zheng, D., Li, L., Zheng, S., Chai, X., Zhao, S., Tong, Q., Wang, J., & Guo, L. (2021). A defect detection method for rail surface and fasteners based on deep convolutional neural network. *Computational Intelligence and Neuroscience*, 2021. <https://doi.org/10.1155/2021/2565500>
- Zhou, Y., Luo, H., & Yang, Y. (2017). Implementation of augmented reality for segment displacement inspection during tunneling construction. *Automation in Construction*, 82, 112–121. <https://doi.org/10.1016/j.autcon.2017.02.007>