Innovations Deserving
Exploratory Analysis Programs

NCHRP IDEA Program

Mixed Reality Assisted Infrastructure Inspections

Final Report for
NCHRP IDEA Project 222

Prepared by:
F. Necati Catbas
Joseph L. Laviola
Enes Karaaslan
Mahta Zakaria
University of Central Florida

February 2023

NATIONAL ACADEMIES Sciences Engineering Medicine

TRB TRANSPORTATION RESEARCH BOARD

Innovations Deserving Exploratory Analysis (IDEA) Programs Managed by the Transportation Research Board

This IDEA project was funded by the NCHRP IDEA Program.

The TRB currently manages the following three IDEA programs:

- The NCHRP IDEA Program, which focuses on advances in the design, construction, and maintenance of highway systems, is funded by American Association of State Highway and Transportation Officials (AASHTO) as part of the National Cooperative Highway Research Program (NCHRP).
- The Safety IDEA Program currently focuses on innovative approaches for improving railroad safety or performance. The program is currently funded by the Federal Railroad Administration (FRA). The program was previously jointly funded by the Federal Motor Carrier Safety Administration (FMCSA) and the FRA.
- The Transit IDEA Program, which supports development and testing of innovative concepts and methods for advancing transit practice, is funded by the Federal Transit Administration (FTA) as part of the Transit Cooperative Research Program (TCRP).

Management of the three IDEA programs is coordinated to promote the development and testing of innovative concepts, methods, and technologies.

For information on the IDEA programs, check the IDEA website (www.trb.org/idea). For questions, contact the IDEA programs office by telephone at (202) 334-3310.

IDEA Programs
Transportation Research Board
500 Fifth Street, NW
Washington, DC 20001

Mixed Reality Assisted Infrastructure Inspections

NCHRP IDEA Program Final Report

IDEA Project NCHRP-222

Prepared for

The NCHRP IDEA Program
Transportation Research Board
National Academies of Sciences, Engineering, and Medicine

by

F. Necati Catbas, *PhD, PE, Professor, Principal Investigator*

Joseph J Laviola, Jr, *PhD, Professor, co-Principal Investigator*

Enes Karaaslan, *PhD, Post-doctoral Research Associate*

Mahta Zakaria, *PhD Student, Graduate Assistant*

University of Central Florida,

February 28, 2023

TABLE OF CONTENTS

ACKNOWLEDGMENT

This research reported herein was performed under NCHRP IDEA 222 Project by Civil, Environmental and Construction Engineering (CECE) Department at the University of Central Florida (UCF). Dr. Necati Catbas, P.E. served as the Principle Investigator, while Dr Enes Karaaslan started this research as graduate student and then continued as post-doctoral research associate. For the idea of mixed reality assisted infrastructure inspections, the PI and his team closely collaborated with Dr. Ulas Bagci of Northwestern University (formerly at UCF Computer Science). This collaboration led to two awards and a US Patent. Dr. Joe LaViola of UCF provided key insights for the development and execution of the project.

The research team would like to thank the project advisory panel for their time to meet, for reviewing project progress and providing insightful input. The advisory panel consists of Dr. Hoda Azari and Dr. Ping Lu (Federal Highway Turner Fairbanks Research Center); Mr. Felix Padilla, P.E. (Florida Department of Transportation), Mr. Masato Matsumoto, P.E. (Nexco West USA Inc) and Dr Ulas Bagci (Northwestern University). Their contributions and support are greatly appreciated. The comments and valuable input from the advisory panel are greatly appreciated. The research team would also like to thank Dr. Inam Jawed, Senior Program Officer of NCHRP IDEA program, for offering invaluable guidance, assistance and support during the course of this project. Mr. Ahmad Abu-Hawash, P.E. (formerly with Iowa Department of Transportation) started as the IDEA program advisor and then Mr. Darryll Dockstader (Florida Department of Transportation) took over as the IDEA program advisor. They are to be acknowledged for their input for the project coordination and direction.

The PI thanks UCF Preeminent Postdoctoral Program (P3) for providing additional support to Dr. Karaaslan during the project. Finally, several graduate and undergraduate students of CITRS (Civil Infrastructure Technologies for Resilience and Safety research lab directed by Dr. Catbas), made contributions to this research for data collection, field work, and research discussions. Dr. Chuanzhi Dong (former doctoral student and then post-doc at CITRS) made significant contributions to computer vision related research activities. Mr. Furkan Luleci (doctoral student at CITRS) is also acknowledged for supporting editing efforts and also for significant contributions to a parallel study that led to an extensive literature review paper on Extended Reality (XR) for condition assessment of civil engineering structures.

GLOSSARY

AI	Artificial Intelligence
AP	Average Precision
AR	Augmented Reality
BIM	Building Information Model
FN	False Negative
FP	False Positive
FPS	Frame per Second
HMD	Head-Mounted Device
IoU	Intersection over Union
IoT	Internet of Things
mAP	Mean Average Precision
MR	Mixed Reality
SHM	Structural Health Monitoring
TP	True Positive
UI	User Interface

EXECUTIVE SUMMARY

The need for timely and accurate infrastructure inspections increases as America's infrastructure ages. According to the most recent report by ASCE, the overall grade of United States infrastructure has dropped to C-, brought down by the bridge score dropping to C in 2021. There are currently 619,588 bridges in the United States, with 43,578 in poor condition and 291,339 in fair condition. While the number of bridges in poor condition has dropped by approximately 1500, the number in fair condition has increased by twice this value. According to ARTIBA, with this repair rate, it will take 30 years to repair all bridges in the United States.

One important factor in bridge maintenance is visual inspection. Timely and accurate inspections may result in better bridge management and decision-making. However, the process of visual inspection is time-consuming and labor-intensive. Inspectors must conduct site visits to find apparent defects on the structure's surface. Roads must be closed for a complete and thorough inspection of hard-to-reach parts of the structure, such as the deck's surface or under the deck. Furthermore, studies show that inspection reports are not always consistent. The variation in results becomes more apparent for defects in moderate or severe states.

In light of all that has been mentioned, there is a need for more efficient and accurate visual inspections; this can be achieved using emerging technology. Recent advances in computer vision algorithms and learning-based methods have introduced more accurate ways of handling visual data. Convolutional Neural Networks (CNN) are vastly used in medical imaging, transportation, and autonomous driving. Researchers have also explored the possibility of using such methods in visually inspecting structures in recent years. However, CNNs are highly dependent on the size of the training dataset. Unlike other practices with abundant training data, the surface defects dataset is still limited. Crack datasets are among the biggest datasets in civil engineering. They are largely very uniform and only include defect areas. The lack of diversity in such datasets makes it difficult for inspectors to use them in different environments. As a result, most of the available studies are case-specific. In addition, CNN models cannot achieve reliable accuracy with such small datasets, making post-processing necessary and reducing the method's efficiency.

This project introduces a methodology with the aid of edge computation, bringing analysis to the site. The designed methodology uses AI models to localize and quantify surface defects on concrete bridges in real time, allowing inspectors to supervise the results of defect identification. Human supervision improves the accuracy of the results and eliminates the need for post-processing, increasing efficiency. To facilitate human supervision, defect localization and quantification are separated using two separate CNN models. This method allows the inspector to oversee the results of the defect localization model and arrange the threshold for better detection. Furthermore, attention-guided segmentation results in more accurate results from defect quantification.

Many different CNN architectures were tested for defect localization and quantification. For defect localization, different object detection architectures, including SSD-MobileNetV3 (*1*), SSD-MobileDet (*2*), EfficientDet-D0 (*3*), YOLOv4-Tiny (*4*), and YOLOv5s (*5*), were compared. The best model, YOLOv5s, was selected as it performs significantly faster than other architectures, has higher accuracy, and has a smaller footprint when compared to other architectures. For defect quantification, UNet, FCN, LinkNet, and PSPNet with three different classification backbones were trained and compared. Unet, with the

backbone of efficientnetb0, was the most accurate model. Although PSPNet had a higher inference speed, it was eliminated since defect quantification accuracy has greater importance than inference speed.

The proposed AI model was optimized and quantized to obtain high inference speed and a small memory footprint, making it suitable for real-time performance in edge computing devices. The model was then deployed into a Mixed Reality headset, HoloLens 2, to develop a proof-of-concept device. Figure 1 displays the design process of this project. The Mixed Reality platform combines the 3D world with the virtual world, providing an excellent environment for human-computer interaction. A user interface was developed in Unity to facilitate human-AI collaboration. Finally, the proof-of-concept device was tested for inspection of a bridge for evaluation.

Figure 1. Process of design for MR-assisted Inspection.

LIST OF FIGURES

LIST OF TABLES

CONCEPT AND INNOVATION

This project introduces a novel visual inspection methodology that uses a learning-based computer vision algorithm optimized for real-time operation in edge devices (i.e., wearable devices, smartphones, etc.). The explored system aims to assist the inspector by accelerating routine tasks such as localizing and measuring the surface defects of structures (i.e., cracks, spalling). Unlike other studies, where Artificial Intelligence (AI) completes the task of defect localization or quantization, in this method, the human inspector interacts with AI in the system through the human-machine interface of the edge device. The designed AI models in this study are optimized with lightweight architectures that can bridge the AI with the edge device, meaning that data collection, preprocessing, and analysis are all done in real time under the inspector's supervision. The algorithm is designed to be deployed in edge computing devices that facilitate human-computer interaction, such as Augmented Reality (AR) or Mixed Reality (MR) platforms (Figure 2). A wearable holographic see-through headset was chosen for a case study to investigate the system's effectiveness. In this study, the human-AI interaction is facilitated using the Mixed Reality platform in HoloLens 2.

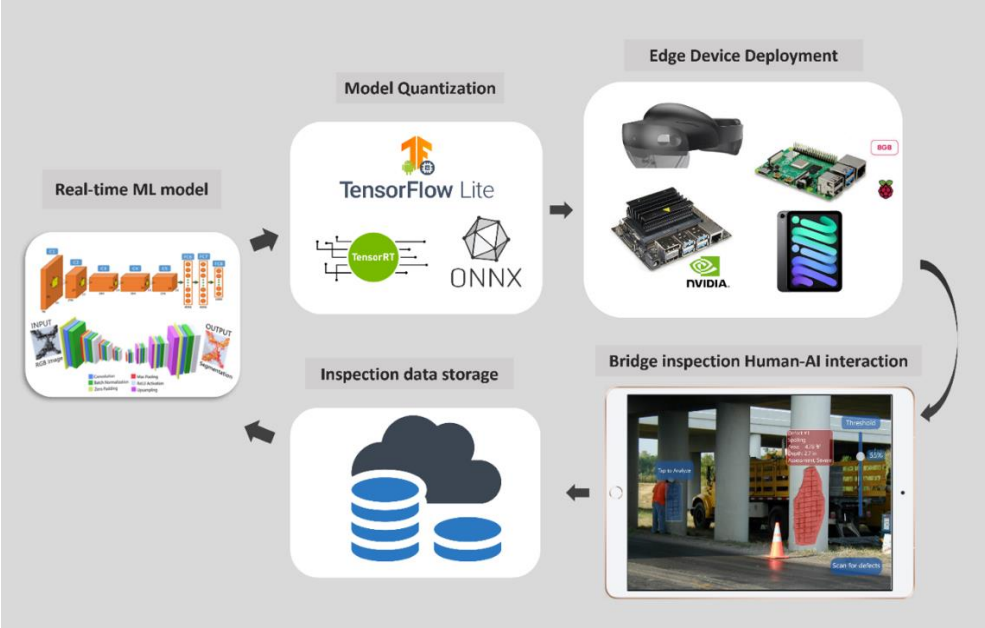

Figure 2. Illustration of Advanced bridge inspection using a real-time learning-based computer vision algorithm on an edge device(*6*)**.**

This project combines the efficiency and accuracy of Machine Learning-based models with the mixed reality platform, which offers a suitable environment for human-computer interactions. The optimized AI models in this study were performed in real-time without a strong processor. Unlike similar studies, the results of the AI are constantly controlled and approved by professional inspectors to ensure its accuracy. After each inspection, the collected annotated data can be used for fine-tuning the AI models and improving their accuracy. The deep learning model's accuracy is directly related to the dataset size and quality. The annotation process is the most labor-intensive task in training the models. In this project, the data collected

during each inspection is annotated by the AI models and verified by the inspector. After each inspection, more data from various bridges in different environments is collected, creating a diverse human-verified annotated dataset. While the current models cannot inspect without post-processing, the dataset generated using MR-based inspection will eventually lead to training a highly accurate AI model capable of performing automated inspections.

IDEA PRODUCT

The product developed in this project is an API conducting infrastructure visual inspection using Artificial Intelligence compatible with edge devices supporting Mixed Reality and Augmented Reality. The main objective of the AI integrated Mixed Reality (MR) system developed in this study is to assist the inspector by accelerating certain routine tasks, such as measuring all cracks in a defect region or calculating the area of spalling. In this system, the human-centered AI interacts with the inspector instead of completely replacing human involvement during inspections. This collective work will lead to quantified assessment and reduced labor time while ensuring human-verified results.

This project uses two CNN models: YOLOv5 for object detection and UNet for segmentation, which is used for localization and quantifying concrete surface defects. The developed technology in this project improves defect quantification results by limiting the search area to localized defects (attention-guided quantification). It also allows the inspector to communicate with AI results by separating the two tasks, making sure that all the localized defects in the view are correctly detected and quantized. This API uses a Mixed Reality platform to facilitate human-AI interaction. The developed technology can be deployed in MR headsets (HoloLens 2) for routine inspections. The developed user interface (UI) allows the inspector to verify AI results, correct bounding boxes, and change thresholds for better results (Figure 3).

The plans for implementing the IDEA product are detailed later in the report. A patent was filed entitled "Artificial Intelligence Collaborative System for Enhanced Inspections of Infrastructures Using Mixed Reality" (US Patent App. 17/247,389, 2021).

Figure 3. Defect localization with Human-AI interaction in the Mixed Reality platform.

INVESTIGATION

INTRODUCTION

Bridges are one of the most critical elements of transportation systems, and therefore their continuous operation is of the utmost importance for safe and efficient transportation. According to the 2022 bridge report on bridge conditions published by the U.S. Department of Transportation (DOT), Federal Highway Administration, out of 619,588 U.S bridges, more than 43,500 are in poor condition, which is 1,445 fewer bridges than in 2020. However, nearly half, 48%, are in fair condition, with 2,916 more than in 2020 (*7*). The ARTBA 2022 Bridge Report estimates that with the current rate of bridge repair, it will take 30 years to repair all bridges in poor condition in the US. Iowa, Pennsylvania, and Illinois have the largest number of bridges in poor condition. On the other hand, the bridge inventory in Florida ranks among the best in the nation, with only 415 bridges in poor condition (*8*). The Florida DOT inspects all public highway bridges in the State, and based on the bridge inventory; all bridges have been inspected at least once in the last two years, highlighting the importance of bridge inspection in the maintenance of bridges. However, conventional methods in bridge inspection are subjective, labor-intensive, and require hours of road closures. Major landmark structures have permanent structural health monitoring (SHM) systems, and non-destructive evaluation (NDE) of different components is carried out. However, most bridges or ancillary structures cannot be efficiently monitored or tested. Some of the common techniques used in ancillary structure inspections are visual inspection, dye-penetrant testing (PT), magnetic particle inspection (MPI), ultrasonic testing (UT), and infrared thermography (IRT). The disadvantages of these methods are their dependencies on human experience, ineffectiveness for rusty and rough surfaces, ineffectiveness for complex shapes, and labor intensity. Different NDE technologies, their usage and comparative analysis can be found in many papers in the literature.

Structural assessment using local and global levels using computer vision technologies has gained much attention in the structural health monitoring community in research and practice. Dong and Catbas (2020) present a general overview of the concepts, approaches, and real-life practice of computer vision–structural health monitoring along with some relevant literature that is rapidly accumulating in an extensive literature review. The computer vision–structural health monitoring covered in this review at local level includes applications such as crack, spalling, delamination, rust, and loose bolt detection. At the global level, applications include displacement measurement, structural behavior analysis, vibration serviceability, modal identification, model updating, damage detection, cable force monitoring, load factor estimation, and structural identification using input–output information (*9*).

This study, presented in this report, investigated effective, accurate, and practical methods of structural inspection using non-contact, real-time Machine Learning-based (ML) Computer Vision (CV) algorithms in conjunction with structural models and considerations. ML-based CV algorithms have advanced extensively in civil engineering applications for local and global structural characterization and assessment over the last decade. As such, these approaches were the first to be investigated to inspect concrete surface defects in real-time on an interactive edge computing device. This study generated annotated high-quality datasets for improving the performance of trained ML models. Bringing the processing to the edge increases inspection efficiency by producing analyzed data that incorporates human expertise. Furthermore, the method's advantages and disadvantages were investigated using a proof-of-concept Mixed Reality device.

The proposed framework applies to any family or population of structures with intermittent, portable, and rapid monitoring approaches, addressing a societal need. From a technological point of view, the real-time ML-based algorithms in this project can be deployed in any edge device, assisting inspectors during routine visual inspections of concrete structures. The method generates supervised-annotated data to fine-tune available ML models and improve their accuracy. Over time, the inspections can be conducted unsupervised using UAVs and inspection robots.

Current Practice in Concrete Bridge Inspection

By U.S. federal regulations, all structures that carry traffic and span over 20 feet are subject to comprehensive inspection at least every two years. The regulations define eight types of bridge inspection. Three are periodic: routine inspection, fracture-critical member inspection, and underwater inspection. State DOTs, on the other hand, establish more detailed guidelines for the periodic use of hands-on inspection, close-up access, and collection of quantitative data. State DOTs define guidelines for short-interval, interim inspections in response to bridge defects, conditions, or load posting. State DOTs also establish guidelines for in-depth long-interval inspections for selected bridge types and elements (*10*).

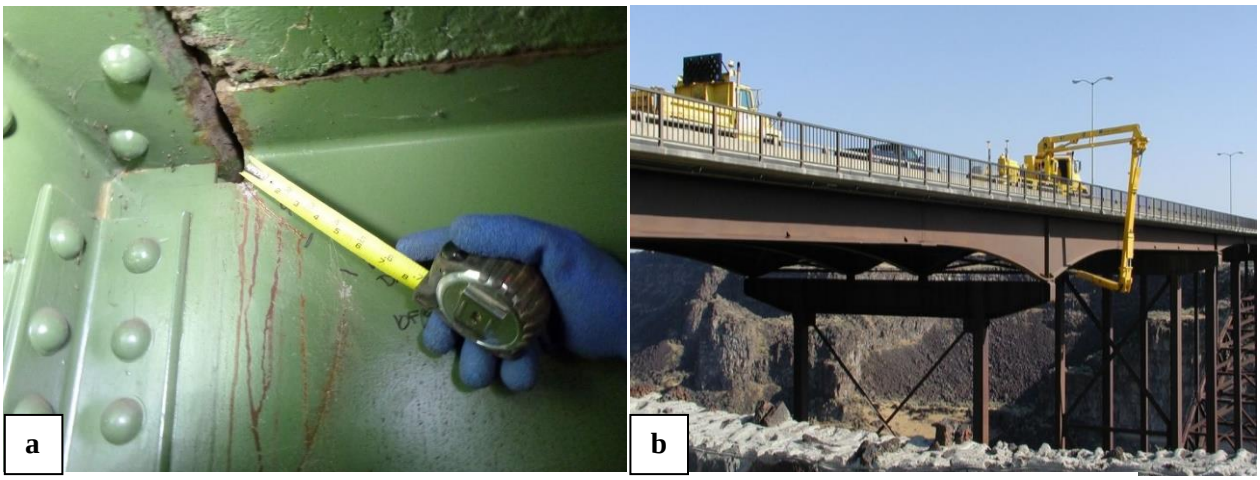

Figure 4. Visual Inspection using basic hand tools and snooper truck (photos from Canva)

In a routine biennial inspection, trained bridge inspectors check for obvious damage, which may take varied forms, such as spalled concrete, corroded steel, and even insect and fungus damage on timber elements. They also examine bearings, deck drains, and expansion joints for proper operations; evaluate the serviceability of bridge substructures, decks, approaches, and appurtenances; and, for waterway crossings, inspect the channel for scouring and flow obstructions. These elements are inspected visually, using basic hand tools where appropriate (Figure 4-a). A snooper truck is commonly used if the side deck or under deck needs to be inspected, as shown in Figure 4-b.

These federally mandated inspections aim to assess and document the condition of essential bridge elements to ensure safety and serviceability and to facilitate the timely programming of maintenance and

repairs. Some bridges are also subject to special inspections and/or in-depth evaluations of the safety and serviceability of particular elements known to have specific problems or present particular risks. For example, special inspections are conducted on fracture-critical bridges, which are those with non-redundant steel tension components, the fracture of which would likely cause catastrophic failure. Traditional slow and subjective methods used in assessing deck conditions, which include impact sounding (Figure 5-a), chain dragging (Figure 5-b), half-cell potential, and core analysis, are being replaced by more modern techniques.

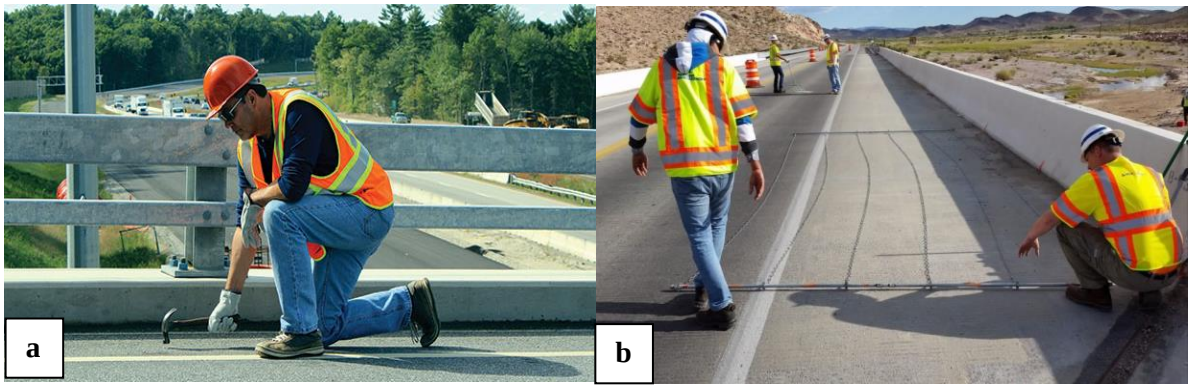

Figure 5. Traditional methods of assessing deck condition (*88*)**.**

Concrete bridge decks deteriorate faster than other bridge components due to their direct exposure to traffic along with environmental effects and other factors such as salting for de-icing. Thus, the FHWA's LTBP Program considers them the highest priority when fixing bridge performance issues. Since most DOTs spend 50-80% of their bridge expenditures on repair, rehabilitation, and replacement of concrete bridge decks, better methods are needed to detect defects and quantify the extent and severity of bridge deck conditions early, accurately, and rapidly, with minimal traffic impact, ideally, without lane closures for bridge deck inspections. Under these circumstances, NDE techniques such as HD image-based crack detection, impact echo (IE), ultrasonic surface waves (USW), electrical resistivity (ER), ground-penetrating radar (GPR), and IRT have been developed to inspect and monitor aging and deteriorating structures rapidly and effectively in place of visual and sounding inspection methods. Some of these NDE techniques are shown in Figure 6.

Figure 6. Different NDE technologies used in bridge deck inspection (*11*)**.**

Current Practice in Bridge Visual Inspection Using Computer Vision Algorithms

Efforts to use computer vision technology to detect visual surface defects on structures have been ongoing for almost a decade. The early approaches included edge detection, template matching, and segmentation (*12*). German et al. (2012) used an entropy-based thresholding algorithm in conjunction with image processing methods in template matching and morphological operations to detect spalling and cracks. Nguyen et al. (2014) also generated a filter-based algorithm for edge detection to extract crack edges on concrete surfaces (*13*). Recent advances in artificial intelligence (AI) and Machine learning-based computer vision algorithms have revolutionized many computer vision tasks, including detecting and measuring structural defects.

To process the image data of concrete defects, researchers in the literature implemented Convolutional Neural Network (CNN) to perform automatic crack detection on concrete surfaces. Combined with transfer learning and data augmentation, Yokoyama and Matsumoto (2017) developed a CNN-based crack detector with 2000 training images (*14*). Similarly, Zhang et al. (2020) used a UNet-based algorithm to detect cracks pixel-wise in a more precise manner (*15*) (figure 10). Zhang et al. (2020) used a CNN-LSTM model to detect, in real-time, the concrete bridge deck cracks on a frequency domain (*16*). The algorithm extracts the whole crack from its background. The conventional edge detection-based approaches just segment the crack edges, offering a more feasible solution for crack thickness identification. Deng et al. (2020) use a faster Region-based CNN architecture to automatically detect concrete cracks even with handwriting (*17*). Adhikari et al. (2014) used 3D visualization of crack density by projecting digital images and neural network models to predict the crack depth and necessary information for the condition assessment of concrete components (*18*). A study by Ren et al. (2020) introduced a CNN model named CrackSegNet to conduct a dense pixel-wise segmentation of cracks in tunnels. The authors use 409 images collected from a single tunnel as the training dataset (*19*). Multiple studies have also been conducted on measuring and classifying pavement cracks. One recent study by Eslami and Yun designed an attention-based CNN network that improves the classification results of pavement cracks compared to automated systems(*20*).

Another study uses YOLO (You Only Look Once) algorithm for real-time detection of cracks. The sizes of the detected cracks are calculated based on the positions of the projected laser beams on the structural surface (Figure 7&8). While much of the available research focuses on detecting and measuring cracks on concrete surfaces or pavements, more recent studies also focus on detecting and quantifying other surface defects. A study by Xu et. al. uses Mask R-CNN for automatic defect detection in tunnels. The authors endow a path augmentation and edge detection branch to the network to improve its accuracy, decreasing the computational efficiency significantly. The authors focus on detecting leakage, spalling, and other objects in tunnels, such as pipes and edges. Another study by Zhang et al. uses simultaneous object detection and segmentation to segment cracks, spalling, and exposed rebar on the surface of the concrete. The model consists of two separate parts: YOLOv3 and Sub Mask architecture, where the bounding box task is conducted separately from the segmentation. The study uses a total of 1440 images for training and testing. The proposed method has a relatively good speed but is not fast enough for real-time inspection. Similarly, the accuracy is acceptable; however, the results depend on the bounding boxes obtained from the object detection model and, therefore, are unsuitable for automated inspection.

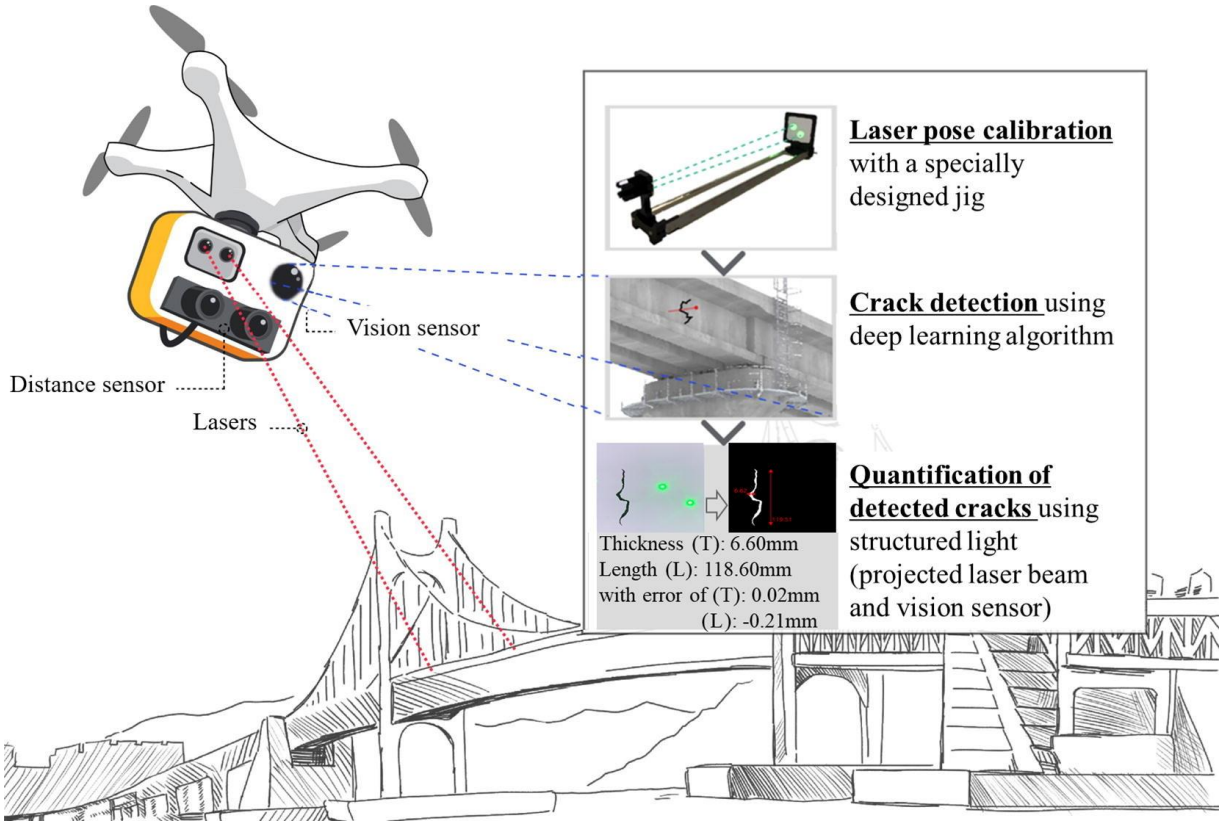

Figure 7. crack detection and quantification using deep learning and structured light(*21*)**.**

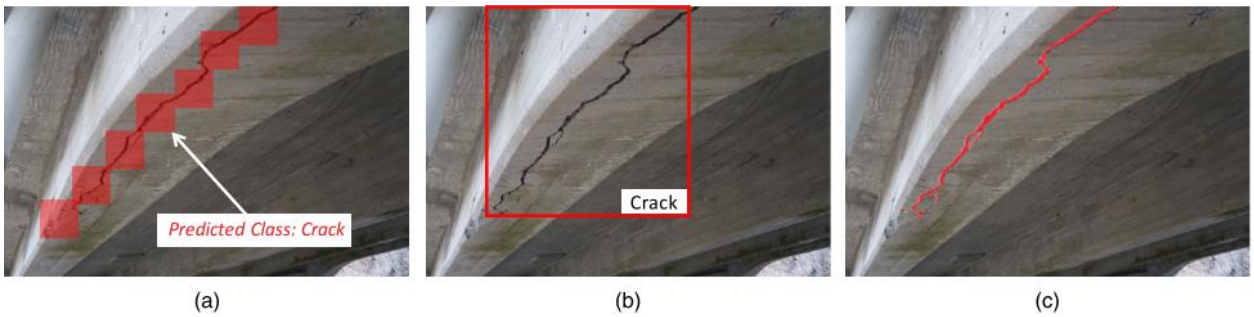

Figure 8. Visual recognition formulations of structural defect detection: (a) image classification; (b) object detection; and (c) semantic segmentation (*21*)**.**

One review paper on computer vision-based detection and condition assessment of concrete infrastructure defects emphasizes the importance of sufficiently large, publicly available, and standardized datasets to leverage existing supervised machine learning methods for damage detection (*22*). There are limited open-source datasets for concrete defects available, including SDNET2018 (*23*), which includes 58000 cropped images of cracks on concrete surfaces; CODEBRIM with approximately 1700 original images of surface defects, including cracks, spalling, efflorescence, and rusting, as well as cropped images of the same defects (*24*); and finally COCO-Bridge, which includes 774 images with 2500 instances of bridge elements (figure 9) (*25*). While the available datasets are adequate for classification purposes, the number of open-source original images needs improvement for automated defect localization and quantification.

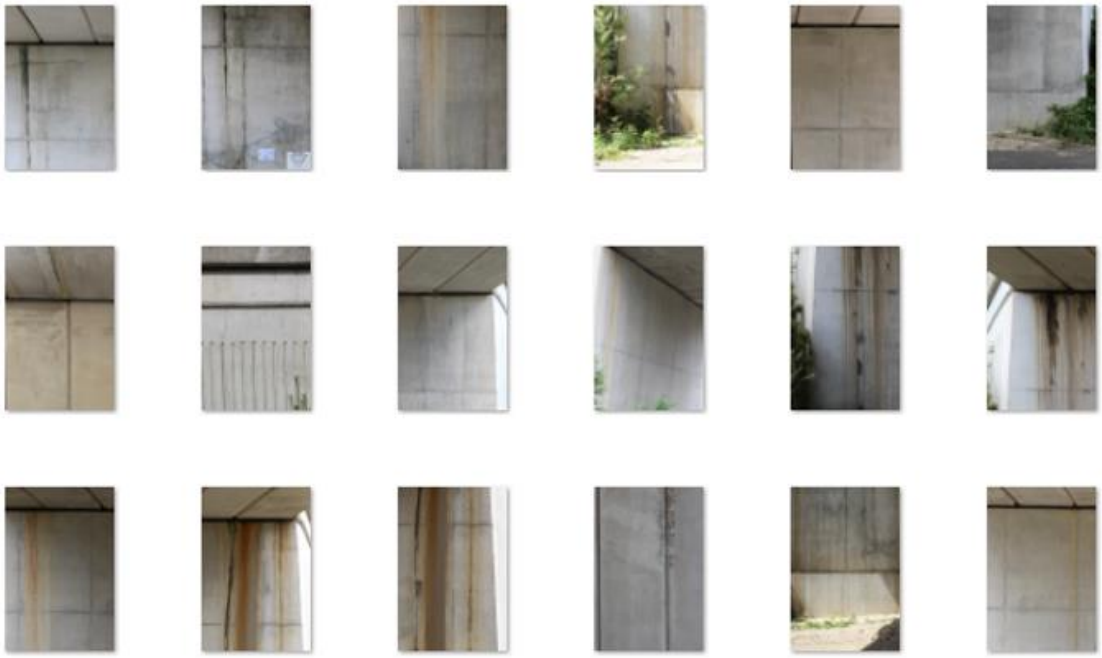

Figure 9. A batch of CODEBRM dataset (*24*)**.**

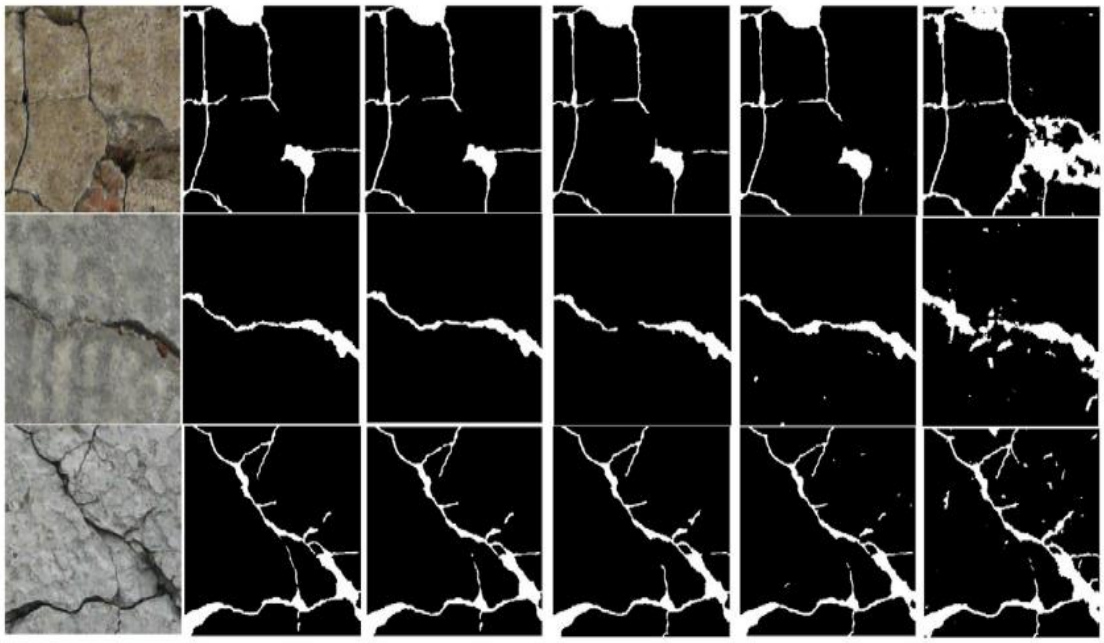

Figure 10. UNet-based concrete crack detection, Zhang et. al 2020 (*26*)**.**

Virtual, Augmented and Mixed Reality

Virtual Reality (VR) is a computer-simulated reality replicating a physical environment or imaginary world through immersive technology. VR replaces the user's physical world with a completely virtual environment, isolating the user's sensory receptors (eyes and ears) from the real world (*27*). VR is observed through a system that displays the objects and allows interaction, thus creating a virtual presence (*28*). VR headsets have gained vast popularity, especially in the gaming industry. On the other hand, augmented Reality (AR) is an integrated technique that often leverages image processing, real-time computing, motion tracking, pattern recognition, image projection and feature extraction. It overlays computer-generated content in the real world. An AR system combines real and virtual objects in a real environment by interactively registering virtual objects with real objects in real-time (*29*). The beginning of AR dates back to Ivan Sutherland's see-through head-mounted display to view 3D virtual objects (*30*).

The initial prototype was only able to render a few small line objects. AR research has recently dramatically increased. Visualizing very complex virtual objects in an augmented environment is now possible. The recent developments of AR and VR technology helped companies produce holographic headsets that benefit Mixed Reality (MR) technology, in which one can experience a hybrid reality where physical and digital objects co-exist and interact in real-time. Mixed Reality was originally introduced in a 1994 paper, "A Taxonomy of Mixed Reality Visual Displays"(*31*). In the paper, a Virtuality Continuum (VC), in other words, a mixed reality spectrum, was explained in detail. A schematic representation is shown in Figure 11. Extended Reality (XR), an umbrella term for VR, AR, and MR technologies, has found many diverse use cases in the civil infrastructures system. A comprehensive literature search is presented by Catbas et al (2022) provide essential information and guidelines for practitioners and researchers on using XR technologies to maintain the integrity and safety of civil structures (*32*).

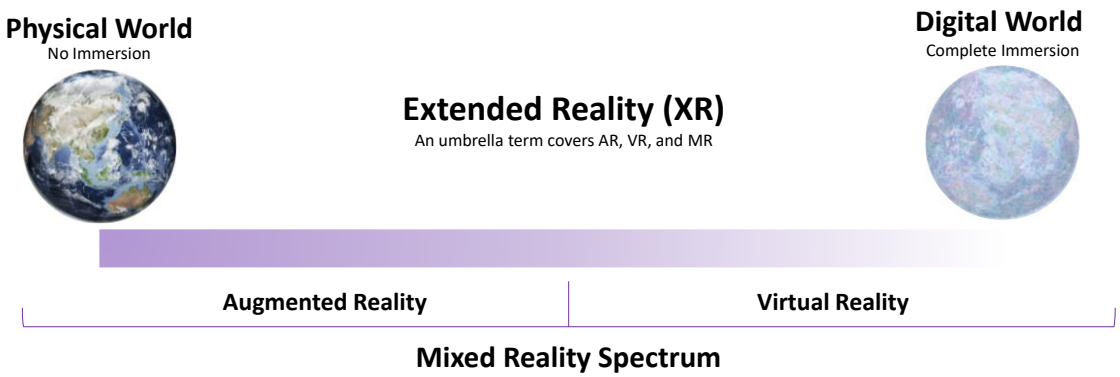

Figure 11. Mixed reality spectrum (*32*)**.**

MR technology has breakthrough applications, especially with successfully deploying 3D user interfaces, such as computer-aided design, radiation therapy, surgical simulation, and data visualization(*33*). The next generation of computer games, mobile devices, and desktop applications also will feature 3D interaction (*34*). There are also other efforts for using MR technology in the construction industry and maintenance operations. Kamat and El-Tawil (2007) discuss the feasibility of using AR to evaluate earthquake-induced building damage(*35*). Behzadan and Kamat (2007) investigated the application of the global positioning system and 3-degree-of-freedom (3-DOF) angular tracking to address the registration problem during interactive visualizations of construction graphics in outdoor AR environments (*36*). Vision-based mobile AR systems are vastly used in 3D reconstructions of scenes in architectural, engineering, construction, and facility management applications. Bae et al. (2013) developed a context-aware AR system that generates 3D reconstructions from a 3D point cloud. Important efforts for the use of AR in infrastructure inspections are also shown by several researchers (*37*).

Researchers at the University of Cambridge collaborated with Microsoft to develop an effective bridge inspection practice in which the data collected from the field is visualized in an MR environment at an office (*38*). Moreu et al. (2017) developed a conceptual design for novel structural inspection tools for structural inspection applications based on the HoloLens (*40*) device (*39*). The experiments conducted with the HoloLens for taking measurements and benchmarking the obtained measurements are shown in the study. Fonnet et al. proposed using MR HMD (Hololens) to overcome difficulties in inspection, data collection, and integration into the BIM model. Brito et al. employed the MR approach to ease the maintenance tasks of inspectors by using HoloLens glasses(*41*). Dang and Shim addressed the challenge of storing damage and repair records and the in-situ structural behavior of bridge structures during inspections. Thus, the paper proposes a real-time Bridge Management System (BMS) that employs a BIM approach in cooperation with a HoloLens device to automate inspection tasks(*42*). The proposed method is applied on an existing cable-supported bridge and demonstrates good potential for enhancing the performance of maintenance activities. Another study addressed the various challenges of visual inspection procedures which cause ambiguity in structural assessments. Therefore, the authors proposed using HoloLens to improve the ability of inspectors during infrastructure inspections. Using HoloLens enables inspectors to make decisions faster and more accurately, assess risks on the site, measure damage growth, and create inspection documentation to store the data collected during the inspection. Maharjan et al. developed and validated a novel human-machine Interface using MR headsets which assistsinspectors in the field during

data collection and decision-making processes (*43*). A more recent study explored the use of a BIM-based MR application where HoloLens is used to improve and facilitate bridge management and potentially used for a defect inspection overlay in the BIM model (Figure 12) (*44*).

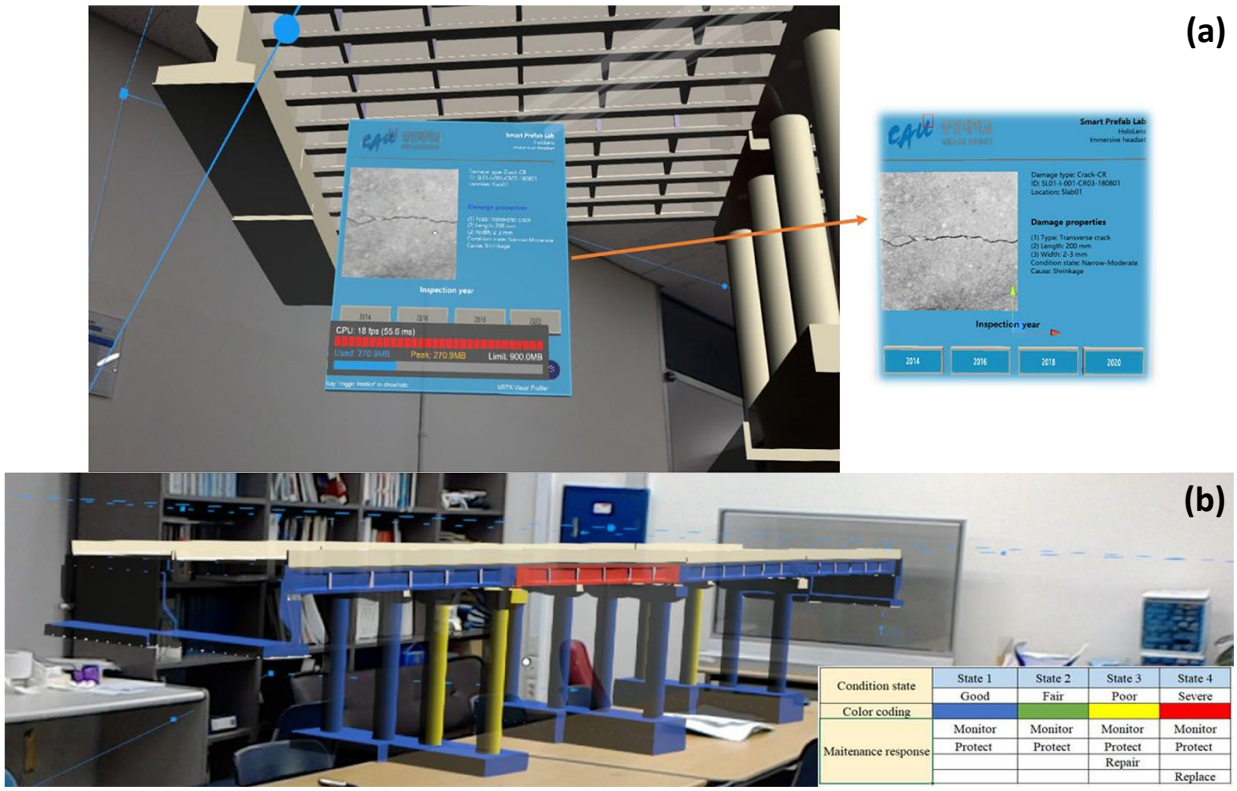

Figure 12. (a) Damage visualization and (b) condition rating visualization of bridge structure in HoloLens (*44*)**.**

Al-Sabbag et al. introduced a study on Human-Machine Collaboration Inspection (HMCI) to allow collaboration between inspectors who wears MR HDM and robots for structural inspection procedures. While the inspector can obtain the meta-information about the defect on the site via MR HDM, the assistant robot gathers that metadata and processes it to an offsite computational server in real time. The workflow of HMCI begins with the robot generating a 3D map of the site, and the spatial coordinates are calibrated with MR HDM (HoloLens 2) to be projected in the HoloLens 2's view. The produced 3D map and pictures are then sent to the server for damage analysis, and subsequently, the damage results are received by the HoloLens2 and overlaid on the real scene where it is visualized in the HoloLens 2 (Figure 13). The proposed study was tested in a laboratory environment. The authors indicated that this study is one of the first works for a human-machine collaborative system integrating robots, inspectors, and MR HDM for bridge inspections. Mixed Reality platform combines the virtual world with the outside environment, where inspectors can continuously interact with AI and other application resulting in a more efficient practice (*45*).

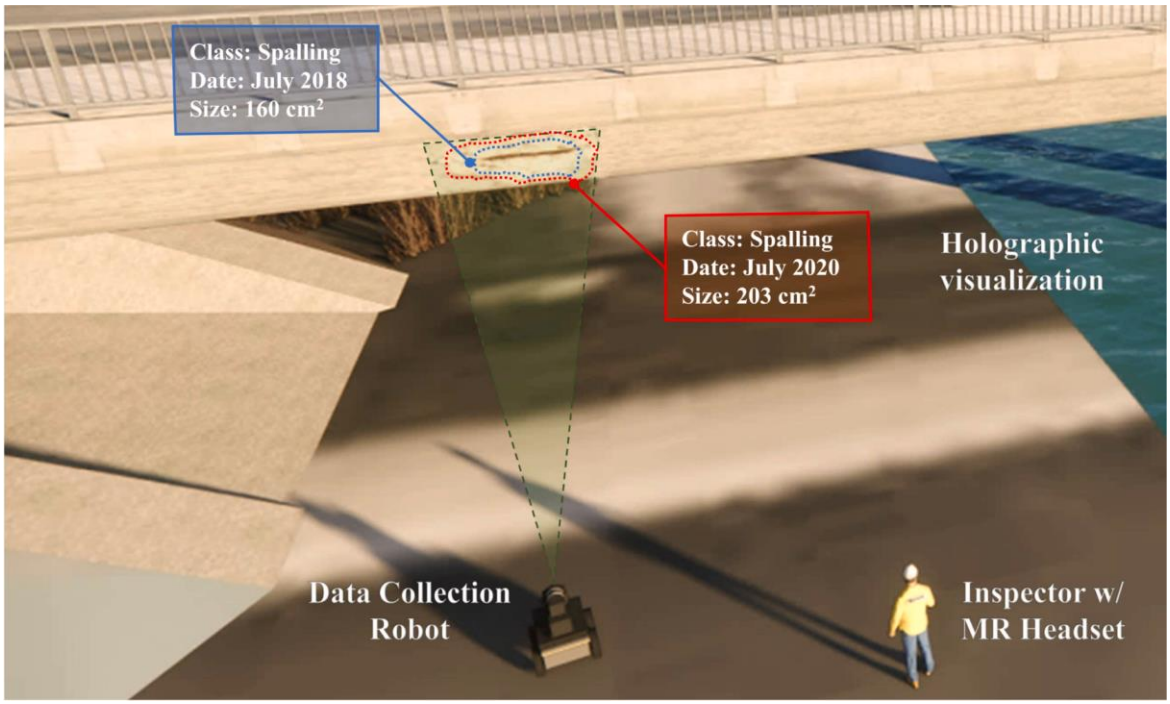

Figure 13. Descriptive visualization of HMCI for bridge inspection (*46*)**.**

The proposed methodology takes goes on to combine AI implementation with MR technology. In this system, the embedded AI architecture predicts the location/region of cracks and spalling on the infrastructure in real-time along with condition information, but also augments the information in the holographic headset for the improved human inspector - AI interaction.

AI-BASED DEFECT ASSESSMENT WITH HUMAN COLLABORATION USING MIXED REALITY

The proposed AI-assisted infrastructure assessment using MR technology employs state-of-the-art methods and algorithms from interdisciplinary practices. Machine learning is vastly used to robustly detect cracks and spalls in infrastructures. In contrast, human-computer interaction concepts are employed for improving assessment performance by including the professional judgment of human inspectors. MR is an excellent platform to maintain this interaction since it augments virtual information into the real environment and allows the user to alter information in real time. In this proposed methodology, the bridge inspector uses an MR headset during a routine infrastructure inspection. While the inspector performs routine inspection tasks, the AI system integrated into the headset continuously guides the inspector and shows possible defect locations. Suppose the human inspector confirms a defect location. In that case, the AI system starts analyzing it by executing defect segmentation and then characterization to determine the specific type of defect. If the defect boundaries need any correction or the segmentation needs to be fine-tuned, the human inspector can intervene and calibrate the analysis. The alterations made by the human inspector (e.g., change of defect boundary, minimum predicted defect probability, etc.) will be used later for retraining the AI model by following a semi-supervised learning approach. Therefore, the accuracy of the AI is improved over time as the inspector corrects the system.

Another advantage of the system is that the inspector can analyze defects in a remote location while reducing the need for access equipment. Even though, in some cases, hands-on access is inevitable (i.e. determining sub-concrete defects), the system can still be effective for quick assessments in remote locations. Suppose the defect location is far or in a hard-to-reach location. In that case, the headset can zoom in and perform assessments without needing any access equipment, such as a snooper truck or ladder. Figure 14 displays the framework of this research.

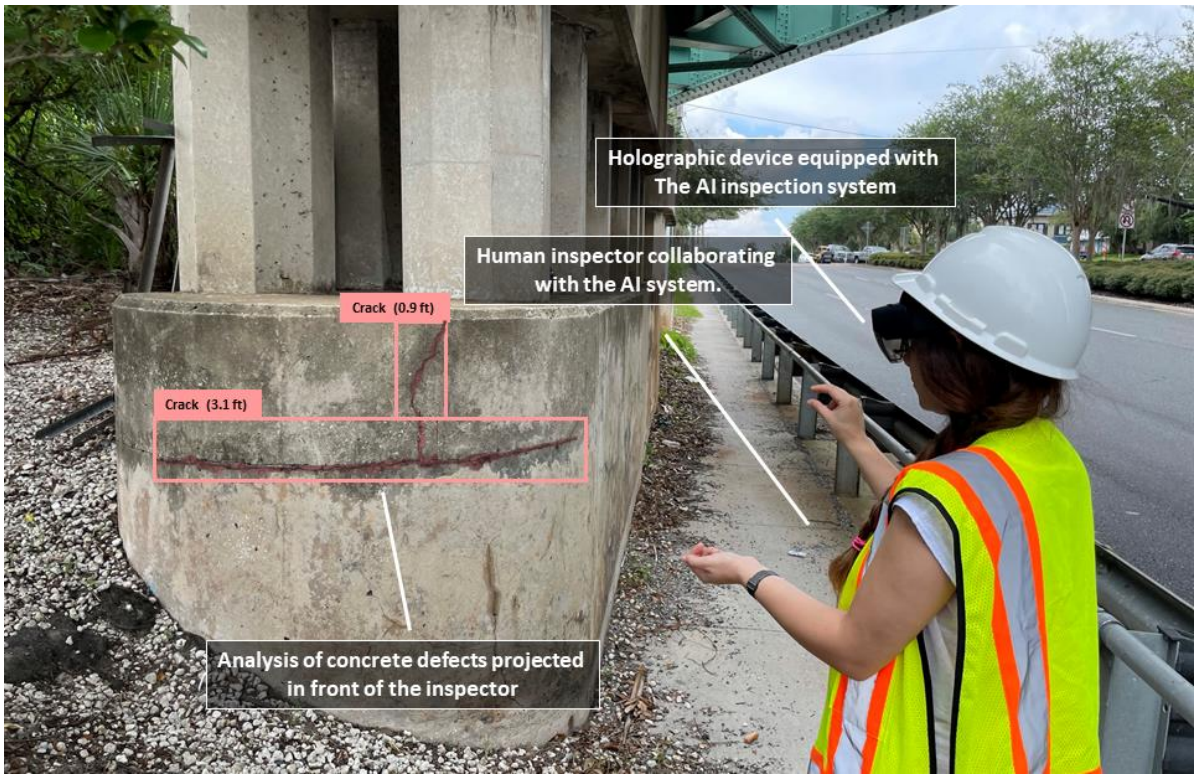

Figure 14. Visual representation of the AI-powered mixed reality system. (The user sees crack detection and segmentation in the headset).

The proposed methodology of AI-assisted infrastructure assessment using MR systems differs from the state-of-practice of current learning-based approaches and mixed reality implementations. Table 1 compares the proposed method with major works in the literature. The major difference of the proposed method from current mixed reality approaches is that the system performs automatic detection and segmentation of defect regions using real-time deep learning operations instead of manually marking the defect regions in the MR platform. In this way, the system can save significant time in defect assessment compared to marking defects in current MR implementations.

Table 1. Comparison of this project with the major literature work.

Ioannis (2017)	Moreu et al. (2017)	Bae et al. (2013)	Xie et al. (2017)	Proposed Method
Remote bridge inspections with HoloLens	Structural inspection and measurement using HoloLens	Mixed reality for structure 3D scene reconstruction	CNN-based crack detection	Mixed reality-assisted bridge condition assessment
Data collection is monitored from a remote location	On-site measurement of defects	Image data is reconstructed after data collection	Aims post-processing of images to identify defects	On-site system to augment bridge inspector performance
Focuses on visualization and post-processing of data	Mainly relies on a human operator obtaining measurements	No detection of defects is implemented; a 3D model is used for inspection	Detection performance relies on AI systems only	Aims to create a collective intelligence with human-AI collaboration
Views high-resolution defect images on a real-size bridge model	Uses 3D projective geometry for measurement estimation	Uses 3D projective geometry to register images	Uses basic data augmentation techniques to increase the training dataset	Uses an extensive data augmentation system that generates many variations of defect images

Defect Characteristics and Data Collection Procedure

Automated detection of defects in concrete structures requires training each defect type individually by processing many training images. First, commonly available infrastructure defects are determined, and their condition assessment procedure is investigated using infrastructure inspection guides (*48–50*). According to reference guides, infrastructure defect types were determined as shown in Figure 15: a. Cracking, b. Rusting, c. Spalling, d. Efflorescence, e. Joint Damage, and f. Delamination (detected by infrared).

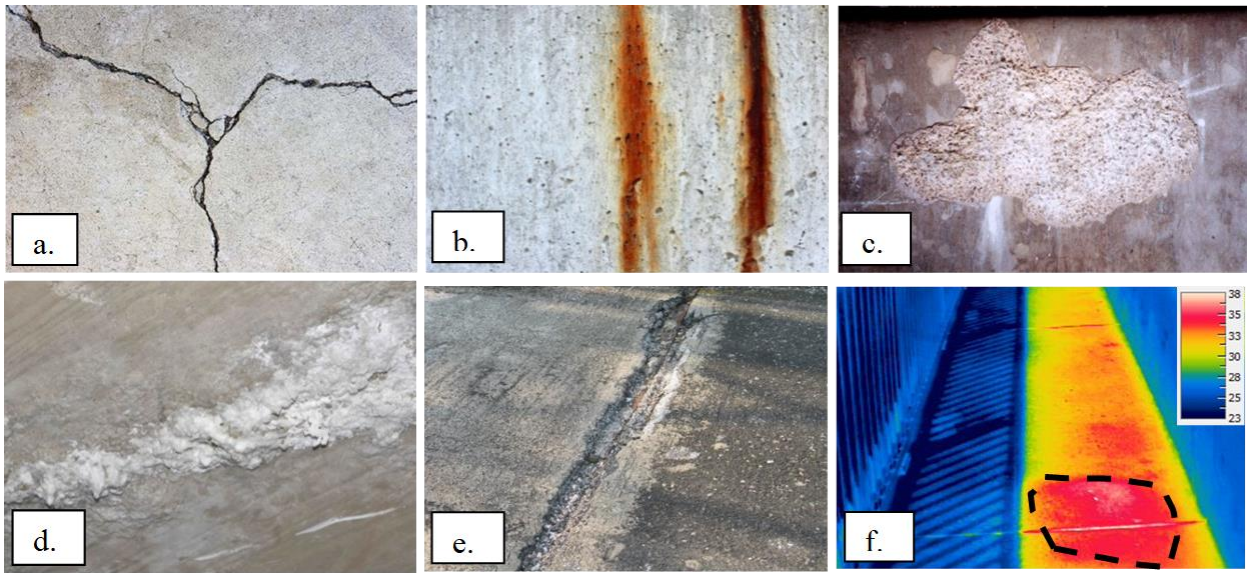

Figure 15. Example defect images for each defect classification.

This study focuses on crack and spalling defects. The defect images are gathered from various sources, including industry partners, transportation agencies, and other academic institutions. Some of the data were only categorized but not annotated; a considerable portion was annotated with bounding box pixel coordinates, and a relatively small dataset was annotated for segmentation. However, extensive data augmentation was applied to the datasets to increase AI prediction accuracy (Figure 16). The data augmentation included rotation, scaling, translation, and Gaussian noise.

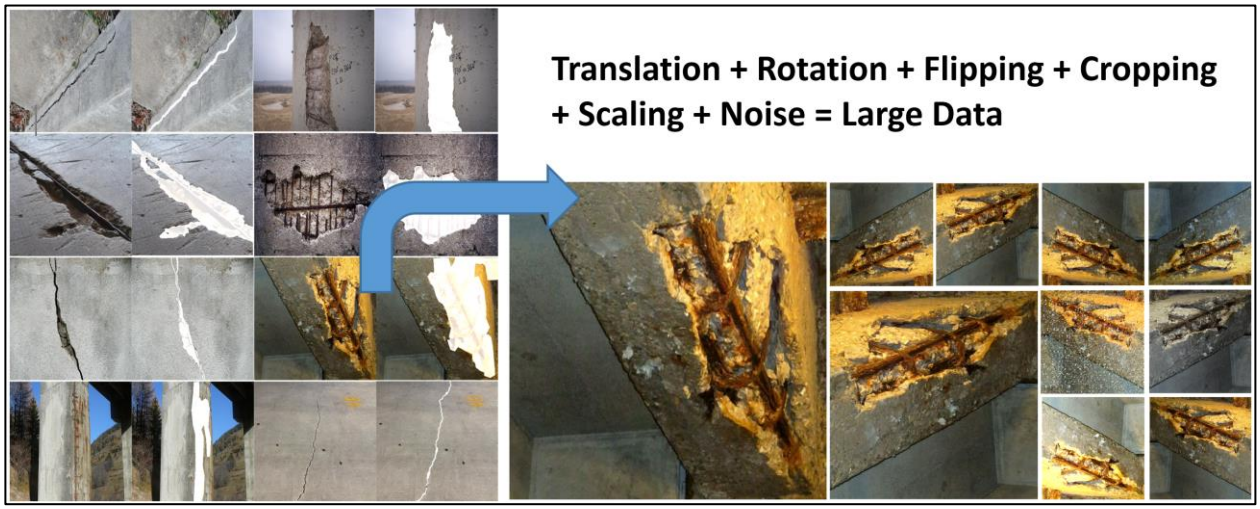

Figure 16**. Annotation of dataset and augmentation with translation, scaling, rotation, and noise.**

The datasets from other studies were sieved into a clean, relevant, and compatible dataset for the deep learning methodology described here. The available data was split into 70% training, 15% validation, and 15% test data. The test data was never used during training to validate the performance metrics; therefore,

the models would have no familiarity with the test dataset before the final evaluation. The datasets used in this project were changed and cleaned up multiple times. The reason behind this is that many of the available images from the defects are either cropped, such as images in SDNET2018, or too close to the defect, as in CODEBRM, and therefore they do not provide the diversity required for the training. Moreover, the publicly available crack datasets for segmentation are much greater than those for spalling. However, for the model to perform well, the annotations must be uniform. For instance, if an object detection model is trained with 20,000 instances of crack and 500 instances of spalling, the weights in the CNN model will be adjusted based on the cracks, and spalling will be more likely to remain undetected. Figure 17 displays some examples of the data instances removed from the dataset.

Figure 17. Images not suitable for defect localization training due to lack of diversity.

As previously mentioned, SDNET2018 has over 50,000 cropped images of cracks. While those images are not suitable for defect localization training, they are suitable for training defect quantification. These images, however, are not annotated for segmentation models. The dataset was initially generated to classify concrete surfaces: as damaged or intact. Since the annotation process for defect quantification is very labor-intensive, a semi-supervised approach was used instead to label these defects. While some relatively good annotation results were obtained, only a few images of the dataset were used in the model. The images in SDNET2018 are all 256x256 in size and very uniform. Many of these images result in a biased dataset, which results in an inaccurate model. Even with a single class of damage (crack), the size and quality of these images cause uniformity in the dataset, which is undesirable since the same uniformity does not exist over all the datasets. Therefore, only around 20 images of this dataset were used in the segmentation dataset. Figure 18 displays some images from the dataset. The authors also gathered 300 images from the NASA Causeway bridge for defect localisation and added them to the existing dataset. The images are taken from girders and below the deck. They have diverse backgrounds making them suitable for the defect localization model. Ideally, more images must be collected from different bridges, with different materials and

environments. The team also visited some other bridge sites and collected data. The final dataset includes 1500 images of cracks and spalling.

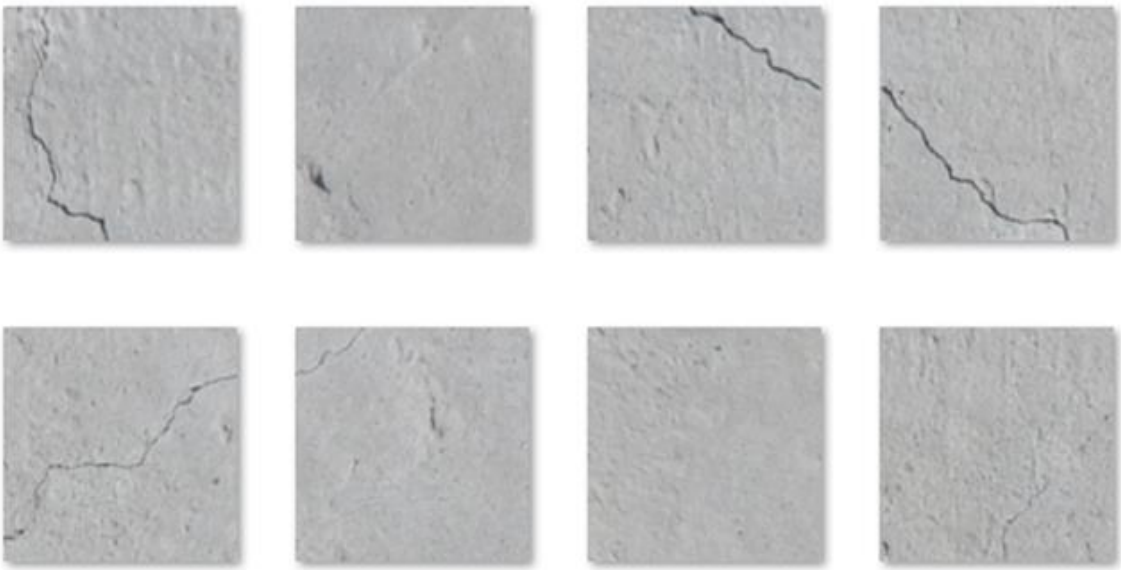

Figure 18. Some of the images from SDNET 2018

The open-source dataset used to segment the defects was also modified to better represent the inspection process using an MR device. To do so, the defects were cropped from the images to support the attention-guided procedure in defect segmentation. Figure 19 displays an example of the prepared dataset for segmentation.

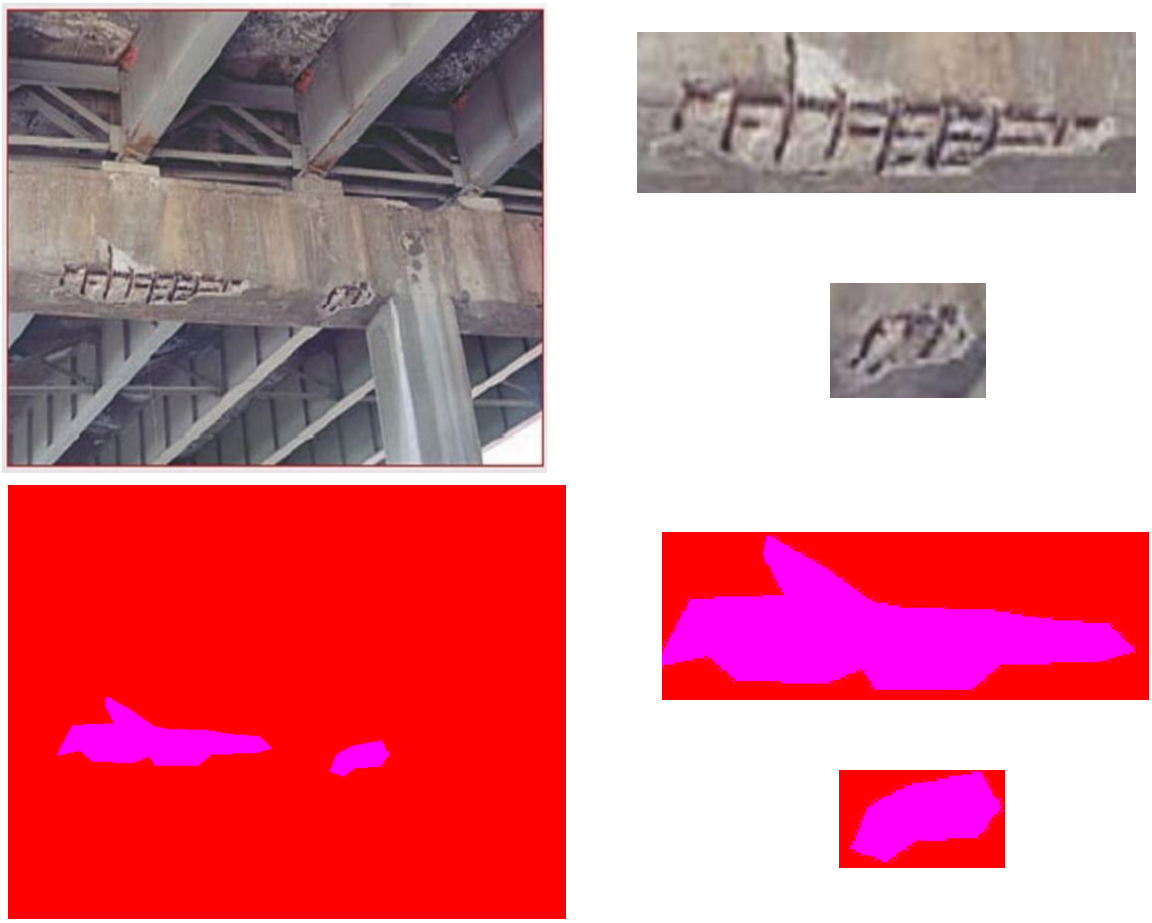

Figure 19. Original Images and masks vs. Cropped images with AI models for Attention guided segmentation.

Human Centered Semi Supervised learning

The proposed deep learning methodology for concrete defect analysis is designed for human-computer interaction environments such as wearable holographic headsets and handheld mixed reality (MR) devices. Using technologies that integrate the proposed methods, an inspector can continuously communicate with the AI system. The human-computer interaction in MR will entail practical human-AI collaboration to create collective intelligence. The AI models for damage detection and segmentation in the proposed methodology will allow the inspector to adjust the prediction threshold values, model inference parameters, and even the attention regions in real-time. This human-centered system can easily outperform a fully automated robotic technique (*51*). Similar systems are commonly seen in automated vehicle technologies, visualization systems in the health industry, and video game engines. The semi-supervised approach referred to in this methodology consists of generating fine-tuned data by using inspector adjustments in detection boxes and segmentation regions during routine inspection performance. The inspector only provides minimal input to the system during the inspection, and the annotated training data is automatically

generated. The system periodically schedules fine-tuning in the cloud and optimizes the weights of the last six convolutional layers. The fine-tuned model weights are automatically updated in the deployed device. The AI framework thereby improves its prediction accuracy as the inspector uses the device without doing any data preparation. Figure 20 describes the overall system.

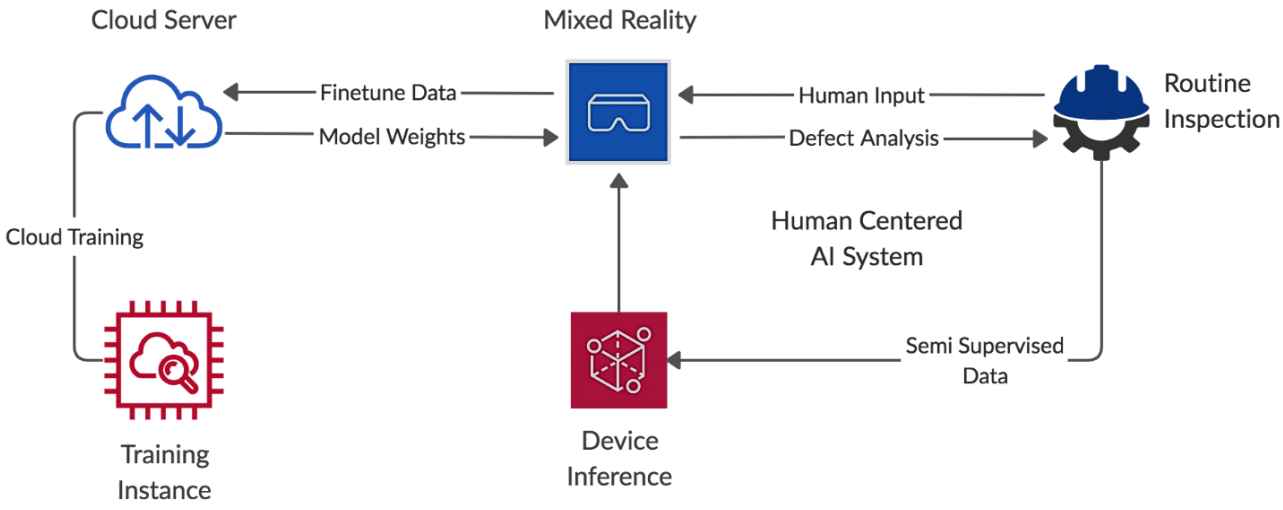

Figure 20. The interaction diagram of the system components in the proposed methodology.

During a bridge inspection, asking a human inspector to modify prediction thresholds will help improve the detection accuracy and determine the segmentation's boundary region. In Figure 21, real-time damage detection did not show one of the spall regions to the inspector when the prediction threshold was set to 0.8; when the inspector adjusted the value to 0.7, the missing spall region was also detected. (The value represents the probability of an accurate prediction.)

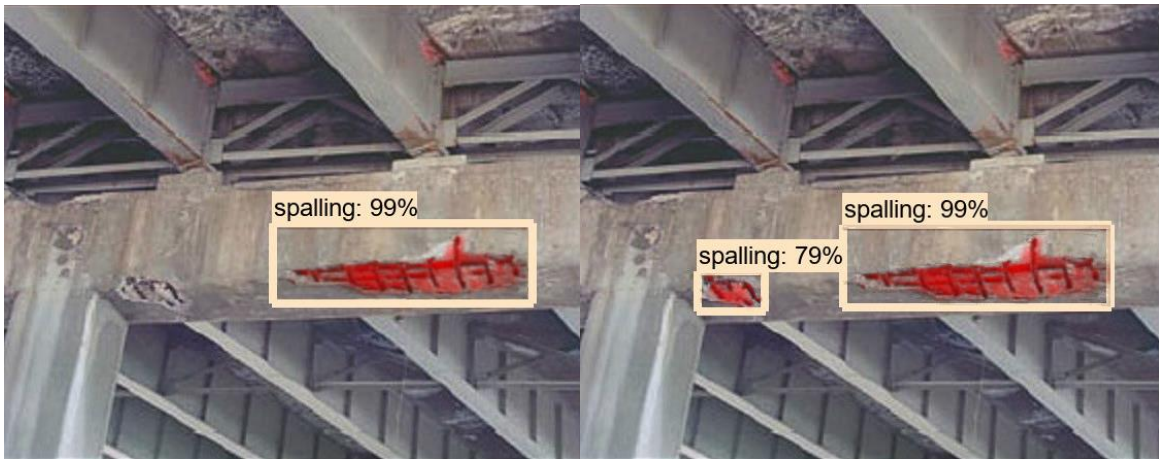

Figure 21. Example of human-AI collaboration in the proposed methodology (detection AI alone on the left misses a spall, while human-assisted AI detects all spalls on the right with threshold adjustment by the inspector).

CONVOLUTIONAL NEURAL NETWORKS

The machine learning system introduced in this study comprises a multi-step process for localizing and quantifying surface defects. First, the defect localization model locates the surface defects in the inspector's view, creating an attention region for the subsequent defect quantification model. After the inspector verifies the localized defects, the defects are quantified and measured for condition assessment (e.g., maximum crack width or area of spalling). A deep convolutional neural network (CNN) was trained for defect localization. CNNs are widely used in object detection tasks, including autonomous driving to locate road objects (*52*) and with medical imaging for early detection of abnormalities (*53*). Another type of CNN model widely used for semantic segmentation tasks was trained for quantifying surface defects. Similarly, semantic segmentation models have shown reliable results in other fields, such as identifying salient elements in medical scans (*54*) or finding lane lines in autonomous driving (*52*). CNNs have also already shown promising results in classifying and detecting concrete surface defects (*26*, *55*, *56*).

CNNs work well with two-dimensional data and are great tools for analyzing images and videos. These models typically comprise convolution, pooling, and activation layers to extract features, reduce dimensions for efficient computation, and introduce nonlinearity (*13*). In the convolutional layers, the input images are multiplied by small distinct feature matrices called kernels, and their summations are normalized by matrix size (i.e., kernel size). By convolving images, similarity scores between every region of the image are assigned to generate the image feature matrix. After convolution, the negative values of similarity in the image matrix are removed in the activation layer. A pooling matrix is used to reduce the size of the resultant matrix. For classification, the resultant matrix is then passed through a fully connected layer to obtain class scores. Finally, the image vectors of the trained images are compared with the input images, and a correspondence score, the highest score value, will indicate the classified label. A typical CNN architecture is shown in Figure 22.

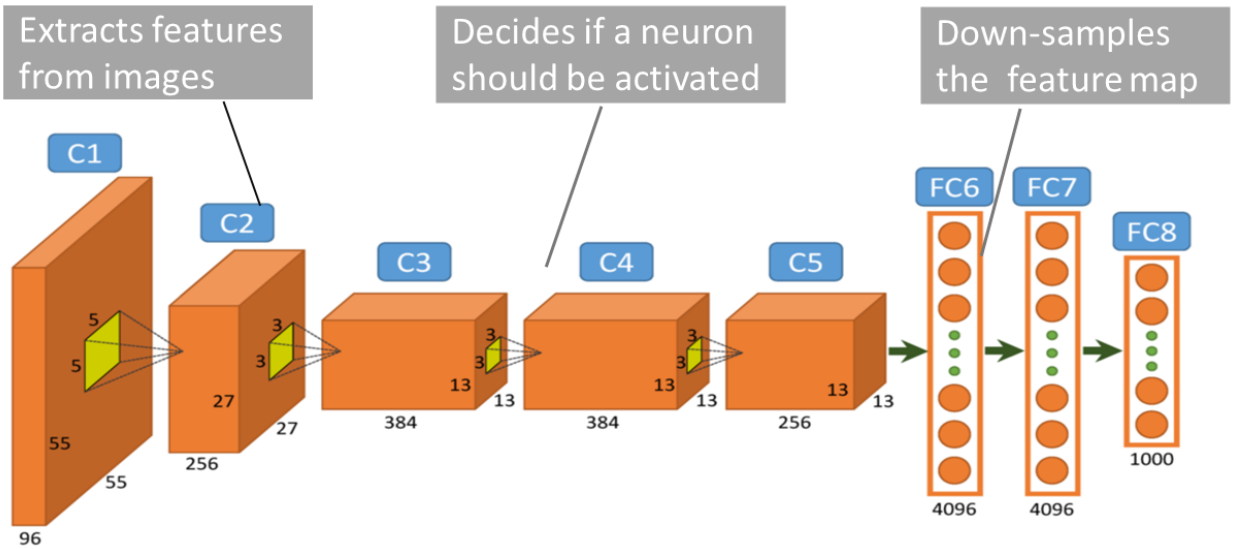

Figure 22. A commonly used CNN architecture, – AlexNet (*57*)

There are critical challenges when effectively training a deep learning model. One major challenge is overfitting. Overfitting usually occurs when the data is either too small when compared to the size of the neural network architecture or when the data is large but not diverse enough. If a dataset is too small, more data can be added to the training using data augmentation techniques, or the complexity of the network architecture may be reduced (e.g. decreasing the convolutional layers). If the dataset is large but not diverse enough, regularization may be added to the network (e.g. adding dropout layers or L1/L2 regularization) (*58*). Early stopping is another way to tackle the overfitting problem.

Evaluation of CNN Models

The performances of the damage detection and segmentation AI models were evaluated using the accepted evaluation procedures used in the literature. The evaluations were carried out on only the AI models without the human-centric framework in which the human inspector assists the AI with minimal input. Therefore, real-life performance results from the human-AI collaboration are expected to be superior to the evaluation results of the individual AI models in this study.

When evaluating machine-learning models, it is common to classify predictions into four categories: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). However, evaluating object detection segmentation models will require additional metrics to measure the accuracy of the detected or segmented areas. Mean Average Precision (mAP) is a performance indicator that finds the average of maximum precisions at different recall values based on a confidence threshold. Average precision (F1 Score) is calculated as follows:

$$Presicion = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$Average\ Precision\ (F1) = 2 \times \frac{precicion \times recall}{precision + recall} \tag{3}$$

To calculate precision and recall, TP, FP, and FN need to be determined from evaluation metrics. One of the common metrics is the Intersection over Union (IoU). IoU, also known as Jaccard Coefficient, was first introduced by Jaccard (1912). IoU measures region overlap without concentrating on true boundaries; hence, IoU is the preferred method for detection measures, not for segmentation. This metric is the ratio between the intersection and the union of the predicted boxes and the ground truth boxes.

$$IoU = \frac{area\ of\ overlap}{area\ of\ union} = \frac{TP}{TP + FP + FN} \tag{4}$$

A similar approach is the Dice Similarity Coefficient (DSC), first proposed by Dice (1945). DSC is reliable and the most commonly used metric for delineation accuracy. It measures regional overlap in the boundary of segmented and ground truth contours. DSC was calculated by Equation (5).

$$DSC = \frac{2TP}{2TP + FP + FN} \quad \textbf{(5)}$$

Hyperparameters in Training CNN models

Another cause of poor training performance is the wrong selection of training hyperparameters. A small learning rate is often chosen to reduce training loss. However, validation loss then becomes much larger. This indicates that the model fits over the branch of the loss function and does not converge around the local minima (*61*). Increasing the learning rate, on the other hand, yields the nonconverging loss function. The effect of hyperparameters was investigated during a preliminary study by the authors. The summary of the best values is stated in Table 2. The comparisons were conducted using SSD: Single Shot MultiBox Detector and SegNet.

Table 2. Hyperparameter evaluation

Hyper Parameters	Detection Model	Segmentation Model
Learning rate scheduling	*Initial Epochs<100 Epochs<200* 0.001 0.0001 0.00001	*Initial Epochs<25 Epochs<50* 0.01 0.001 0.0001
Model Optimizer	Adam Optimizer ($\beta=0.9, \varepsilon=10^{-8}$)	RMSprop Optimizer
Batch Size	Training = 32, Evaluation = 4	Training = 16, Evaluation = 4
Validation Loss	Categorical Cross-entropy	Binary Cross-entropy
Regularization	Dropout, L2 Regularization	Batch Normalization
Monitoring	Early Stopping	Early Stopping

\

The effect of batch size selection on training performance was also investigated during trainings. Even though some studies indicated that large batch sizes (more than 64) in a Stochastic Gradient Descent based model might impact training performance by causing less gradient noise in training, hence leading to poor generalization behavior (*62*), a contrary result was observed when comparing batch sizes smaller than 32. During the training of a limited dataset, when very small batches were fed through the network, the network failed to provide a stable enough estimate of the maximum gradient of the full dataset when averaging the gradients of small batches (*63*). The comparison of different batch sizes in terms of total loss values (the summation of classification loss and localization loss) for the damage detection model is shown in Figure 23.

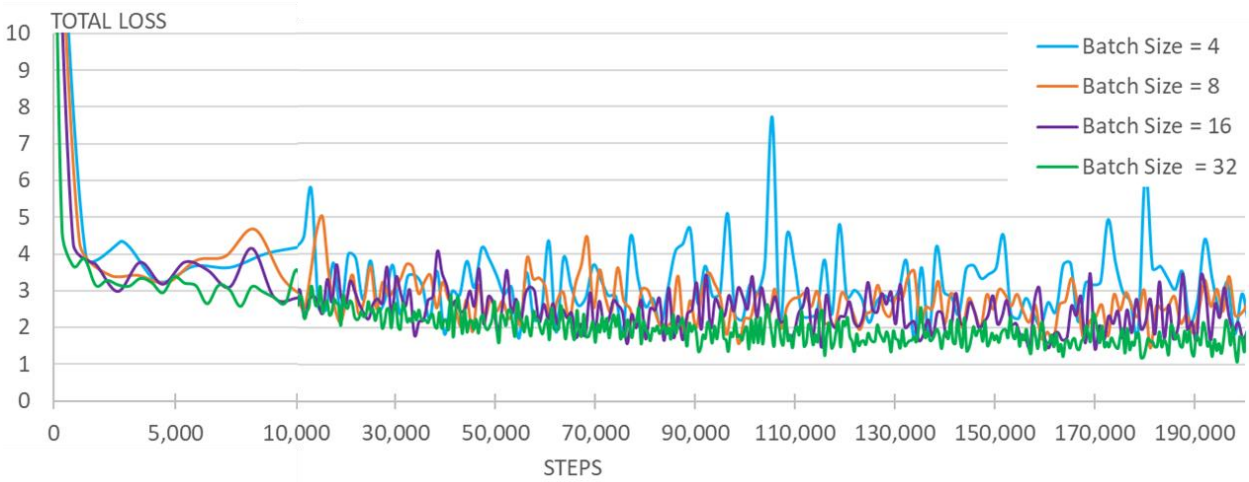

Figure 23. Effect of batch size selection on the training performance (the detection model is used for reference, and the steps are shown four times more frequently after the 10,000th step).

The selection of the model optimizer is also very important. Model optimizers search for local minima and the maxima points of the training model's cost function. Commonly used optimizers are the Adam Optimizer, Stochastic Gradient Descent (SGD), and RMSprop Optimizer (*64*). All optimizers were experimented on both models, and the best performance was observed when the Adam Optimizer was used in the detection model, and the RMSprop Optimizer was used in the segmentation model. The training performance comparison of these model optimizers on this dataset is shown in Table 3. For damage detection, minimum obtained classification and localization loss values were used; for damage segmentation, minimum obtained Dice loss values were used as comparison metrics.

Table 3. Comparison of model optimizers

	Adam Optimizer	Stochastic Gradient Descent (SGD)	RMSprop Optimizer
Damage Detection[1]	classification loss = 0.69 localization loss = 0.48	classification loss = 0.92 localization loss = 0.64	classification loss = 0.78 localization loss = 0.57
Damage Segmentation[2]	Dice loss = 0.55	Training Failed	Dice loss = 0.31

[1]Evaluation loss values obtained at the end of the 193rd epoch (114,835 training steps)
[2]Evaluation loss calculated by Dice loss function obtained at the end of 112th epoch (33,936 training steps)

As shown in Table 3, the Adam Optimizer yielded the lowest evaluation loss values when the concrete defect dataset was fully trained on the object detection model. The Adam Optimizer uses exponentially weighted averages just like RMSprop, but also integrates the idea of momentum optimization; thus, it converges faster and maintains stability (*65*). However, the Adam Optimizer will not converge to an optimal solution for more complex networks, as opposed to RMSprop, as shown in the segmentation model's results despite the long training duration. The Stochastic Gradient Descent (SGD) performed poorly in both models and even failed to train the segmentation model. SGD uses random searches to escape the local minima/maxima points but sometimes causes major spikes when converging the cost function (*66*).

Attention-Guided Damage Segmentation

For concrete defect assessment, it is not solely enough to detect the damage in a bounding box; the damage also needs to be segmented from entire regions to quantify necessary defect measurements for understanding the extent of the defects. Unlike the general practice available, defect localization and quantification in this study are conducted by two different AI models. Initially, a CNN object detection model localizes the defects. Then, a complementary deep learning model segments the damaged regions. As a unique approach for damage segmentation, an attention guidance approach was used in this research, as previously investigated by the authors (*67*). A sequential connection was created between detection and segmentation models. First, images were fed into the damage detection pipeline, and after the human inspector verified the bounding box, damage segmentation was executed only for the region inside the detected bounding box. This approach significantly improved the accuracy of segmentation and successfully prevented outliers. Figure 24 shows qualitatively how attention-guided segmentation was superior to segmentation without attention guidance. On the left image in the figure, the model performs only pixel-wise segmentation operations to find the damaged regions and subtracts them from the background. On the right image, the model first performs detection and then immediately inputs the detection results into the segmentation pipeline

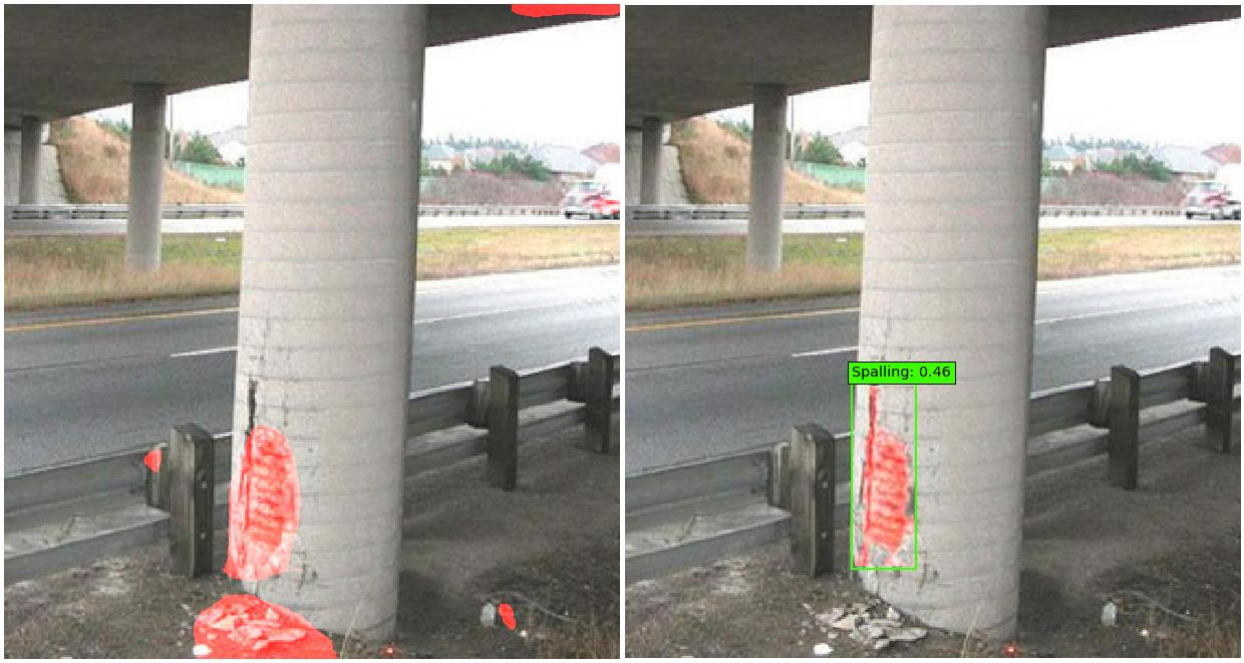

Figure 24. The effectiveness of attention-guided segmentation is shown in red highlighted areas (Left: Segmentation resulted in some false positive results; Right: Attention guidance readily removes misclassified pixels).

The attention-guided approach reduces the need for diversity for the segmentation model since the bounding boxes mostly cover the defect area.

The concrete damage detection and segmentation were evaluated using IoU and DSC metrics in mAP calculations. In addition to the mean average precision calculation, the average prediction speed (in milliseconds) was also monitored during the evaluations. The results are shown in Table 4.

Table 4: Performance Evaluation of the Segmentation Model

	mAP using IoU	mAP using DSC	Speed (ms)
Defect detection using SSD with VGG16 backbone	0.74	-	0.17
Segmentation with no attention guide (*SegNet* only)	-	0.64	0.68
Segmentation with attention guide (SSD + *SegNet*)	-	0.88	0.72

According to the evaluation results, the damage detection model could predict the damage boundaries with 74% mean average precision using the IoU metric and 85% precision using the DSC metric. The segmentation model, on the other hand, showed significant improvement when the detected boundaries were used as the initial region of interest. Segmentation without attention guides predicts the damaged

regions in 52% mAP using the IoU metric and 68% mAP using the DSC metric. When the segmentation model was coupled with the detection model (attention-guided segmentation), the precision increased up to 79% with the IoU metric and 88% with the DSC metric. As for inference speeds, the models were tested on the Pascal GPU, a mobile chipset commonly used by mixed-reality devices. The damage detections were performed at 0.17 milliseconds per frame. The segmentation could be predicted in under a second. The segmentation model operated at 0.68ms average speed when executed alone on the input image. However, it operated 13ms faster when the attention guide provided an initial region of interest (attention). Therefore, the sacrifice of the computational speed was minimal when damage detection and segmentation were coupled (only 0.04ms).

The evaluation results clearly compare the Segmentation Only operation and *Sequential Operation* (Detection + Segmentation). The sequential pipeline resulted in approximately 30% higher precision in subtracting the spall and crack regions than the segmentation pipeline alone, with only a slight sacrifice in computation time.

Evaluation of Different CNN Models for Defect Localization

The first step towards building AI models is data collection. First, concrete defect images were gathered from public datasets and through image scrapping from internet image search engines. Some real-world images were also collected from several bridge structures the team had visited for inspection. During data preparation, a significant portion of the images was filtered out as they were cropped too closely and unsuitable for defect localization tasks. They were kept for defect quantification as they can be labeled using a semi-supervised learning approach. The data labeling process for defect localization includes manually annotating the defect images by drawing bounding boxes around the defects. These boxes serve as ground truth during training. Each annotation consists of the defect type and the pixel location of the box corners for each training image. For data labelling, an open-source image annotation tool called LabelMe was used. 8 shows examples of bounding box annotation of training images. Aside from the open-source datasets, 400 images were collected from the Indiana River Bridge. The images were obtained during a thorough inspection and labelled using AI-assisted programs Roboflow and LabelImg (Figure 25). The process of data cleaning in this study is important. Although many images are available, only those with diverse backgrounds were useful, as they represented the area from the view of an inspector wearing an MR headset. A total of 1500 images were gathered and used for defect localization training.

Figure 25. Image Annotation of data collected from NASA Causeway Bridge in LabelImg.

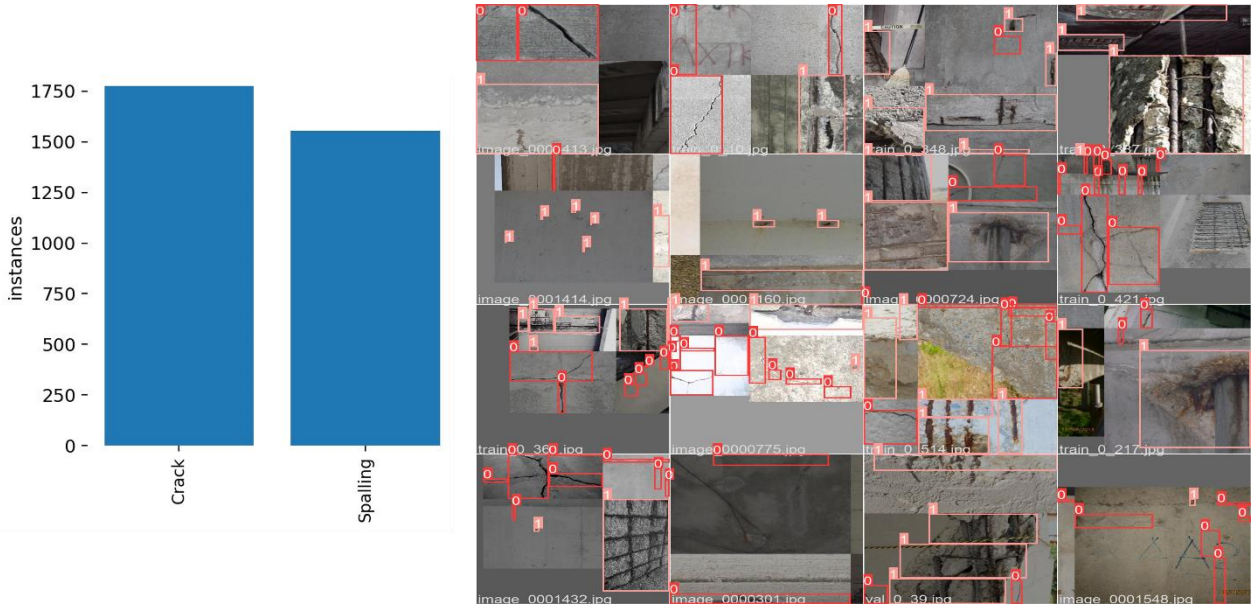

Figure 26**. A sample Training batch and the class distributions for the defect localization model.**

The number of instances in the dataset should also be uniform to obtain the best-trained model. If one instance has many more elements than another, the model does not learn the defects properly and will be biased. Therefore, the dataset was cleaned and preprocessed to make the number of instances as similar as possible. (Figure 26)

The authors tested and compared multiple publicly available object detection architectures to select the best-performing CNN algorithm for defect localization tasks. In order to facilitate real-time processing of the defects, the selected object detection model needs to satisfy two other criteria in addition to high

detection accuracy: high inference speed and small memory footprint (*6*). Several state-of-the-art CNN architectures, including SSD-MobileNetV3 (*1*), SSD-MobileDet (*2*), EfficientDet-D0 (*3*), YOLOv4-Tiny (*4*), and YOLOv5s (*5*), were tested for mean precision accuracy, inference speed, and quantization capability for edge computing devices.

SSD-MobileNetV3 and SSD-MobileDet have a lightweight architecture suitable for edge computation and therefore were selected for evaluation in this study. However, none of the models demonstrated the desired accuracy during the test on real-world images collected from a bridge structure. mPA values were also very low, and many instances of cracks and spalling remained undetected (Table 1). YOLOv3 is another strong CNN architecture which is also popular in autonomous driving (*68*) and has also been used in the localization of defects with good accuracy (*26*); however, the model architecture is large and requires more computational power, which makes it unsuitable for the real-time performance on targeted edge devices in this study. Moreover, the dataset in this study is relatively small for training YOLOv3. YOLOv4-tiny is another version of YOLO models that is more suitable for smaller datasets; however, the model requires a large memory size compared to other CNN architectures and therefore was not selected for this study. The comparison between the models is made in Table 5. On the other hand, the training results of EfficientNetD0 were consistently overfitted. This was observed by following the training graph and evaluating the model on test images. Figure 27 displays the loss values from the training and validation sets. Overfitting is observed by comparing the loss function between the training and validation datasets. The loss value, which represents the summation of errors in a model, usually decreases with each epoch. If the loss decreases only on the training set but remains the same on the validation set or increases, then most likely overfitting occurred.

Table 5 Comparison between object detection models

Meta Architecture	mAP (IoU)	Speed (ms)	Speed (FPS)
SSD MobileDet	0.18	0.053	19
SSD MobileNet V3	0.28	0.050	20
YoloV4-tiny	0.37	0.043	23
EfficientNetD0	--	0.250	25
YoloV5s	0.51	0.143	30

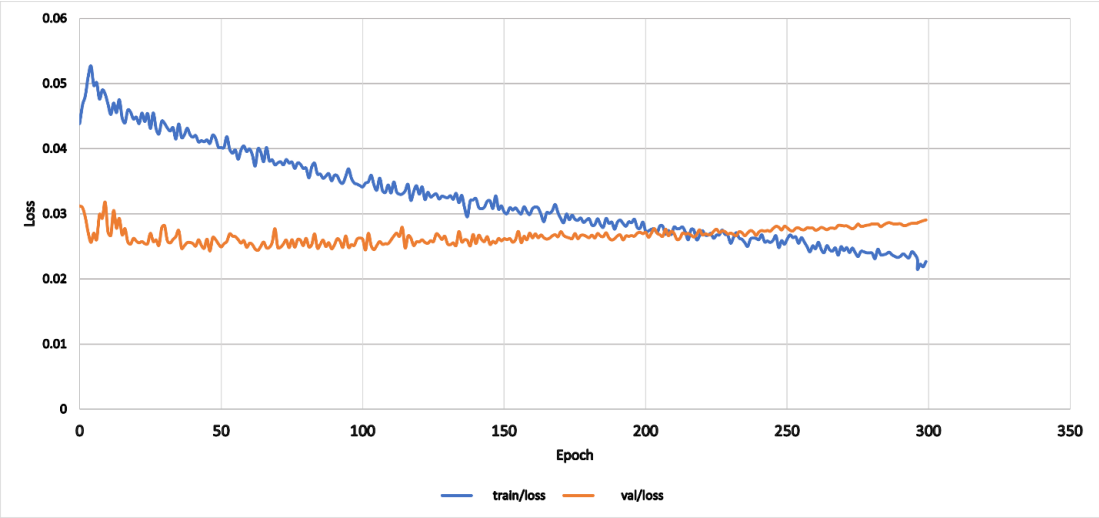

Figure 27. EfficientNeTd0- Loss values per epoch

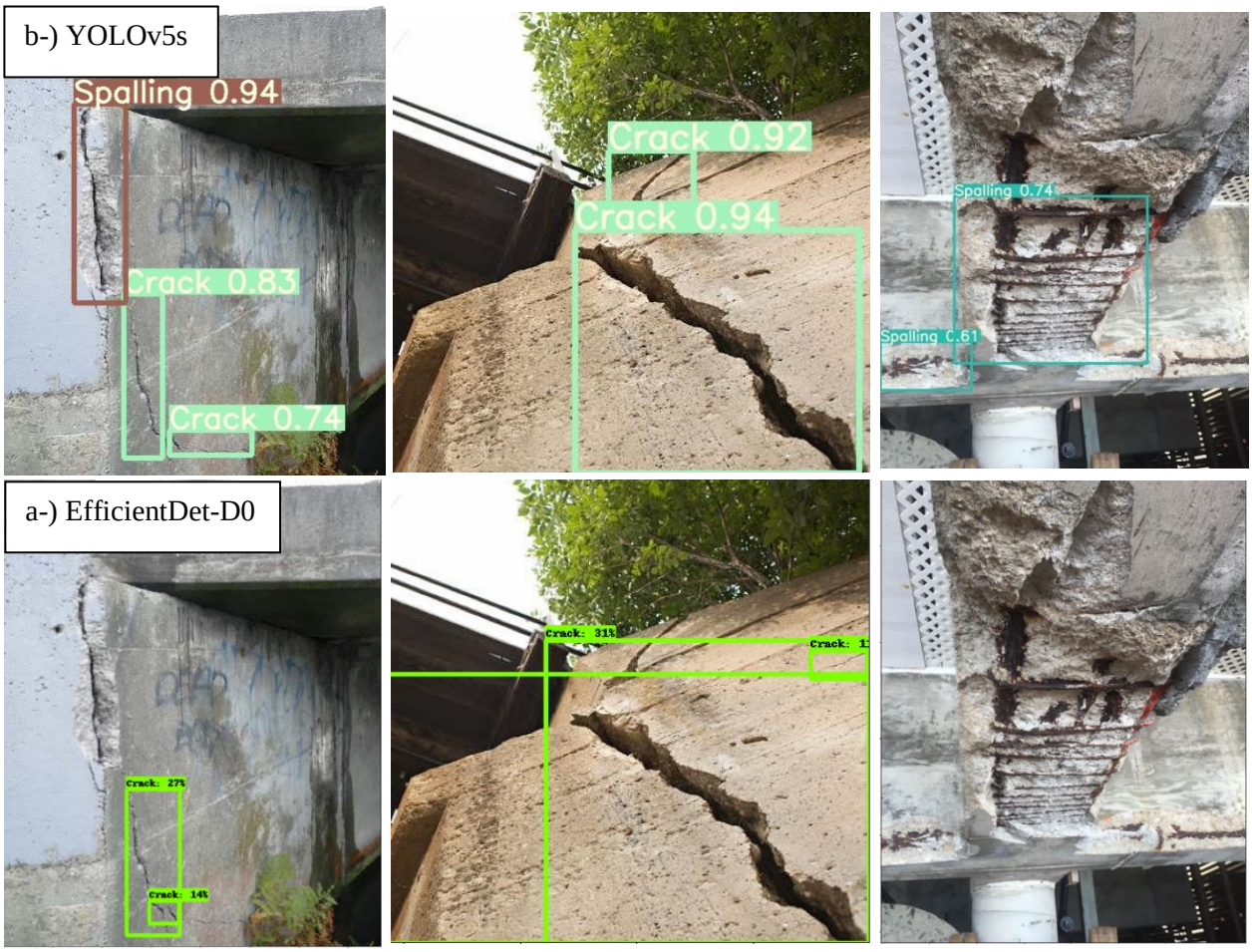

Figure 28. The training loss graphs of the EfficientDet-D0 model and tests on real-world images

Yolov5 is another powerful object detection model with high inference speed and a small memory footprint. This object detection model also supports different sizes of datasets. Yolov5S (small) was trained and evaluated in this study. This algorithm also allows for the optimization of its hyperparameters for better accuracy. Initially, pre-determined hyperparameters from a previously trained model on the COCO dataset were used using transfer learning. The hyperparameters were adjusted for the surface defect dataset using a generic optimisation algorithm. To improve the mAP values, a weighted combinationof metrics was used for optimization: mAP@0.5 contributes 10% of the weight and mAP@0.5:0.95 contributes the remaining 90%. After training for ten epochs, the new optimized hyperparameters were adjusted to train the defect quantification model. Unlike EfficientNetD0, the loss values in this model dropped consistently after each epoch for both the training and validation set. The final mAP value reached 0.65. After approximately 400 epochs, the model stopped learning and reached its highest performance. Figure 28 shows some real-life comparisons between EfficientDet0 and Yolov5. Figure 29 displays the training graphs of YOLOv5s object detection, and figure 30 shows the performance of Yolov5 on the test dataset.

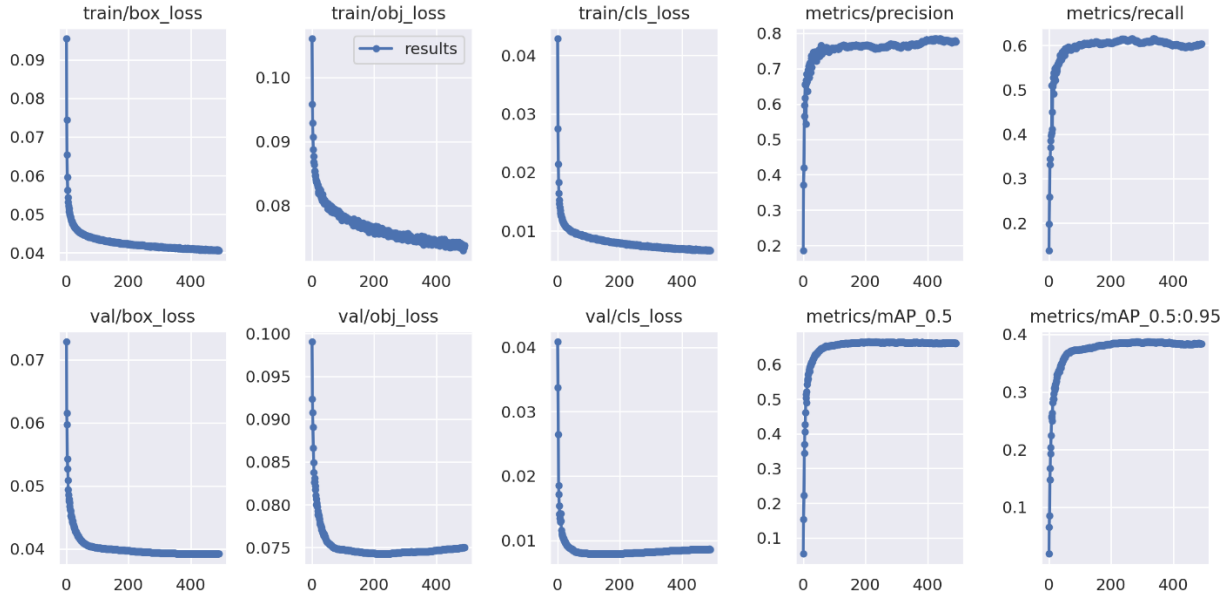

Figure 29**. Yolov5 object detection training metrics.**

Figure 30**. Yolov5 performance on concrete bridge defects in various environments.**

Evaluation of Different CNN Models For Defect Quantification

For concrete defect assessment, it is not enough to detect the damage in a bounding box; the area also needs to be segmented from the intact regions to perform quantification, including necessary measurements for understanding the extent of the defects. Therefore, another deep learning model is used sequentially with the defect localization model YoloV5 to segment defect regions. The segmentation model classifies each pixel of the cropped image in the bounding box as “Damage” or “No Damage”. Segmentation CNNs in this study consist of an encoder and a decoder section. The encoder is a classification network that conducts the classification, and the decoder semantically projects the features of each class determined by the encoder on the pixel area conducting segmentation. Initially, multiple segmentation models were investigated for their accuracy and inference speed. Popular segmentation models such as FCN (*69*), UNet (*70*), SegNet (*71*), and SegCaps (*72*) were previously investigated by Karaaslan et al. (2021) (*73*). The study showed that model size (i.e., memory allocation) is the main challenge in deploying these models in edge computing devices. Hence, the authors only evaluated segmentation models with lightweight classifier backbones. Twelve different CNN architectures with different classification backbones were selected for

training (4 model architectures with three classifier backbone options). UNet, LinkNet, FPN, and PSPNet segmentation architectures with the backbones of Efficientnetb0, Densenet121, and Inceptionv3 were trained and tested. Figure 31 displays the UNet architecture and an example of its output after the semantic segmentation.

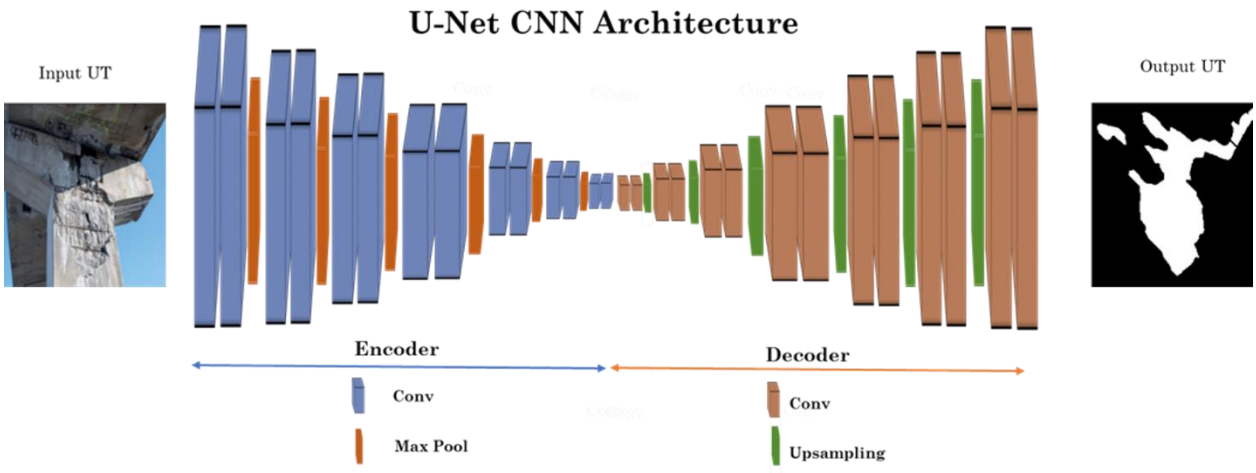

Figure 31**. UNet Architecture for semantic segmentation**

Annotation for the segmentation models is highly sensitive and time-consuming. Each pixel on the defect area needs tobe annotated separately. Therefore, an open-source annotated dataset was initially used for training. Prior to the training, the dataset was improved using data augmentation. This method artificially increases the dataset by modifying the images, including cropping, rotation, or adding noise or blur. The images were also resized based on the requirements for each segmentation model. The method improves the model's accuracy by forming new images for training. Similar to defect localization training, transfer learning was also used to train the defect quantification model. In order to improve the training results, a pre-trained weight from ImageNet (*74*) was used for the encoder (transfer learning). Initially, the model was trained using only ImageNet encoder weight and the decoder from scratch; however, satisfactory results were not obtained in the first run. Therefore, instead of training the decoder from scratch, its layers were frozen for the first few epochs and then unfrozen to complete the training. Consequently, the obtained results significantly improved. This process is called fine-tuning. Fine-tuning is useful when the dataset is small and is used to adjust the already trained weights for the new dataset. Rather than finding the decoder weights from scratch, the pre-trained weights are adjusted for the new dataset and used to complete the training. Table 6 compares the 12 models based on the metrics that were explained. Figure 32 also visually compares the four models and the backbone of efficientnetb0.

Table 6. Comparison between segmentation models

Backbone	Efficientnetb0			Densenet121			Inceptionv3		
CNN Model	IoU score	F1 score	Fps speed	IoU score	F1 score	Fps speed	IoU score	F1 score	Fps speed
UNet	0.74	0.84	16.66	0.55	0.66	16.39	0.41	0.51	15.38
LinkNet	0.56	0.68	12.5	0.45	0.58	11.11	0.58	0.7	14
FPN	0.59	0.71	7.69	0.62	0.74	7.14	0.54	0.67	7.43
PSPNET	0.51	0.63	27	0.56	0.68	26	0.5	0.62	27

FSP: frame per second

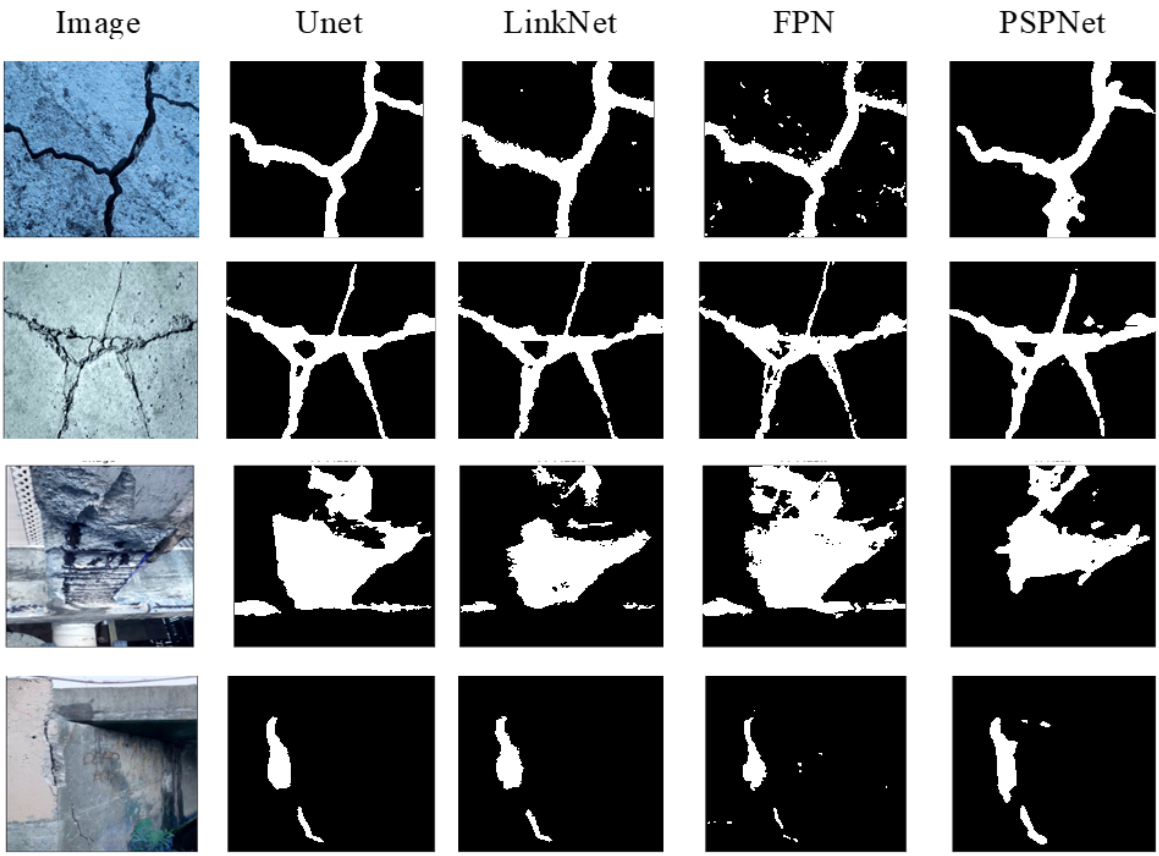

Figure 32. Segmentation model comparisons with inceptionb0 backbone.

Table 6 displays the comparison between the models. Another metric, F1, which is the weighted average of the precision and recall indicating the model's accuracy, was also used to evaluate the defect quantification model. The model speeds were initially tested on Colab Pro GPU. The training results show that UNet with the backbone of efficientnetb0 is the most accurate model, and PSPNet with the same backbone had the highest inference speed. Both criteria are highly important for defect quantification. While inference speed depends on the model architecture and cannot be improved, the model accuracy can be improved by adding more data to the dataset. As previously mentioned, publicly available annotated image data for segmenting concrete defects is limited. On the other hand, defect annotation for segmentation models is very labor intensive, making it prohibitively time-consuming to generate a large and diverse dataset manually to perform well for all defects. Therefore a semi-supervised approach was followed to generate more annotated data. First, a small set of the existing data, which already included segmentation labels, was prepared for the correct format for the selected training procedure. Then, the segmentation architectures selected for testing were trained for 50 epochs until the models could not improve their precision. The model with the highest accuracy (UNet) was then run over an unlabeled dataset to generate more labelled data. The newly labelled dataset was added to the initial dataset for another round of training. This process was repeated multiple times to obtain accurate annotations. Figure 33 displays the improvement in annotations after three rounds of training.

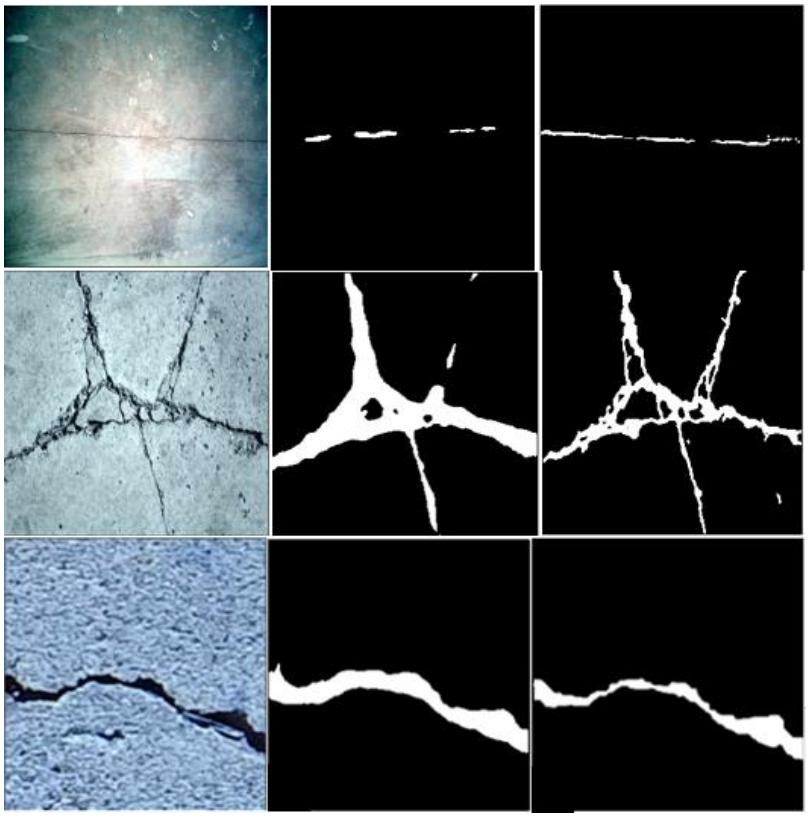

Figure 33**. Semi-supervised annotation improvement.**

AI-BASED MIXED REALITY ASSISTED INSPECTION

Collaborative work of defect localization and quantification

Collaborative work allows human-computer interaction. With this, the developed framework did not solely rely on the computer system output but interacted with inspectors. This platform provides continuous improvement and learning for the system as it is exposed to several cases. During a routine visual inspection, the inspector runs the defect localization model. The model runs in real time on the selected edge device. All the defects in the camera view are detected in real-time. If there are missing or incorrect detections, the inspector can manually correct the bounding boxes. Figure 34 displays the backend performance of defect localization and quantification.

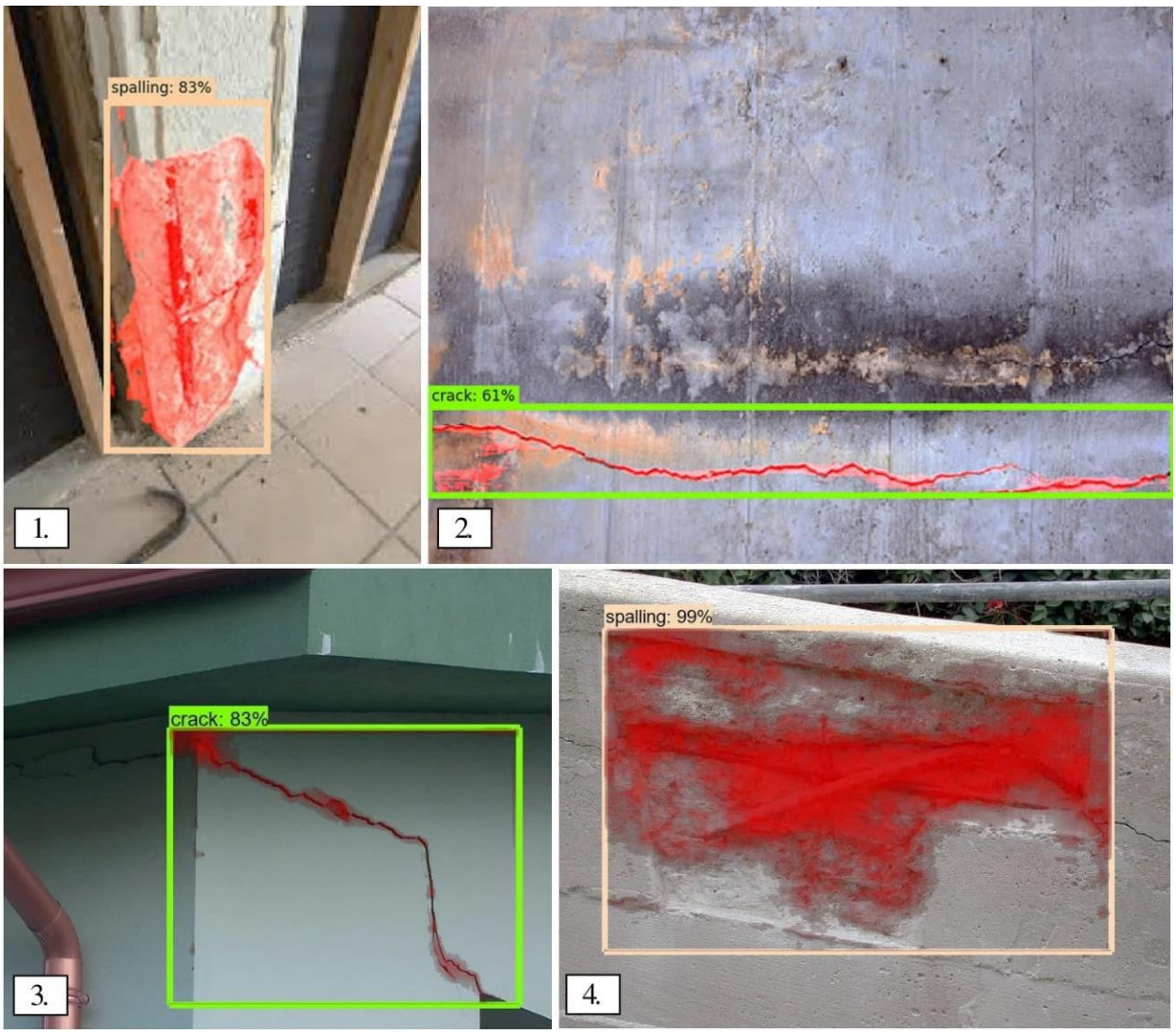

Figure 34. Defect localization and quantification performance on some test images.

Edge Computing for MR Platform

What makes this project unique is the integration of learning-based models into the MR platform to facilitate human-AI interaction. The MR platform requires a headset device to perform. Such devices have low computational power, so running machine learning algorithms on them is not trivial. This is one of the biggest challenges in this project. Unlike previous studies, where researchers use different AI models and make comparisons only based on their accuracy, in this study, it is important to have an accurate model that is also fast enough to perform in real-time and does not occupy much memory, so that it can be deployed in MR headsets. Consequently, many available deep learning architectures were initially eliminated due to their high computational footprint. The remaining were also tested for their inference speed. With a good processor, most of the deep learning models can perform in real-time; however, that is not the case for edge computation in mobile devices including MR headsets.

The edge is part of a distributed computing topology where information processing is located close to the edge, where things and people produce or consume that information. Edge computing is a distributed computing framework that brings enterprise applications closer to data sources such as IoT devices or local edge servers. In essence, edge computing is about geographic proximity. For years, data has been held in the cloud, which has been and will continue to be an asset in data storage. By geographic proximity, the cloud is off-premises, closer to the core of the data center. By comparison, edge technology processes data much closer to the source, such as an individual user or a connected device, for speed, latency, and security, rather than depending on the unknown of the cloud. Edge computation has many benefits, including boosting performance, enhancing privacy protections and data security, reducing operational costs, helping meet regulatory and compliance requirements, enhancing reliability and resiliency, and supporting AI/ML applications. Figure 35 shows some edge computing devices in the market.

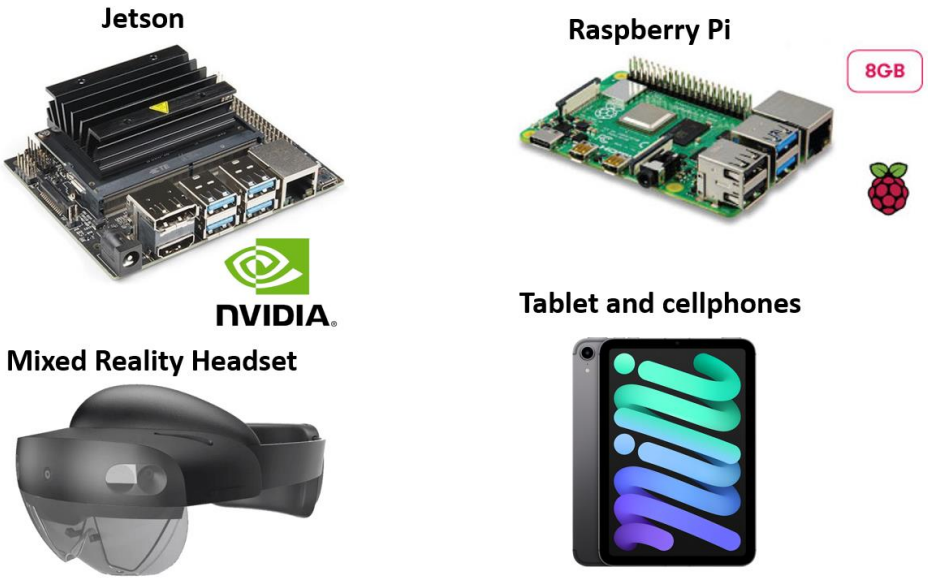

Figure 35. Edge computing devices.

With edge computation, rather than saving the raw data on a hard drive or on the cloud and processing it later, the data is collected and processed on the edge, saving time and energy, and meaningful results are saved in the cloud for decision-making. In this study, the results of the edge computation are the details of the defects on the concrete surface, which may be potentially used by bridge owners and decision-makers. Moreover, the data collected by each inspector becomes processed annotated data obtained during an inspection, ready to be added to datasets for AI training. The designed system in this study can improve its accuracy after each usage, making it unique compared to other studies available. Figure 36 displays the process.

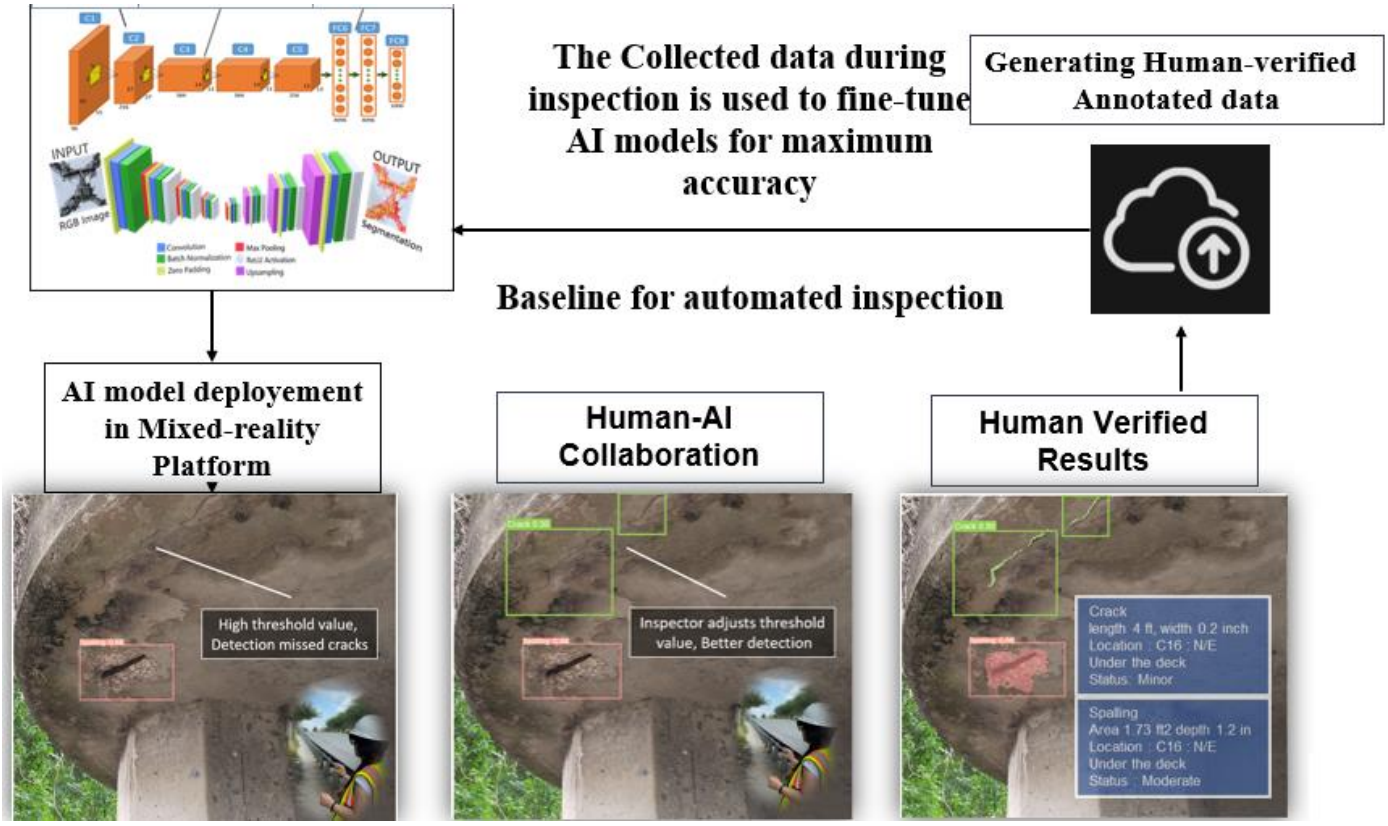

Figure 36. Bridge Inspection Using Edge Computation, IoT-based bridge inspection.

The introduced inspection method collects the data using AI, processes it in real-time on the edge, and stores it in the cloud for future use, making it an example of the Internet of Things (IoT) in the visual inspection of bridges.

Model Optimization, model deployment, and Edge Computation

Edge devices have limited memory and computational power. Although the developed AI algorithms in this study are lightweight and have a high inference speed, model optimization is still necessary for the real-time performance of the models on edge devices. The three most important properties of the model that need to be reduced for edge computation are storage size, memory usage (RAM), and latency. Model quantization is an optimization method that reduces the size of the model. It can also reduce the time it

takes to run an inference through the CNN model (latency). There is a small trade-off between memory size and accuracy. Generally, by reducing the latency and memory size, a small amount of accuracy is lost. Quantization is usually conducted post-training. There are different types of post-training quantization: post-training dynamic range quantization, post-training full integer quantization, and post-training float16 quantization. Some libraries, such as TensorFlow, offer techniques that can be applied to trained weights for quantization. Depending on the processor on the edge device, the quantization methodologies change. Dynamic range quantization reduces the model size to 25% of its original size and improves the model speed by 2-3 times the original value. Dynamic range quantization is suitable for edge devices that run on a CPU. Full integer quantization is another method of CNN quantization that decreases the model size by four and increases its speed to over three times its original value. The integer quantization method suits devices using CPU, edge TPU, and Microcontrollers. Float 16 quantization is also used for CPUs and GPUs.

Another important factor in determining the quantization method is the computational power and the required model speed. By using float 16 quantization, the size of floating points is halved: therefore reducing the size of the model by up to half while causing a minimal loss in accuracy. However, it does not reduce the latency of the model as much as quantization to fixed point math. With full integer quantization, more improvement in latency is acquired by quantizing the whole model to integers. As a result, more accuracy is lost. This method needs to convert all variables to integer variables, and therefore, the conversion requires a representative dataset for calibration.

In this study, given the designed ML models, different quantization methods were evaluated for three different edge devices: Raspberry PI with an edge-tpu processor, Jetson Nano with NVIDIA Maxwell GPU, and Microsoft HoloLens using Qualcomm Snapdragon 850 CPU figure(37). To deploy the models on Jetson nano with NVIDIA, TensorRT float 16 quantization was used to reduce inference. Float 16 quantization is supported by both Jetson and HoloLens 2. Jetson nano, however, has a more powerful processor and can handle multiple models performing in real-time with Float 16. Raspberry PI does not support float 16 quantization with edge-tpu. Instead, tflite-edgetpu integer 8 provides satisfactory real-time performance. Some platforms, such as the MR platform in this study, run in Unity and Barracuda and therefore require an ONNX model. ONNX (Open Neural Network Exchange Format) is a format which can be used for CNN models. This format allows conversions from different libraries, including Pytorch and Keras, used in this study. Float 16 quantization on HoloLens 2 provides a good inference speed; however, adding more models to the device requires more computational power; therefore, full integer quantization may be more appropriate.

The final decision on the quantization model depends on the criteria mentioned above and the performance necessary for the task. The mentioned devices and edge devices, such as tablets and cellphones, can be used for visual inspection. MR/AR platforms offer a better human-AI interaction among these models and therefore are recommended. Figure 38 shows the post-training quantization flow chart.

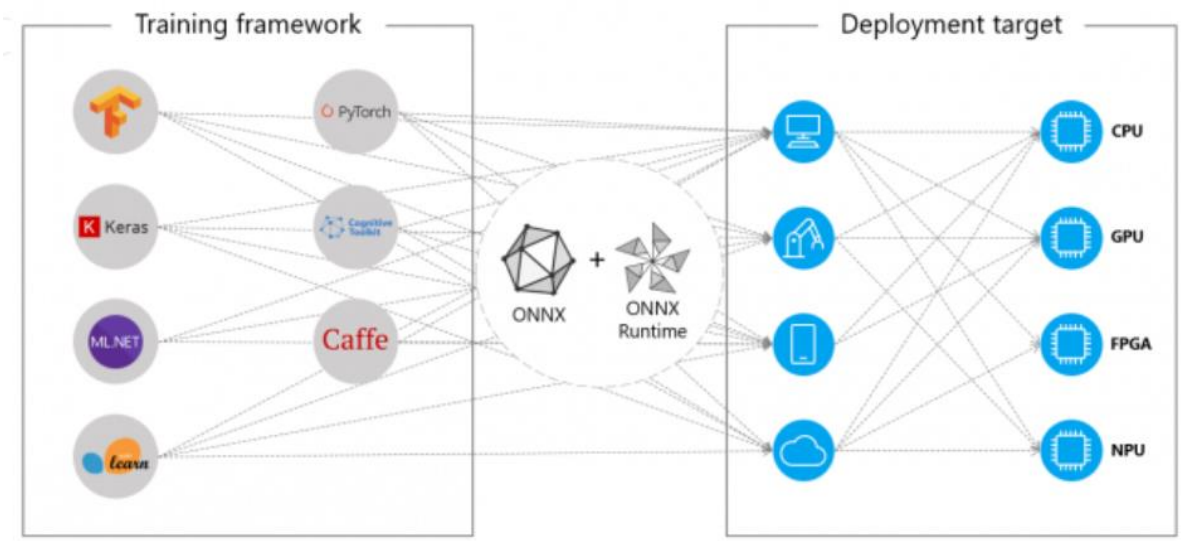

Figure 37**. Transforming ML models to ONNX (Open Neural Network Exchange) for deployment** (*75*)**.**

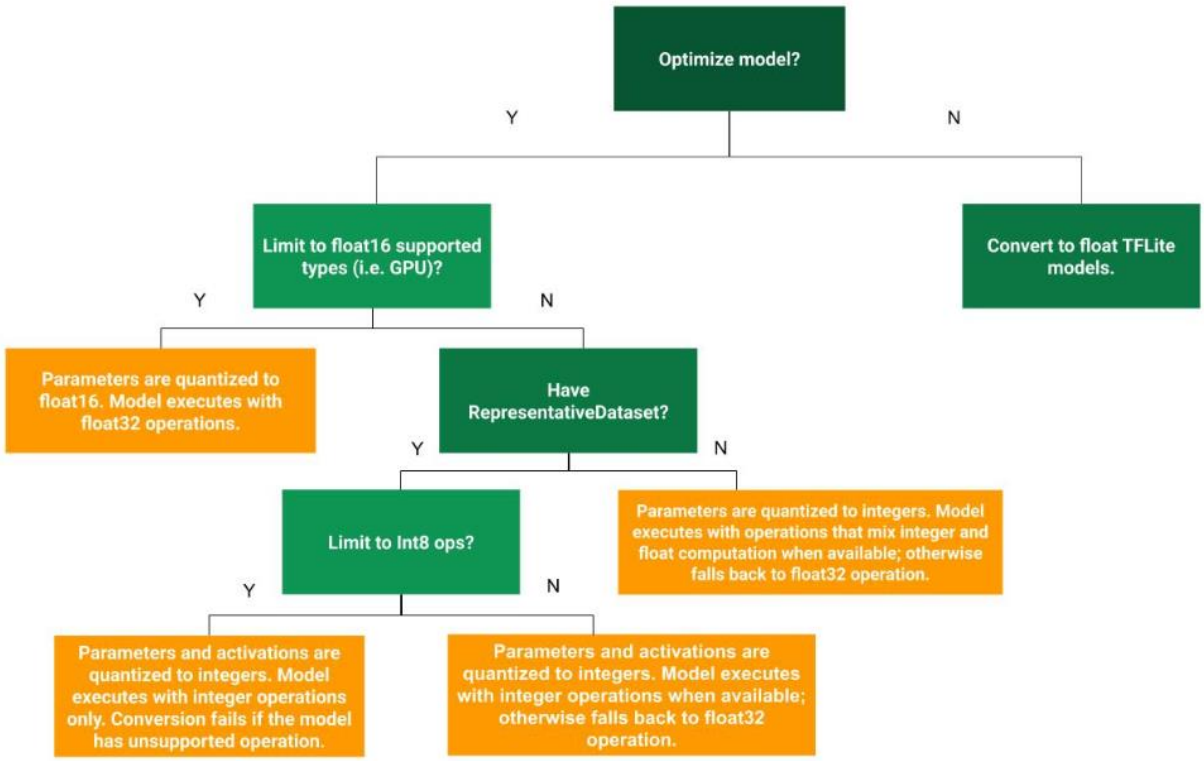

Figure 38. Post Training quantization flowchart (*76*)**.**

Deep Learning model deployment in HoloLens 2

Since the edge device used in this study is HoloLens 2, Barracuda, a lightweight cross-platform, needs to be used to run neural networks in HoloLens 2. The original weights obtained from the defect localization model, YOLOv5, use a library called Pytorch. Barracuda do not support Pytorch. Moreover, to run each model in real-time, the CNN models need to be quantized and converted to a suitable framework for operation in Barracuda. Quantization is a process that brings the neural network to a reduced size while maintaining accuracy. Quantization for deep learning approximates a neural network that uses floating-point numbers by a neural network of low-bit width numbers. This is essential for edge-device applications, where memory size and the number of computations are limited. As a result, the operation dramatically reduces the memory requirement and computational cost. The trained CNN architecture can be quantized during or after training. In this project, all CNN architectures were trained prior to quantization. Many attempts were made to obtain the optimal quantized model for deployment of the defect localization model (YoloV5). First, the trained YOLOV5 weight was converted to an ONNX graph. YOLOv5 has a complicated post-processing script. There are some operations that Barracuda do not support. There are two methods to convert the YOLOv5 weights to ONNX. The first method converts the weight and simplifies it to be deployed in HoloLens 2. The other method converts the model with post-processing into an ONNX file. The team obtained both of these models for testing purposes. For the ONXX model without post-processing, a script is being written that converts the original post-processing script from Pytorch to C#, a language supported by Unity and Barracuda. The steps that were taken to generate the algorithm for deploying the defect localization CNN model in the MR platform are as follows:

- Acquire the webcam frame using Unity OpenCV,
- Generate a function named Letterbox image to convert all the images to the appropriate size, known by the Yolov5 model,
- Normalize the pixel values to be recognized by the deep learning model,
- Run the image through the model using Barracuda,
- Extract the tensors with Barracuda,
- Extract the bounding boxes, confidence scores, and class numbers,
- Run non-max suppression to obtain the best detection,
- Display on the screen with bounding boxes.

The defect quantization receives the cropped image from the bounding boxes obtained in the previous steps and runs the model on the area. The size and the state of the defect then show up next to the defect. Figure 39 shows an example of the performance of a photo in the CITRS laboratory.

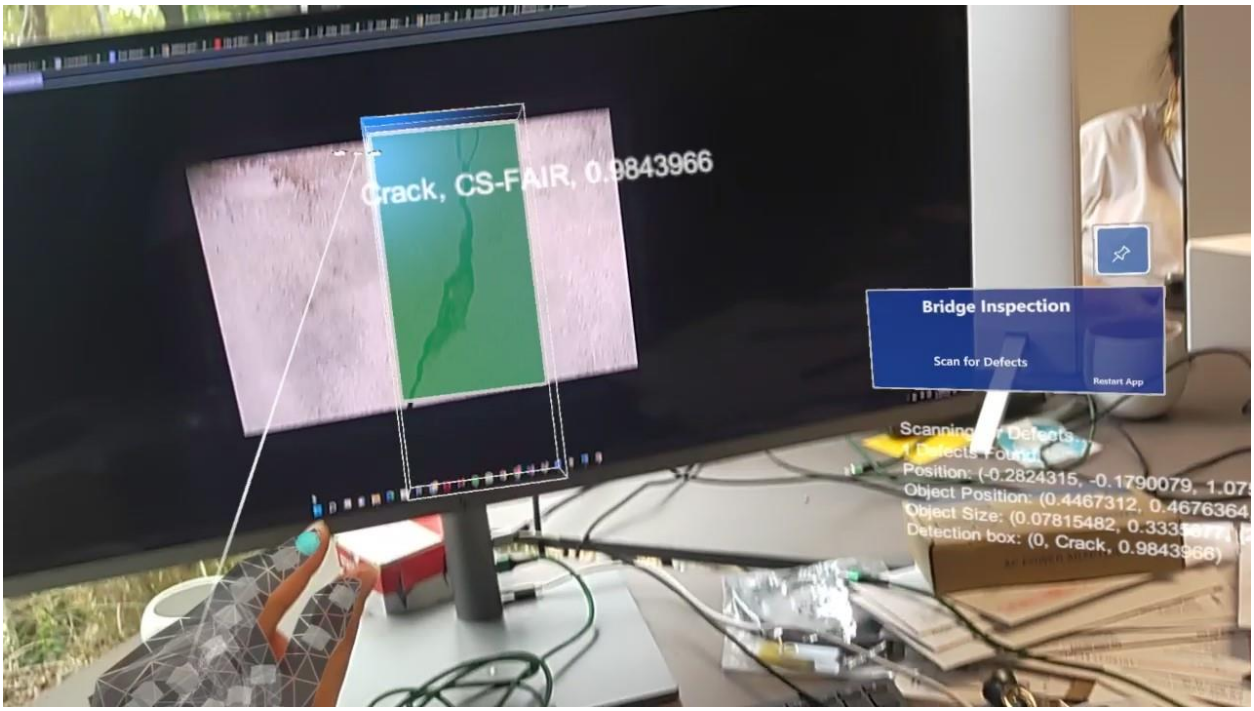

Figure 39**. Defect Localization and quantification results in the MR platform.**

Real-time Image Target Tracking

Conventionally, the camera localization for augmented reality (AR) relies on detecting a known pattern within the captured images, namely a marker. The first AR tracking applications used markers placed on the object to robustly register an image at different camera angles (*77*). However, depending on the application, placing markers in the scene was not always possible. Simon et al. (2000) used planar structures in the camera's scene to perform markerless AR tracking (*78*). In a different study, Ferrari et al. (2001) introduced markerless AR with a real-time affine region tracker (*79*). Recent works in AR tracking use the Visual SLAM algorithm (simultaneous localization and mapping) to perform robust markerless tracking (*80–82*). This study used an open-source AR tracking library called EasyAR to perform markerless tracking (*83*). EasyAR has a third-party plugin for Unity 3D, a widely used platform development environment for cross-platform applications (*84*). The ML-Agents plugin was also configured to implement the deep learning models in Unity (*85*).

After a crack or spall region is detected and accurately segmented from the scene, an image target is automatically created in the platform environment. The image targets work with feature-based 3D pose estimation using the calculated projection matrix (*86*). The projection matrix can be calculated by following the stereo camera calibration procedure provided by the headset manufacturers. After successful calibration, camera intrinsic and extrinsic parameters such as focal length, location, and camera orientation are retrieved in Unity using the headset sensors, gyroscope and head-position-tracker. EasyAR in Unity can create on-the-fly image targets from the damage-detection output and perform fast, robust markerless tracking using Visual SLAM. The 3D pose is estimated accurately at different angles and distances; the inspector still sees

the overlay information on the correct location, as shown in Figure 40.

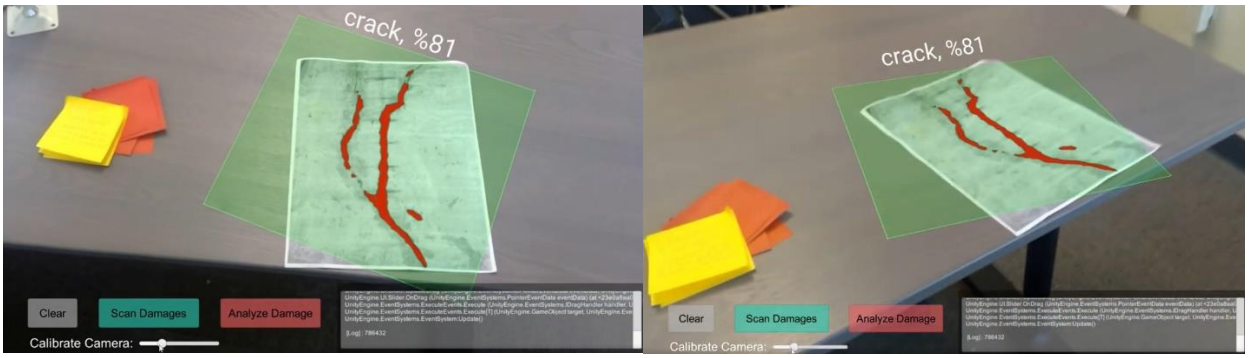

Figure 40. Markerless tracking of on-the-fly image targets created from AI analysis results

One of the important limitations of AR tracking is that segmentation results are only projected onto planar surfaces since the created image targets are two-dimensional. Therefore, volumetric calculations from curved surfaces (e.g. circular columns) have intrinsically large errors. In the future improvement of this work, 3D image targets will be created to perform AR projection onto curved concrete surfaces (figure 41).

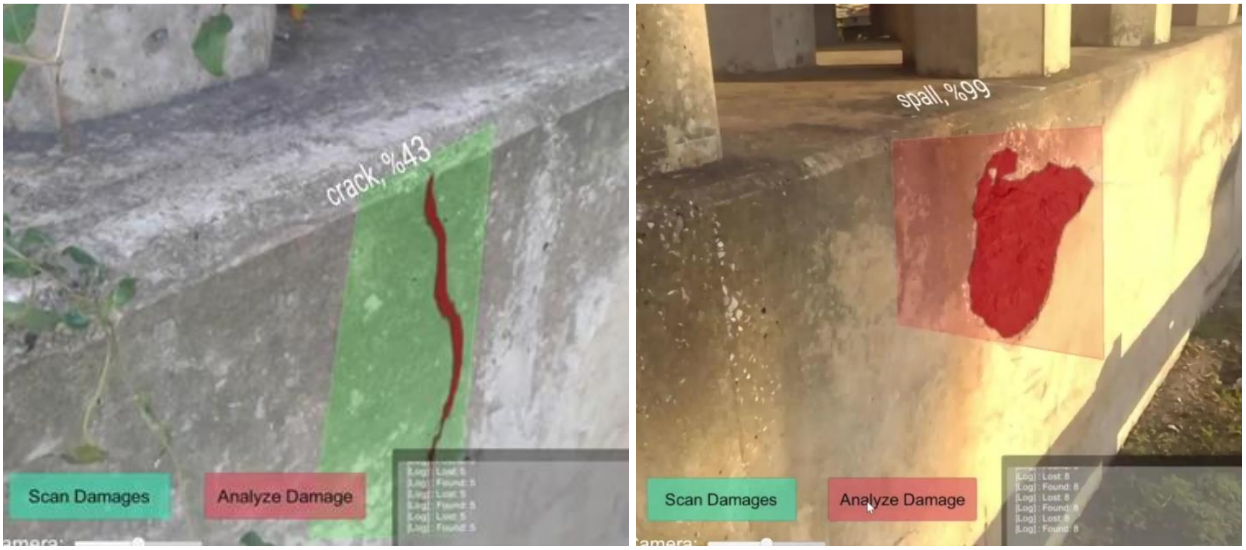

Figure 41: Real-world examples from the headset showing AI analysis results projected on the concrete defects (left: concrete crack; right: concrete spalling).

Retrieval of Dimensional Properties

Accurate retrieval of real-world dimensional properties from the AR projection is the most important step in estimating the condition of concrete damage in the proposed system. Depending on the stereoscopic video see-through technology, several techniques can be used to estimate the real geometric distance. The proposed methodology in this study uses spatial mapping techniques based on binocular disparity.

After a successful calibration, basic proportioning of image pixel size to a known real-world dimension (camera offset from eye focus is known) is used to calculate the area of a spall or length of a crack. Figure 42 shows calibrated image targets in the Unity platform.

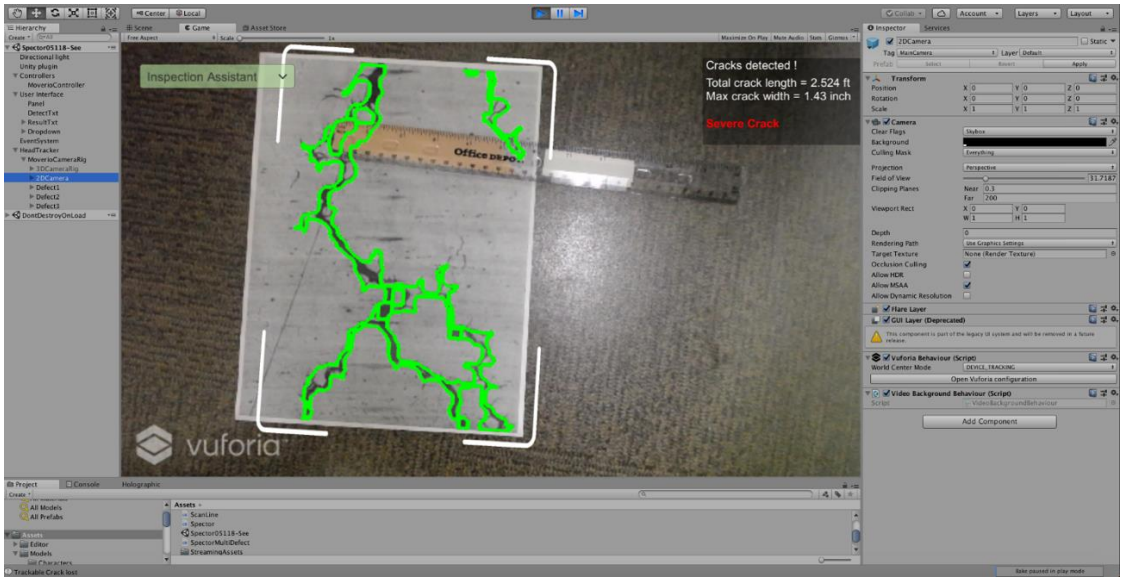

Figure 42: Calibrated image target that estimates maximum crack width in Unity.

In order to improve the estimation accuracy of the geometric properties, the AR target object projected onto the defect surface is continuously calibrated using non-linear least square fitting. The necessary data points were obtained in real-time from the headset's different camera positions as the inspector got closer to the object or looked at the defect from different angles. Figure 43 shows the details of the non-linear least square fit calculation of an example calibration of the image target to the estimated area of spalling.

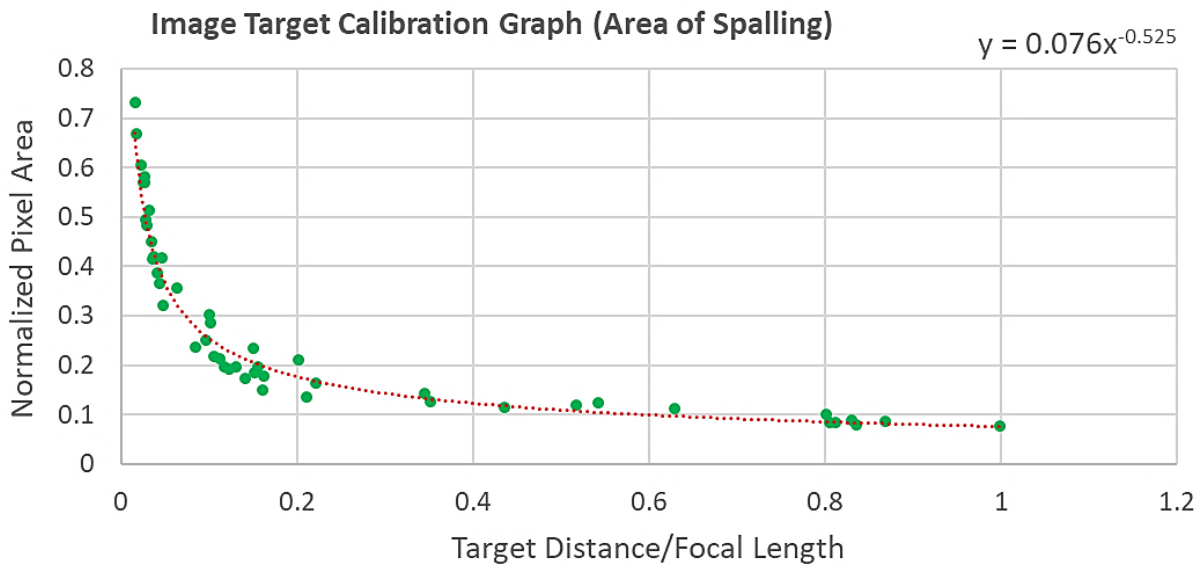

Figure 43: Calibration of image target for more accurate prediction of geometrical property

In the horizontal axis of the calibration, the estimated target distance was normalized by the focal length, and in the vertical axis, the pixel area of the target normalized by the camera resolution was used. The fit equation corrects the known distance parameter in the dimension proportion to predict the area at higher accuracy.

Evaluation of Factors Affecting the Geometric Estimation

Experimental work was conducted in the CITRS Lab to determine the factors affecting the retrieval accuracy of the dimensional properties. The experiment aimed to investigate the effect of environmental factors and camera specifications on geometric estimation. The described calibration method was repeated in the experiment multiple times for different ambient illuminations, crack widths, target distances, and camera resolutions using Moverio BT-300 smart glasses in a laboratory environment. Synthetically generated crack images with different thicknesses, illumination levels, and patterns were printed on letter-size papers and placed on a white platform. The experiment setup is shown in Figure 44, and the laboratory experiment results are tabulated in Table 7. In the experiment, it was assumed that the cracks were perfectly segmented using the methodology previously described.

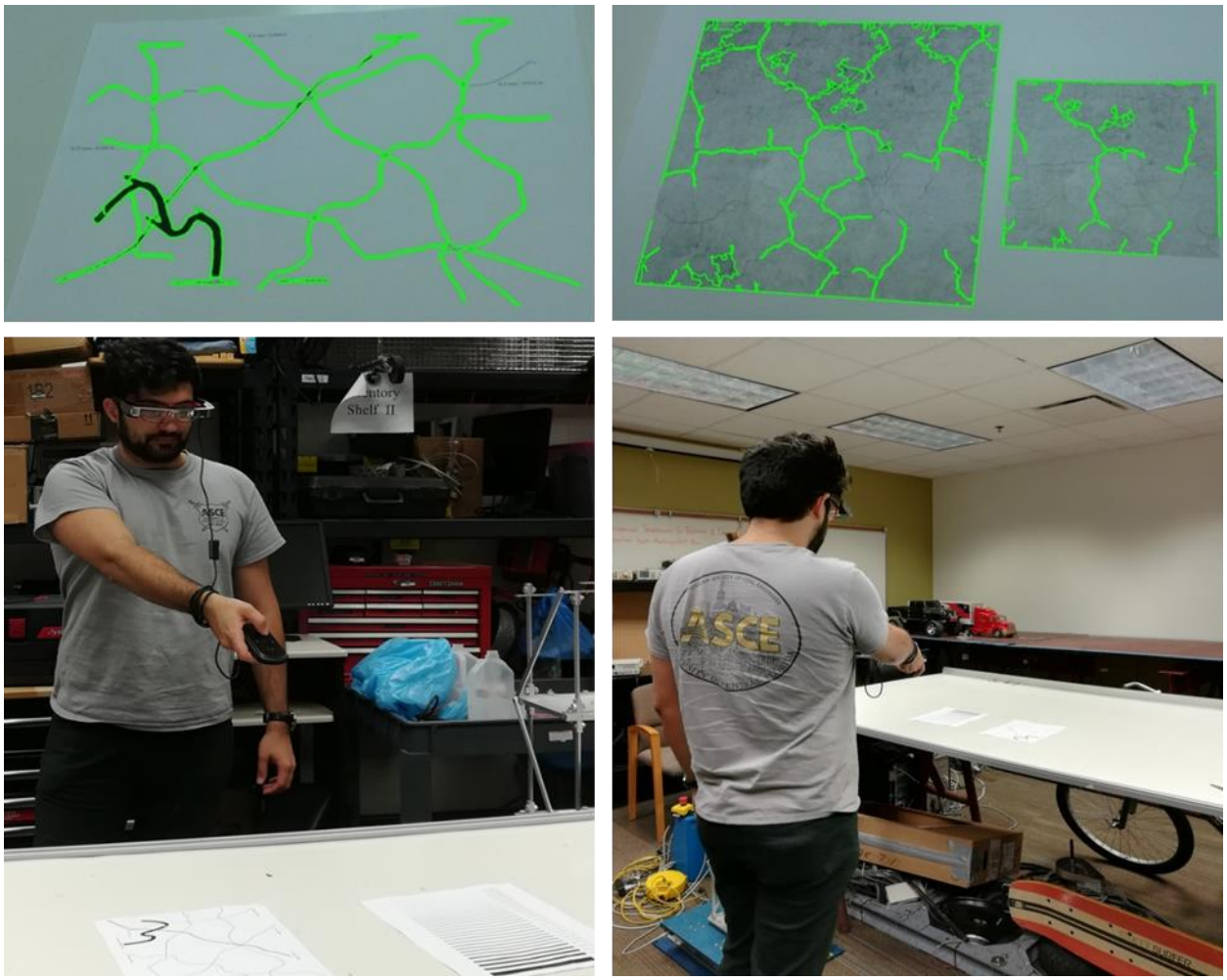

Figure 44: Experiment setup to evaluate factors affecting the performance of the geometric estimation in the MR headsets.

Table 7. Calculation of average error in geometric estimation under different conditions

Illumination	Camera Res.	Target Dist. (ft)	Crack Width (in)	Average Error (%)
1 light on	720p	3ft	1/6”	8.62%
1 light on	720p	3ft	1/2”	5.04%
1 light on	720p	3ft	1”	3.10%
1 light on	720p	5ft	1/6”	18.17%
1 light on	720p	5ft	1/2”	8.60%
1 light on	720p	5ft	1”	6.09%
1 light on	1080p	3ft	1/6”	5.35%
1 light on	1080p	3ft	1/2”	4.04%
1 light on	1080p	3ft	1”	2.76%
1 light on	1080p	5ft	1/6”	13.10%
1 light on	1080p	5ft	1/2”	6.27%
1 light on	1080p	5ft	1”	3.55%
2 lights on	720p	3ft	1/6”	8.02%
2 lights on	720p	3ft	1/2”	4.85%
2 lights on	720p	3ft	1”	2.87%
2 lights on	720p	5ft	1/6”	16.99%
2 lights on	720p	5ft	1/2”	7.07%
2 lights on	720p	5ft	1”	4.71%
2 lights on	1080p	3ft	1/6”	4.14%
2 lights on	1080p	3ft	1/2”	2.45%
2 lights on	1080p	3ft	1”	1.23%
2 lights on	1080p	5ft	1/6”	9.92%
2 lights on	1080p	5ft	1/2”	5.43%
2 lights on	1080p	5ft	1”	3.03%

The experiment showed that the geometric estimation accuracy largely depends on the distance of the crack from the headset camera. Due to the limited capabilities of the headset used in the experiment, the procedure was repeated only at short distances of 3ft and 5ft. At larger distances, the headset could not create the AR tracker objects of the targets. The camera resolution was also an important factor affecting the accuracy of the geometric estimation. Setting the camera resolution to 1080p (1920x1080) yielded acceptable results at both 3ft and 5ft. According to results, minor cracks (widths less than 1/6") can be reliably measured only at a 3ft distance and 1080p resolution. The illumination level was also a signification factor affecting the performance of the geometric estimations in the MR system. One of the two fluorescent lamps was turned off in the simulated environment to create a darker ambient light. However, the experiments were not repeated when both lights were turned off since the targets became completely invisible in the camera, and therefore, the headset failed to create the AR trackers. When both lights were turned on and the resolution was set to 1080p, the 1-inch crack can be measured at approximately 98% accuracy from a 3ft distance.

HUMAN-AI USER INTERFACE

The results of the CNN models in this study are 2D images. Initially, the defect localization detects the defects from the camera view. In the backend, the images obtained from the camera are all two-dimensional. However, in order to locate the exact position of the defect in the 3D view of the inspector, there is a need to obtain the 3rd dimension of the defect. In order to view the object in the MR environment, the depth information of the defect also needs to be obtained. HoloLens 2 has a Kinect camera, which can also measure depth. The Kinect camera captures RGB color images by RGB camera and measures depth information with a depth camera, which has the same camera model as a standard monocular camera. The real dimensions of the surroundings need to be defined in the correct 3D coordinates. The imaging principle based on perspective projection is shown in Figure 45.

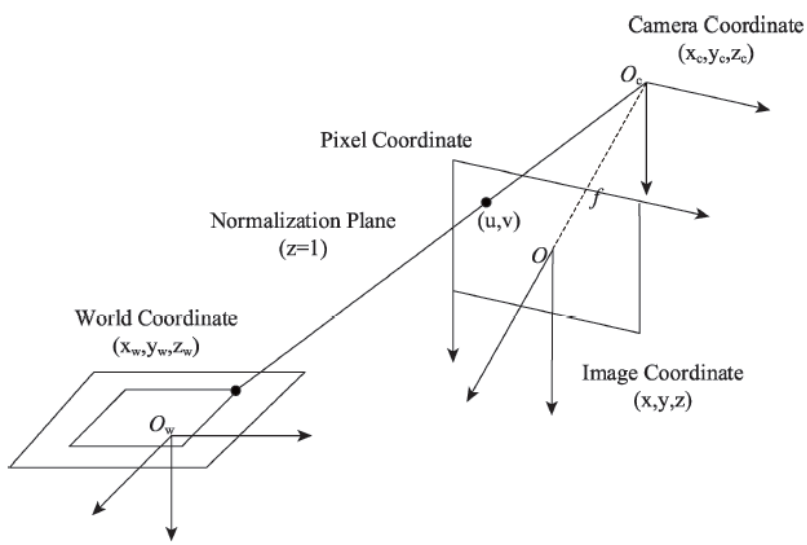

Figure 45 Principle of perspective projection imaging.

According to the principal formula of perspective projection, the transformation relationship between

the image coordinate system and camera coordinate system can be obtained as follows:

$$\frac{X}{X_C} = \frac{Y}{Y_C} = \frac{f}{Z_C} \quad \textbf{(6)}$$

The rigid body transformation relationship between the camera coordinate system and the world coordinate system is expressed as follows:

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix} = \begin{bmatrix} X_\omega \\ Y_\omega \\ Z_\omega \end{bmatrix} + T \quad \textbf{(7)}$$

Because the conversion relationship between the image coordinate system and the pixel coordinate system is (3), where α and β are the zoom ratios between the two coordinate axes of the image coordinate system and the pixel coordinate system, respectively:

$$u = \alpha X + u_0 \quad \textbf{(8)}$$

$$v = \beta Y + v_0 \quad \textbf{(9)}$$

By transforming the expressions into homogeneous expressions, the transformation relationship between any point in the world coordinate system and the corresponding point in the pixel coordinate system can be obtained as follows:

$$Z_C \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{pmatrix} R & T \\ 0 & 1 \end{pmatrix} \begin{bmatrix} X_\omega \\ Y_\omega \\ Z_\omega \\ 1 \end{bmatrix} = M_{int} M_{ext} \begin{bmatrix} X_\omega \\ Y_\omega \\ Z_\omega \\ 1 \end{bmatrix} \quad \textbf{(10)}$$

The mapping relationship between the spatial points in the three-dimensional world and the plane pixels in the two-dimensional image is obtained by introducing the pinhole camera model. Usually, for calculating the 3rd dimension from an RGB camera, another factor needs to be accounted for distortion. However, an RGB-D camera is generally more precise, so the distorted model is not considered here and is negligible. The distortion error usually increases with a high distance. The distance between the inspector and the structure does not reach a significant amount; therefore, not accounting for distortion does not significantly affect the result.

Another step in locating the objects in the 3D world is spatial mapping. Spatial mapping provides a detailed representation of real-world surfaces in the environment around the HoloLens. This allows the device to recognize the surrounding world accurately, including surfaces, dimensions, etc. The application can make holograms appear realistically by merging the real world with the virtual world. It also allows the inspector to interact with the virtual objects in a way that simulates their existence in the real world. Figure 45 is an example of spatial mapping in HoloLens 2.

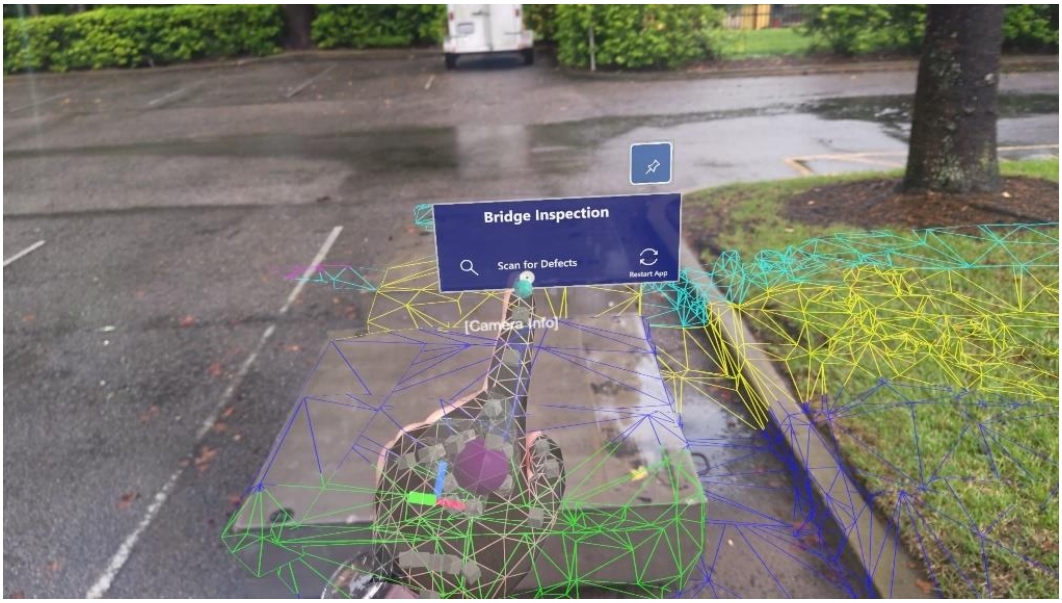

Figure 46. Designed user interface and spatial meshing in the MR platform

Spatial mapping makes it possible to place virtual objects on real surfaces (figure 46). The mapping relationship between the spatial points in the three-dimensional world and the plane pixels in the two-dimensional image is obtained for each detected defect. The located object displays each detected defect. Figure 47 displays a defect detected on the screen. The 3D box around the crack is the 3D object generated from the 2D image using the methodology explained. The inspector can tap on the box for defect quantification, and the results appear on the box after analysis. The defects condition is approximated using the information given in the next section.

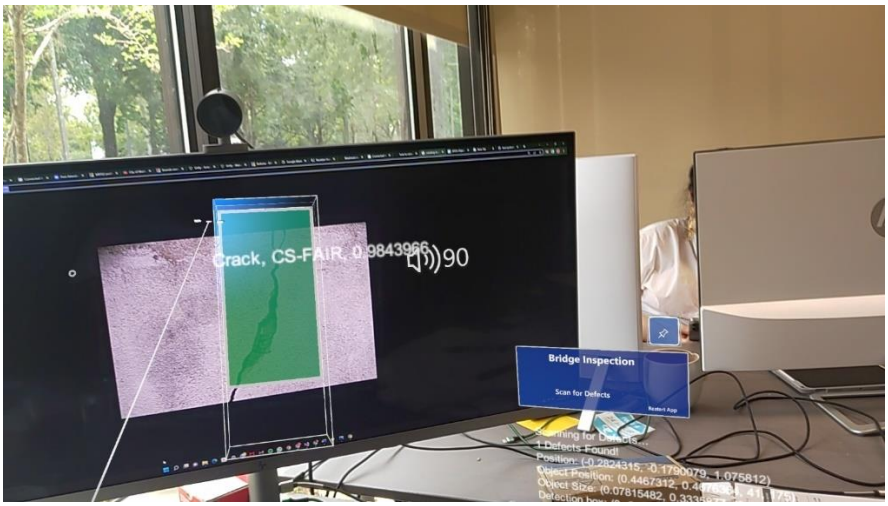

Figure 47. 3D object generated using spatial mapping and depth camera on the location of the synthetic defect.

CONDITION ASSESSMENT

The condition assessment methodology based on the AI system's damage analysis will require answers: "how wide is this crack?" "Which one of the bridge piers is closer?" "What is the camera height, rotation, or focal length?" This information is required for identifying actual measures of defects for accurate assessment of infrastructures and also for augmenting certain objects onto 3D views or highlighting defects in MR headsets. Using projective geometry and camera calibration models, it is possible to perform correct projections of objects in 3D, achieve scene reconstruction, and accurately predict the actual dimensions of objects. However, performing transformations in 3D spaces requires using 4D projective geometry instead of conventional 3D Euclidian geometry (*87*). The projection matrix allowing camera rotation is defined as in Equation (11).

$$x = K[R \quad t]X \tag{11}$$

where x: Image coordinates, K: Intrinsic matrix, R: Rotation matrix, t: Translation, X: World coordinates. The projected coordinate vector x is calculated by multiplying the world coordinates by the rotation and translation-free projection matrix. The coordinate parameters are then put into equations as in Equation (12).

$$w\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K\begin{bmatrix} \alpha & s & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix}\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{12}$$

U and v represent the local coordinates on the image plane v; w defines the scale of the projected object; α and β stand for rotation angles concerning coordinate axes and are short for sinus function. Unity allows camera control that helps developers perform correct projections onto image planes from a 3D view. The projection is described in Figure 48.

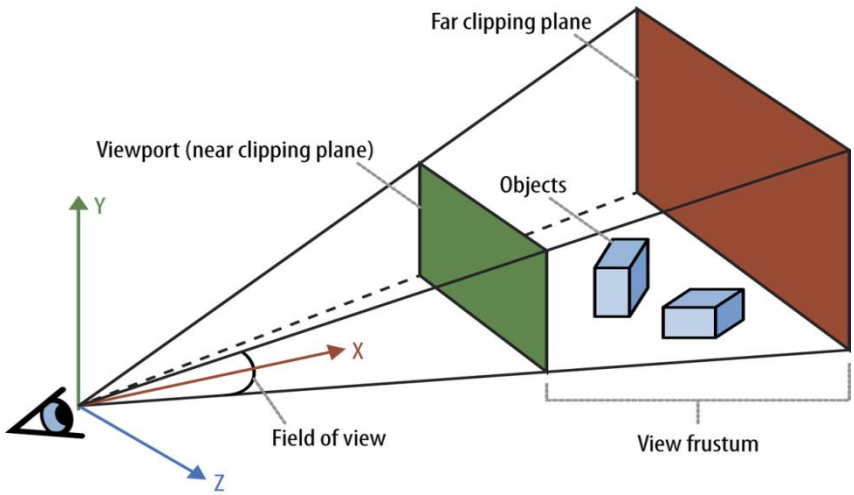

Figure 48**. Camera, viewport, and projection of real-world objects onto a 2D image plane [71].**

The first important factor in the condition assessment of each defect is its location. In the MR platform, either the AR model of the bridge needs to be deployed prior to the inspection, with details of the bridge elements and location, or the inspector needs to input the defect location manually. Since building an AR model of a bridge is not in this project's scope, building infinite models for different bridges is also time-consuming and not feasible for condition assessment; human involvement was increased to obtain the best results. During the inspection, the inspector has access to the design details of the bridge (this is already part of a routine inspection); therefore, the inspector can input the defect location manually after analysis of each defect. In general, crack severity is either measured based on its width or its density. The crack width in this project is measured using a defect quantification model and a conversion algorithm obtained in the last portion of this project to transform dimensions from pixel coordinates to global coordinates. The area of the spalling is the second variable calculated using the defect quantification model and transformation algorithm. However, the area of the spalling is not the only factor in assessing the condition of the defect. The spalling depth and the amount of exposed rebar are also important when assessing spalling. To tackle this issue, the Human-AI interface allows the inspector to input other important factors for condition assessment. HoloLens 2 has a depth camera and sensors and can estimate the depth of the spalling as well. The inspector can use the MR platform to estimate the depth of the spalling or use an approximation using handheld tools. It is important to develop an approach and an algorithm for condition assessment based on concrete defect conditions in Figure 49.

CONCRETE		CONDITION STATES			
DEFECT		1	2	3	4
		GOOD	FAIR	POOR	SEVERE
1080	Delamination/ Spall/Patched Area	None	Delaminated. Spall 1 in. or less deep or 6 in. or less in diameter and/or a patched area that is sound	Spall greater than 1 in. deep or greater than 6 in. diameter. Patched area that is unsound or showing distress. Does not warrant structural review.	The condition warrants a structural review to determine the effect on strength or serviceability of the element or bridge; OR a structural review has been completed and the defects impact strength or serviceability of the element or bridge
1090	Exposed Rebar	None	Present without measurable section loss.	Present with measurable section loss but does not warrant structural review.	
1100	Exposed Prestressing	None	Present without section loss.	Present with section loss but does not warrant structural review.	
1110	Cracking (PSC)	Width less than 0.004 in. or spacing greater than 3 ft.	Width 0.004 - 0.009 in. or spacing 1.0 - 3.0 ft.	Width greater than 0.009 in. or spacing less than 1 ft.	
1120	Efflorescence/ Rust Staining	None	Surface white without build-up or leaching without rust staining.	Heavy build-up with rust staining.	
1130	Cracking (RC and Other	Width less than 0.012 in. or spacing greater than 3.0 ft.	Width 0.012 - 0.05 in. or spacing of 1.0 - 3.0 ft.	Width greater than 0.05 in. or spacing of less than 1 ft.	

Figure 49. Defect Condition states-girders (from VDOT Bridge Element Inspection Report)

BRIDGE INSPECTION USING MR AND REAL-TIME ML: CASE STUDY

The performance of the designed system was initially tested on flat areas, such as sidewalks with cracks, for performance evaluation. As previously mentioned, the 2D objects are converted to 3D objects and placed at the location of the crack. One benefit of this system is that these objects remain even if the inspector leaves the area to be reviewed later. This is also useful for structures with BIM or digital twin models. The objects can be overlayed to BIM models later on. Figure 50 shows the initial detection of a crack on the sidewalk of Rev Kenneth Bridge. The inspector moves; however, the detected object is still in the same place. The 3D view of picture 2 is due to the change in the inspectors' location.

Figure 50. Mixed Reality assisted infrastructure inspection.

The designed mixed-reality headset was tested on a bridge to evaluate its performance. A railway bridge, Reverend Kenneth C. Crossman Bridge, was selected for this purpose. This bridge is located in Orlando, 20 miles from the University of Central Florida CITRS lab, connecting Winter Park to Maitland. The deck of the bridge is a steel truss with concrete abutments. The bridge is considered a landmark in Maitland, FL (figure 51).

Figure 51. Rev. Kenneth Bridge

The bridge is located on Orange Avenue, connecting Maitland to Winter Park. The upper section of the bridge is not accessible. However, the bridge pier can be easily accessed. The structure has visible cracks and some minor spalling.

Initially, the inspector runs the defect localization by tapping on the icon in the view. The icon can be pinned to a specific location or moved with the inspector depending on the inspector's need. Figure 50 displays the icon. The inspector can tap the icon or start the scanning by voice command: 'scan for defects" (Figure 52). Each detected defect is then analyzed if the inspector taps on the 3D objects generated from each detection (Figure 53). If the bounding boxes obtained during the inspection are too big or too small, the inspector can hold the corners and adjust the bounding box sizes (Figure 53). Figures 54, 55, and 56 show some of the results. The spalling in Figure 56 is detected correctly. However, the size of the bounding box is too big, and its location is also not precise. This problem was observed in multiple locations showing that the depth camera in HoloLens 2 performs poorly in some instances. This problem was mainly observed in more challenging areas with height differences. The issue was not limited to the bounding boxes. The depth of the object was measured incorrectly, and therefore, even with manual changes, the detection did not represent the real defect, and consequently, the defect measurements were unreliable.

Figure 52. By tapping on the green detection box, the crack is analyzed.

Figure 53. The results of the analyzed cracks are written next to each bounding box.

Figure 54. The inspector can modify the bounding boxes for better results.

Figure 55. The spalling localization and quantification are done correctly; however, the object was placed incorrectly.

Figure 56. A well-detected and localized defect in a 3D world.

While MR headsets are ideal for human-AI interaction, the limitations of the MR headsets available cause some difficulties in using MR-assisted inspection. Among these issues is the latency in MR applications and the error that occurred in in-depth measurements, which are sometimes significant. Moreover, the datasets available for the AI models need to be standardized for the MR device camera. More data needs to be collected using MR headsets and added to the AI headset to improve accuracy.

PLANS FOR IMPLEMENTATION

In the US, most DOTs spend 50-80% of their budgets on maintenance, rehabilitation, and replacement of concrete bridge decks. DOTs are expected to be long-term customers due to the many direct benefits of our technology, such as cost-savings on maintenance management of concrete bridges.

In order to understand how novel technologies can be implemented, the PI and his students carried out research as part of the NSF I-Corps program. The U.S. National Science Foundation's Innovation Corps (I-Corps™) program was launched to support researchers in the customer discovery process, allowing teams to assess their inventions' market potential quickly. This program allows researchers to explore industry partnerships and investigate cutting-edge technology commercialization. As part of the NSF I-Corps

program conducted by PI Catbas and his team, it is found that each segment of the customers that engage in bridge inspection and maintenance has similar "pains":

1) The subjectivity of inspection (judgement and condition evaluation) and decision-making;
2) Traffic control/lane closure for inspection (social impact), as well as costs associated with control and closure and
3) Safety of inspectors and drivers.

The market opportunity was validated through interviews conducted during the NSF I-Corps program. For example, there are several Private Infrastructure Owners and Public Private Partnerships around the central Florida area, such as Disney, Universal, NASA, and Orlando International Airport, in addition to highway bridge owners, such as Florida DOT and Turnpike. PI and his team conducted interviews with those potential customers to get input for such technology and if they would be interested in using and purchasing. Moreover, Engineering Service Providers who conduct infrastructure inspections were also included in the interviews, and they indicated their interest in innovating their technologies while reducing the time and costs for inspections. It was also discovered that each customer knew of these NDE techniques but not to the extent of their reliability or which technology should be used for their structures for specific cases. Finally, companies in Human-Computer Interaction (HCI) focusing on human systems integration and simulation, mainly for entertainment and defense industries, are interested in our work. The team has made presentations at conferences such as TRB, IABMAS, and ASCE to reach out to more people, agencies and companies.

The researchers have also presented their work at Smart City meetings and a Smart City conference. This work received the Best Paper Award at the Smart City Conference. In addition, the work also received Kikuchi-Karlaftis Award on Artificial Intelligence (AI) for their work as part of this NCHRP-IDEA research project from the Transportation Research Board Artificial Intelligence and Advanced Computing Methods Committee. These market searches, interest from industry and academic groups and the uniqueness of this work encouraged the researchers to file for intellectual property. Several meetings were held at the University for this, and finally, a patent was filed:

> Karaaslan, E., Bagci, U. and Catbas, F.N. "Artificial Intelligence Collaborative System for Enhanced Inspections of Infrastructures Using Mixed Reality" (US Patent App. 17/247,389, 2021)

As part of this filing, the researchers will explore investors either to partner for the future commercialization of the product or to license the patent. The University of Central Florida's Technology Transfer Office will support these activities. To further promote the activities, PI Catbas and his student Dr Karaaslan attended Technology Venture Symposium to present their work from this NCHRP-IDEAS project to attendees, including established entrepreneurs and venture capitalists. Their presentation is entitled "Mixed Reality-Assisted Smart Concrete Inspection" (Link). The researchers plan to present our work to our Advisory Board at the next TRB, and we will coordinate to make a presentation to appropriate AASHTO Committees for possible use and adoption. They will continue contacting private industry, especially those using such technologies. Finally, the researchers will explore additional funding from government agencies such as USDOT and State DOTs as well as NSF for expanding this work with more field demonstrations and releasing newer versions of our methods and integration for decision-making components.

CONCLUSIONS

The NCHRP IDEA project presented in this report explored the development of novel wearable technologies equipped with interactive capabilities using both experienced human inspectors and AI. Implementation in the laboratory and on real-life structures was also discussed. The technical details of the project are presented in the report. Some of the key conclusions are listed in the following:

- Current scientific approaches have employed various learning-based methods for automatically detecting concrete defects while replacing human involvement. However, the developed method aimed to merge the inspector's expertise with AI assistance using a human-centric machine vision approach, thus yielding a more reliable civil infrastructure visual assessment practice.
- In deep learning-based models, the availability of training data is the most critical aspect of developing a reliable system with good accuracy in recognition. Yet, in infrastructure assessment, creating a large image dataset is particularly challenging. The proposed method, therefore, used an advanced data augmentation technique to generate a synthetically sufficient amount of crack and spall images from the available image data.
- The AI system follows a semi-supervised learning approach and consistently improves itself using verified detection and segmentation data in re-training. Semi-supervised learning successfully addresses the problems of a small data pool in AI training, particularly in damage detection applications where a comprehensive, publicly available image dataset is unavailable.
- The attention guide approach (sequential detection and segmentation) yielded a significant reduction in the computational cost of the segmentation operation since only a region of interest is used, while other comparable models (e.g. MaskRCNN) perform segmentation on the entire image and perform localization in parallel. The sequential model also significantly improved the segmentation performance of concrete defects.
- A mixed reality system is an ideal environment to facilitate human-computer interaction. It enables the human-centered AI to interact with the inspector instead of completely replacing human involvement during the inspection. This collective work will lead to quantified assessment and reduced labor time while ensuring human-verified results.
- Some future work needs to focus on improving the accuracy of MR-assisted inspection based on more data collected with MR headsets and standardized datasets.
- HoloLens 2 is a great platform for human-AI communications; however, the depth camera did not always measure the distances correctly.
- The current MR headsets in the market need to be improved to generate a smooth and efficient platform for MR-assisted visual inspection. With recent advances in MR headsets, it can be expected that this will be available soon.
- Finally, the user experiences in this particular study have been positive; however, human comfort and utilization in various field settings may require additional research.

While this project has accomplished the technical scopes outlined at the beginning of the project, there are other activities needed. Some of the recommendations for the implementation of mixed reality systems are provided in the following:

- Validation studies in real life for concrete structures: More data collection from bridges with defects that are more representative in real life would help fine tune the real-life implementation. In addition, this will also allow to test and consider other issues such as environmental and lighting conditions.
- Investigate user experience: A very important consideration is the safe and comfortable use of such wearable headset technologies in real life. More studies on inspector comfort, and safe usage are to be conducted along with the physical and psychological characteristics to the design of MR devices and systems for human use in sometimes harsh bridge environments. Ergonomics or human engineering research would also be as critical as technology development.
- Integration with BIM/Digital Twin of Bridges (Bridge management software) would be important to support day to day business operations of DOTs. For example, integration in workflow e.g. with AASHTOWare™ Bridge Management software (BrM) would help adoption of these technologies.

INVESTIGATOR PROFILES

Dr. F. Necati Catbas (PI) is an educator and researcher currently serving as a Lockheed Martin St. Laurent Professor at the University of Central Florida. Dr Catbas is the founding director of Civil Infrastructure Technologies for Resilience and Safety (CITRS) (https://www.cece.ucf.edu/CITRS/). Dr Catbas and his team focus on theoretical, experimental, and applied aspects of structural identification, structural health monitoring, nondestructive evaluation, and condition assessment of structural systems. Dr Catbas received several awards and honors for his research, teaching and service activities including the "Aftab Mufti Medal" from the Journal of Civil Structural Health Monitoring and International Society for Health Monitoring of Intelligent Infrastructure, and the "Technical Excellence Award" given by all Professional Engineering Societies in Central Florida region. Dr. Catbas is a registered professional engineer in the State of Florida, and he is an elected Fellow of the ASCE and Fellow of the SEI. Dr Catbas along with his former student Dr. Karaaslan and his collaborator Dr. Bagci also received Kikuchi-Karlaftis Award on Artificial Intelligence (AI) for their work as part of this NCHRP-IDEA research project from the Transportation Research Board Artificial Intelligence and Advanced Computing Methods Committee.

Dr Enes Karaaslan completed both his doctoral studies and his post-doctoral studies at the Civil Infrastructure Technologies for Resilience and Safety (CITRS) Lab of UCF. His research includes computer vision applications, mixed reality, edge computing for civil infrastructure and other applications. He is currently the CEO of the start-up company ConnectedWise LLC at the UCF Business Incubator, where he focuses on advanced technologies such as connected autonomous vehicles.

Ms. Mahta Zakaria is a Ph.D. Candidate at the Civil Infrastructure Technologies for Resilience and Safety (CITRS) Lab of the University of Central Florida (UCF). Her research interests include structural health monitoring using dynamic methods, computer vision, machine learning, and artificial intelligence. She served as Graduate Research and Teaching Assistant.

Key Collaborators:

Dr Joseph J. LaViola Jr. is the Charles N. Millican Faculty Fellow and Professor in the Department of Computer Science and directs the Interactive Computing Experiences Research Cluster of Excellence at the University of Central Florida. He is the director of the Modeling and Simulation graduate program and is also an adjunct associate research professor in the Computer Science Department at Brown University. He has been working in the areas of virtual and augmented reality and he is also the lead author of books such as "3D User Interfaces: Theory and Practice," the only comprehensive book on 3D user interfaces targeted toward virtual and augmented reality.

Dr Ulas Bagci is an Associate Professor currently at Northwestern University. Prof. Bagci teaches machine learning, advanced deep learning methods, computer and robot vision, and medical imaging courses. He has several international and national recognitions including best paper and reviewer awards. He has collaborated on machine learning, artificial intelligence and wearable technologies. Dr Bagci currently serves at Northwestern University's Radiology and Biomedical Engineering Department.

REFERENCES

1. Howard, A., M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam. Searching for MobileNetV3. 2019.
2. Xiong, Y., H. Liu, S. Gupta, B. Akin, G. Bender, Y. Wang, P.-J. Kindermans, M. Tan, V. Singh, and B. Chen. MobileDets: Searching for Object Detection Architectures for Mobile Accelerators. 2020.
3. Tan, M., R. Pang, and Q. V. Le. EfficientDet: Scalable and Efficient Object Detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10778–10787. https://doi.org/10.1109/CVPR42600.2020.01079.
4. Bochkovskiy, A., C.-Y. Wang, and H.-Y. M. Liao. YOLOv4: Optimal Speed and Accuracy of Object Detection. 2020.
5. Jocher, G., A. Stoken, J. Borovec, NanoCode012, A. Chaurasia, TaoXie, L. Changyu, A. V, Laughing, tkianai, yxNONG, A. Hogan, lorenzomammana, AlexWang1900, J. Hajek, L. Diaconu, Marc, Y. Kwon, oleg, wanghaoyang0106, Y. Defretin, A. Lohia, ml5ah, B. Milanko, B. Fineran, D. Khromov, D. Yiwei, Doug, Durgesh, and F. Ingham. Ultralytics/Yolov5: V5.0 - YOLOv5-P6 1280 Models, AWS, Supervise.Ly and YouTube Integrations. 2021. https://doi.org/10.5281/ZENODO.4679653.
6. Zakaria, Mahta ; Karaaslan, Enes; Catbas, F. N. Advanced Bridge Visual Inspection Using Real-Time Machine Learning in Edge Devices. *Advances in Bridge Engineering*, 2022. https://doi.org/10.1186/s43251-022-00073-y.
7. American Road and Transportation Builders Association, "Bridge Report". (https://artbabridgereport.org/reports/2022-ARTBA-Bridge-Report.pdf) 2022.
8. Florida Department of Transporation, Florida Bridge Information, https://www.fdot.gov.
9. Dong, C.Z. and Catbas, F.N., 2021. A review of computer vision–based structural health monitoring at local and global levels. Structural Health Monitoring, 20(2), pp.692-743..
10. Johnson, M. B., W. Casey, L. O'Donnell, D. Soden, P. Allec, and A. R. Marshall. *Manual for Bridge Element Inspection, Second Edition*. AASHTO, American Association of State Highway and Transportation Officials, 2019.
11. Karaaslan, E. Enhanced Concrete Bridge Assessment Using Artificial Intelligence and Mixed Reality. 2019, pp. 2004–2019.
12. Abdel-Qader, I., O. Abudayyeh, and M. E. Kelly. Analysis of Edge-Detection Techniques for Crack Identification in Bridges. *Journal of Computing in Civil Engineering*, Vol. 17, No. 4, 2003, pp. 255–263. https://doi.org/10.1061/(asce)0887-3801(2003)17:4(255).
13. Nguyen, H. N., T. Y. Kam, and P. Y. Cheng. An Automatic Approach for Accurate Edge Detection of Concrete Crack Utilizing 2D Geometric Features of Crack. *Journal of Signal Processing Systems*, Vol. 77, No. 3, 2014, pp. 221–240. https://doi.org/10.1007/s11265-013-0813-8.
14. Yokoyama, S., and T. Matsumoto. Development of an Automatic Detector of Cracks in Concrete Using Machine Learning. *Procedia Engineering*, Vol. 171, 2017, pp. 1250–1255. https://doi.org/10.1016/j.proeng.2017.01.418.
15. Zhang, L., J. Shen, and B. Zhu. A Research on an Improved Unet-Based Concrete Crack Detection Algorithm. *Structural Health Monitoring*, No. 29, 2020. https://doi.org/10.1177/1475921720940068.
16. Zhang, Q., K. Barri, S. K. Babanajad, and A. H. Alavi. Real-Time Detection of Cracks on Concrete Bridge Decks Using Deep Learning in the Frequency Domain. *Engineering*, 2020. https://doi.org/10.1016/j.eng.2020.07.026.
17. Deng, J., Y. Lu, and V. C. S. Lee. Concrete Crack Detection with Handwriting Script Interferences Using Faster Region-Based Convolutional Neural Network. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 35, No. 4, 2020, pp. 373–388. https://doi.org/10.1111/mice.12497.

18. Adhikari, R. S., O. Moselhi, and A. Bagchi. Image-Based Retrieval of Concrete Crack Properties for Bridge Inspection. *Automation in Construction*, Vol. 39, 2014, pp. 180–194. https://doi.org/10.1016/j.autcon.2013.06.011.
19. Ren, Y., J. Huang, Z. Hong, W. Lu, J. Yin, L. Zou, and X. Shen. Image-Based Concrete Crack Detection in Tunnels Using Deep Fully Convolutional Networks. *Construction and Building Materials*, Vol. 234, 2020, p. 117367. https://doi.org/10.1016/j.conbuildmat.2019.117367.
20. Eslami, E., and H. B. Yun. Attention-Based Multi-Scale Convolutional Neural Network (A+MCNN) for Multi-Class Classification in Road Images. *Sensors 2021, Vol. 21, Page 5137*, Vol. 21, No. 15, 2021, p. 5137. https://doi.org/10.3390/S21155137.
21. Park, S. E., S. H. Eem, and H. Jeon. Concrete Crack Detection and Quantification Using Deep Learning and Structured Light. *Construction and Building Materials*, Vol. 252, 2020, p. 119096. https://doi.org/10.1016/j.conbuildmat.2020.119096.
22. Koch, C., K. Georgieva, V. Kasireddy, B. Akinci, and P. Fieguth. A Review on Computer Vision Based Defect Detection and Condition Assessment of Concrete and Asphalt Civil Infrastructure. *Advanced Engineering Informatics*, Vol. 29, No. 2, 2015, pp. 196–210. https://doi.org/10.1016/j.aei.2015.01.008.
23. Dorafshan, S., R. J. Thomas, and M. Maguire. SDNET2018: An Annotated Image Dataset for Non-Contact Concrete Crack Detection Using Deep Convolutional Neural Networks. *Data in Brief*, Vol. 21, 2018, pp. 1664–1668. https://doi.org/10.1016/J.DIB.2018.11.015.
24. Mundt, M., S. Majumder, S. Murali, P. Panetsos, and V. Ramesh. CODEBRIM: COncrete DEfect BRidge IMage Dataset. 2019. https://doi.org/10.5281/ZENODO.2620293.
25. Bianchi, E., ; Amos, L. Abbott, P. Tokekar, M. Hebdon, and M. Asce. COCO-Bridge: Structural Detail Data Set for Bridge Inspections. *Journal of Computing in Civil Engineering*, Vol. 35, No. 3, 2021, p. 04021003. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000949.
26. Zhang, C., C. C. Chang, and M. Jamshidi. Simultaneous Pixel-Level Concrete Defect Detection and Grouping Using a Fully Convolutional Model. *Structural Health Monitoring*, Vol. 20, No. 4, 2021, pp. 2199–2215. https://doi.org/10.1177/1475921720985437.
27. Behzadan, A. H., S. Dong, and V. R. Kamat. Augmented Reality Visualization: A Review of Civil Infrastructure System Applications. *Advanced Engineering Informatics*, Vol. 29, No. 2, 2015, pp. 252–267. https://doi.org/10.1016/j.aei.2015.03.005.
28. Mihelj, M., D. Novak, and S. Beguš. *Virtual Reality Technology and Applications*. 2014.
29. Azuma, R., R. Behringer, S. Feiner, S. Julier, and B. Macintyre. Recent Advances in Augmented Reality. *IEEE Computer Graphics and Applications*, Vol. 2011, No. December, 2001, pp. 1–27. https://doi.org/10.4061/2011/908468.
30. Sutherland, I. E. The Ultimate Display. 1965.
31. Milgram, P., and F. Kishino. Taxonomy of Mixed Reality Visual Displays. *IEICE Transactions on Information and Systems*, Vol. E77-D, No. 12, 1994, pp. 1321–1329. https://doi.org/10.1.1.102.4646.
32. Catbas, F.N., Luleci, F., Zakaria, M., Bagci, U., LaViola, J.J., Cruz-Neira, C. and Reiners, D., 2022. Extended Reality (XR) for Condition Assessment of Civil Engineering Structures: A Literature Review. Sensors, 22(23), p.9560.
33. Lungu, A. J., W. Swinkels, L. Claesen, P. Tu, J. Egger, and X. Chen. A Review on the Applications of Virtual Reality, Augmented Reality and Mixed Reality in Surgical Simulation: An Extension to Different Kinds of Surgery. *https://doi.org/10.1080/17434440.2021.1860750*, Vol. 18, No. 1, 2020, pp. 47–62. https://doi.org/10.1080/17434440.2021.1860750.
34. LaViola, J., E. Kruijff, D. Bowman, and I. Poupyrev. *3D User Interfaces: Theory and Practice, Second Edition*. 2017.
35. Kamat, V. R., and S. El-Tawil. Evaluation of Augmented Reality for Rapid Assessment of Earthquake-Induced Building Damage. *Journal of Computing in Civil Engineering*, Vol. 21, No. 5,

2007, pp. 303–310. https://doi.org/10.1061/(ASCE)0887-3801(2007)21:5(303).
36. Behzadan, A. H., and V. R. Kamat. Georeferenced Registration of Construction Graphics in Mobile Outdoor Augmented Reality. *Journal of Computing in Civil Engineering*, Vol. 21, No. 4, 2007, pp. 247–258. https://doi.org/10.1061/(ASCE)0887-3801(2007)21:4(247).
37. Bae, H., M. Golparvar-Fard, and J. White. High-Precision Vision-Based Mobile Augmented Reality System for Context-Aware Architectural, Engineering, Construction and Facility Management (AEC/FM) Applications. *Visualization in Engineering*, Vol. 1, No. 1, 2013, p. 3. https://doi.org/10.1186/2213-7459-1-3.
38. Ioannis, B. Mixed Reality Constructs a New Frontier for Maintaining the Built Environment. *Proceedings of the Institution of Civil Engineers - Civil Engineering*, Vol. 170, No. 2, 2017, p. 53. https://doi.org/10.1680/jcien.2017.170.2.53.
39. Moreu, F., B. Bleck, S. Vemuganti, D. Rogers, and D. Mascarenas. Augmented Reality Tools for Enhanced Structural Inspection. No. 2, 2017, pp. 3124–3130.
40. Microsoft. The Leader in Mixed Reality Technology | HoloLens. *Microsoft*.
41. Brito, C., N. Alves, L. Magalhães, and M. Guevara. BIM MIXED REALITY TOOL FOR THE INSPECTION OF HERITAGE BUILDINGS. 2019. https://doi.org/10.5194/isprs-annals-IV-2-W6-25-2019.
42. Dang, N. S., and C. S. Shim. BIM-Based Innovative Bridge Maintenance System Using Augmented Reality Technology. *Lecture Notes in Civil Engineering*, Vol. 54, 2020, pp. 1217–1222. https://doi.org/10.1007/978-981-15-0802-8_195/COVER.
43. Maharjan, D., M. Agüero, D. Mascarenas, R. Fierro, and F. Moreu. Enabling Human–Infrastructure Interfaces for Inspection Using Augmented Reality: *https://doi.org/10.1177/1475921720977017*, Vol. 20, No. 4, 2020, pp. 1980–1996. https://doi.org/10.1177/1475921720977017.
44. Nguyen, D.-C., R. Jin, C.-H. Jeon, and C.-S. Shim. BIM-Based Mixed-Reality Application for Bridge Inspection and Maintenance. Vol. 22, No. 3, 2022, pp. 487–503. https://doi.org/10.1108/CI-04-2021-0069.
45. Zakaria, Mahta ; Karaaslan, Enes; Catbas, F. N. Human-AI Collaboration in Bridge Inspection Using Mixed Reality Platform. In *Bridge Safety, Maintenance, Management, Life-Cycle, Resilience and Sustainability*.
46. Al-Sabbag, Z. A., C. M. Yeum, and S. Narasimhan. Distributed Collaboration in Infrastructure Assessment through Mixed and Virtual Reality - A. *Transforming Construction with Reality Capture Technologies*, 2022. https://doi.org/10.57922/TCRC.642.
47. Xie, D., L. Zhang, and L. Bai. Deep Learning in Visual Computing and Signal Processing. 2017, pp. 1–18. https://doi.org/10.1155/2017/1320780.
48. AASHTO. Guide Manual for Bridge Element Inspection. *Bridge Element Inspection Manual*, 2011, p. 172.
49. FHWA. National Bridge Inspection Standards Regulations (NBIS). *Federal Register*, Vol. 69, No. 239, 2004, pp. 15–35.
50. FDOT. *Florida DOT Bridge Inspection Field Guide*. 2016.
51. Hill, M. Overview of Human-Computer Collaboration. Vol. 8, No. June, 1995, pp. 67–81.
52. Karaaslan, E., B. Sen, T. Ercan, H. Laman, and J. Pol. Artificial Intelligence Embedded On-Board Machine Vision System to Support Vehicle to Infrastructure. *Transportation Research Record*, 2020.
53. Mansoor, A., U. Bagci, B. Foster, Z. Xu, G. Z. Papadakis, L. R. Folio, J. K. Udupa, and D. J. Mollura. Segmentation and Image Analysis of Abnormal Lungs at CT: Current Approaches, Challenges, and Future Trends. *Radiographics*, Vol. 35, No. 4, 2015, p. 1056. https://doi.org/10.1148/RG.2015140232.
54. LaLonde, R., and U. Bagci. Capsules for Object Segmentation. 2018.
55. Karaaslan, E., U. Bagci, and F. N. Catbas. Attention-Guided Analysis of Infrastructure Damage with

Semi-Supervised Deep Learning. *Automation in Construction*, Vol. 125, No. February, 2021, p. 103634. https://doi.org/10.1016/j.autcon.2021.103634.
56. Karaaslan, E., M. Zakaria, and F. N. Catbas. Mixed Reality-Assisted Smart Bridge Inspection for Future Smart Cities. *The Rise of Smart Cities*, 2022, pp. 261–280. https://doi.org/10.1016/B978-0-12-817784-6.00002-3.
57. Krizhevsky, A., I. Sutskever, and G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. 2012.
58. Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 2014.
59. Jaccard, P. THE DISTRIBUTION OF THE FLORA IN THE ALPINE ZONE. *New Phytologist*, 1912. https://doi.org/10.1111/j.1469-8137.1912.tb05611.x.
60. Dice, L. R. Measures of the Amount of Ecologic Association Between Species. *Ecology*, Vol. 26, No. 3, 1945, pp. 297–302. https://doi.org/10.2307/1932409.
61. Geron, A. *Hands-On Machine Learing With Scikit-Learn & Tensor Flow*. 2017.
62. De, S., A. Yadav, D. Jacobs, and T. Goldstein. Automated Inference with Adaptive Batches. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, 2017.
63. Torosdagli, N., D. K. Liberton, P. Verma, M. Sincan, J. S. Lee, and U. Bagci. Deep Geodesic Learning for Segmentation and Anatomical Landmarking. *IEEE Transactions on Medical Imaging*, 2019. https://doi.org/10.1109/TMI.2018.2875814.
64. Andrychowicz, M., M. Denil, S. Gómez, M. W. Hoffman, D. Pfau, T. Schaul, B. Shillingford, and N. de Freitas. Learning to Learn by Gradient Descent by Gradient Descent. In *Advances in Neural Information Processing Systems 29* (D. D. Lee, M. Sugiyama, U. V Luxburg, I. Guyon, and R. Garnett, eds.), Curran Associates, Inc., pp. 3981–3989.
65. Kingma, D. P., and J. Ba. Adam: A Method for Stochastic Optimization. 2014.
66. Schmidhuber, J. Deep Learning in Neural Networks: An Overview. *Neural Networks*.
67. Karaaslan, E., U. Bagci, and C. F. Necati. Attention-Guided Analysis of Infrastructure Damage with Semi-Supervised Deep Learning. *Automation in Construction*, 2021.
68. Fan, J., T. Huo, and X. Li. A Review of One-Stage Detection Algorithms in Autonomous Driving. *2020 4th CAA International Conference on Vehicular Control and Intelligence, CVCI 2020*, 2020, pp. 210–214. https://doi.org/10.1109/CVCI51460.2020.9338663.
69. Shelhamer, E., J. Long, and T. Darrell. Fully Convolutional Networks for Semantic Segmentation. 2016.
70. Ronneberger, O., P. Fischer, and T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015.
71. Badrinarayanan, V., A. Kendall, and R. Cipolla. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 12, 2017, pp. 2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615.
72. LaLonde, R., and U. Bagci. Capsules for Object Segmentation. *arXiv*, No. Midl, 2018, pp. 1–9.
73. Karaaslan, E., C. F. Necati, and U. Bagci. A Novel Decision Support System for Long Term Management of Bridge Networks. *Applied Sciences*, 2021.
74. ImageNet. https://www.image-net.org/update-mar-11-2021.php. Accessed Jul. 31, 2022.
75. Execution Providers - Onnxruntime. https://onnxruntime.ai/docs/execution-providers/. Accessed Oct. 29, 2022.
76. Post-Training Quantization | TensorFlow Lite. https://www.tensorflow.org/lite/performance/post_training_quantization. Accessed Oct. 29, 2022.
77. Kato, H., M. Billinghurst, I. Poupyrev, K. Imamoto, and K. Tachibana. Virtual Object Manipulation on a Table-Top AR Environment. 2000.
78. Simon, G., A. W. Fitzgibbon, and A. Zisserman. Markerless Tracking Using Planar Structures in the Scene. 2000.

79. Ferrari, V., T. Tuytelaars, and L. Van Gool. Markerless Augmented Reality with a Real-Time Affine Region Tracker. 2001.
80. Strasdat, H., J. M. M. Montiel, and A. J. Davison. Visual SLAM: Why Filter? *Image and Vision Computing*, 2012. https://doi.org/10.1016/j.imavis.2012.02.009.
81. Gao, Q. H., T. R. Wan, W. Tang, L. Chen, and K. B. Zhang. An Improved Augmented Reality Registration Method Based on Visual SLAM. 2017.
82. Hsieh, C. H., and J. Der Lee. Markerless Augmented Reality via Stereo Video See-through Head-Mounted Display Device. *Mathematical Problems in Engineering*, 2015. https://doi.org/10.1155/2015/329415.
83. VisionStar Information Technology. EasyAR.
84. Unity Technology. Unity 3D. *Unity Technology*.
85. Arthur Juliani. Introducing: Unity Machine Learning Agents – Unity Blog. *Unity*.
86. Marchand, E., H. Uchiyama, F. Spindler, E. Marchand, H. Uchiyama, and F. Spindler. Pose Estimation for Augmented Reality : A Hands-on Survey Pose Estimation for Augmented Reality : A Hands-on Survey. Vol. 22, No. 12, 2016, pp. 2633–2651. https://doi.org/10.1109/TVCG.2015.2513408.
87. Chen, S., L. Liang, W. Liang, and H. Foroosh. 3D Pose Tracking with Multitemplate Warping and SIFT Correspondences. *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 26, No. 11, 2016, pp. 2043–2055. https://doi.org/10.1109/TCSVT.2015.2452782.
88. STRUCTURE Magazine | Inside the Bridge Inspection Toolbox. https://www.structuremag.org/?p=8064. Accessed Oct. 29, 2022.

APPENDIX

Sidebar Info

Program Steering Committee: NCHRP IDEA Program Committee

Month and Year: February 2023

Title: Mixed Reality Assisted Infrastructure Inspections

Project Number: 222

Start Date: April 1, 2022

Completion Date: December 30, 2022

Product Category: New or improved tool or equipment

Principal Investigator: F. Necati Catbas, Ph.D., PE

E-Mail: catbas@ucf.edu

Phone: 407-823-3743

TITLE:

Mixed Reality Assisted Infrastructure Inspections

SUBHEAD:

This project developed an AI-based application in Mixed Reality platforms assisting inspectors in the localization and quantification of concrete surface defects.

WHAT WAS THE NEED?

Conventional methods of inspection are subjective, labor-intensive, time-consuming, and often require a road closure. The inspectors use simple handheld tools for the quantification of the defects. Moreover, studies on inspection reports indicate that the results are often inconsistent between the inspectors, especially for medium to severe defects. AI-based methods of defect localization and quantification can provide more efficient, can potentially reduce the subjectivity of the results, and improve accuracy as a complementary technology. However, in order to reach the target accuracy, there is a need for more data to be collected and used for training the AI. Human involvement is still necessary while using these models due to uncertainties with the performance. On the other hand, post-processing of the inspection results can also be labor intensive and lead to multiple visits to the site which reduces the efficiency of AI-based inspection.

WHAT WAS OUR GOAL?

Our goal was to develop real time AI-based inspection technology to allow the inspector to oversee and approve its results. Rather than removing the human involvement, we wanted to benefit from the inspector's expertise in AI-based inspection using Human-AI collaboration in the Mixed Reality platform.

WHAT DID WE DO?

In this project, an AI-based application was developed that can be deployed in Mixed Reality devices to benefit from the advantages of AI models while allowing the inspector to oversee the results and intervene if necessary. Rather than a fully automated system and post-processing the data, the inspector continuously interacts with the AI model, approves its results, or corrects its error. The developed AI model in this project also follows an attention-guided method where the localization and quantification of the defects are carried out separately, which both improves the accuracy of the results and increases the inspector's involvement.

WHAT WAS THE OUTCOME?

The application developed in this project was successfully deployed in a Mixed Reality headset and tested. One of the main challenges of AI-based inspection in real-time is the processing power. This project successfully displayed that the optimized model can perform in real-time in certain Mixed reality devices. It was also shown that human involvement in the inspection can improve the accuracy of the AI and bring the post-processing time to a minimum.

WHAT IS THE BENEFIT?

The developed methodology improves the accuracy of inspections by facilitating the use of AI models for visual inspection. It removes the subjectivity of the inspection reports by accurately measuring the defect sizes and also improves the documentation of the inspection results by benefiting from the Mixed Reality platform. The device allows the inspector to access hard-to-reach areas, therefore reducing risks of inspection. It is also faster and more accurate than using handheld tools which is the common practice in visual inspection.

LEARN MORE

United States Patent for Methods of Artificial Intelligence-Assisted Infrastructure Assessment Using Mixed Reality Systems

https://patentimages.storage.googleapis.com/a2/a1/9a/021a9900d51e64/US11551344.pdf

2019 Kikuchi Karlaftis Best Paper Award Presentation

https://www.youtube.com/watch?v=mg9tEf7E2ls&t=379s

A field implementation video from the early implementation on a railroad bridge.

https://www.youtube.com/watch?v=d9LPyzHDRq4&t=15s

IMAGES

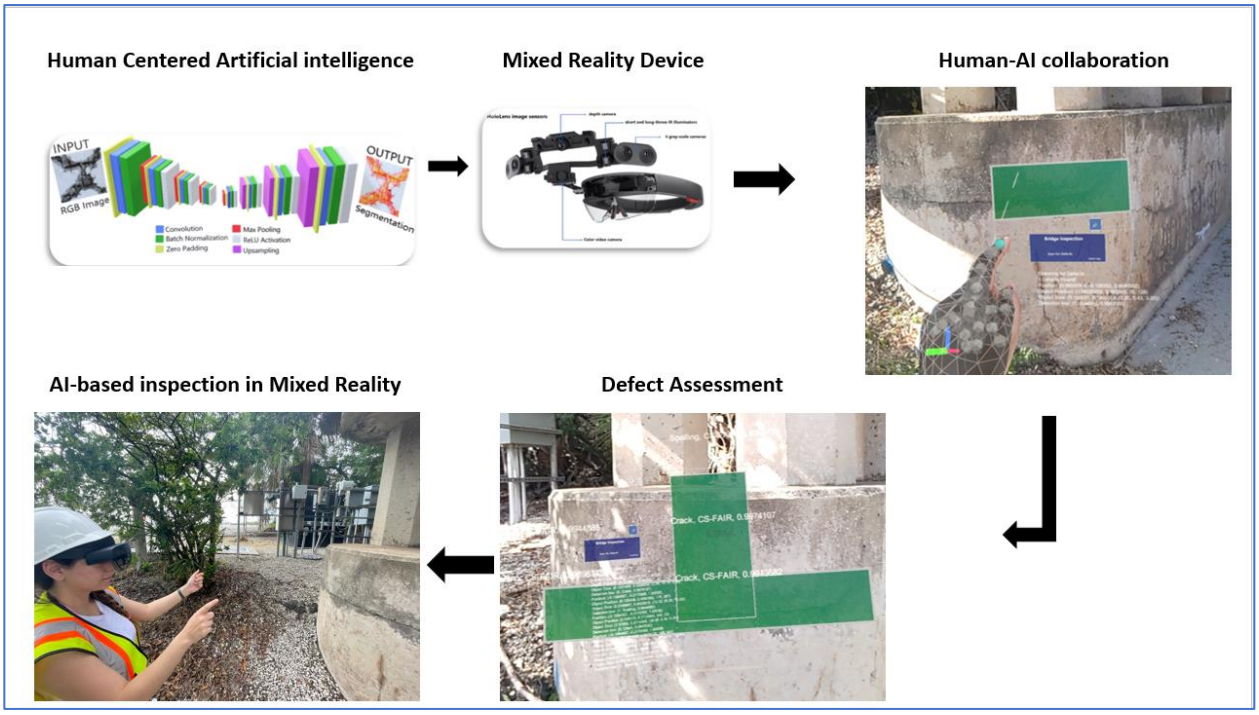

Figure. 1 An overview of the developed technology: AI-based inspection in Mixed Reality Platform.