CTIPS Center for Transformative Infrastructure Preservation and Sustainability
USDOT Region 8 University Transportation Center

Developing Context-Aware Computer Vision Models for Robust Data-Informed Condition Assessment of Bridges

CTIPS-032
Approved 9/15/2024

University

Utah State University

Principal Investigators

Mohsen Zaker Esteghamati, Ph.D.
Assistant Professor
Dept. of Civil and Environmental Engineering
Phone: (435)797-1522
Email: mohsen.zaker@usu.edu
ORCID: 0000-0002-2144-2938

Research Needs

Visual inspection (Figure 1.a) at regular intervals has traditionally been the primary method for assessing the condition of physical assets to ensure they meet intended performance objectives [1]. However, this method is labor-intensive and costly, poses safety risks to inspectors, and might be prone to quality inconsistencies [2]. These challenges have motivated the implementation of new inspection technologies in transportation asset management, generating abundant asset data from various sources such as drones (Figure 1.b), satellite imagery, LiDAR, and point-cloud scanners [3,4]. As such, there is a pressing need to develop "automatable" and "reliable" methodologies to leverage this multi-faceted data for enhancing transportation asset condition and performance.

Computer vision (CV)–based techniques provide an efficient approach to processing such data, enabling a high-level understanding of images and videos. CV models transform images (or videos) into pixel-wise mathematical functions to infer patterns, extract features, and provide a description of the image [5]. As shown in Figure 1.c, CV methods encompass a broad spectrum from low-level processing (e.g., noise reduction, contrast enhancement) to higher-level processing (feature extraction and object detection) [1]. Consequently, CV methods have been widely used in various infrastructure monitoring applications such as bridges, road networks, pavements, tunnels, and pavements. For example, algorithms have been proposed for surface

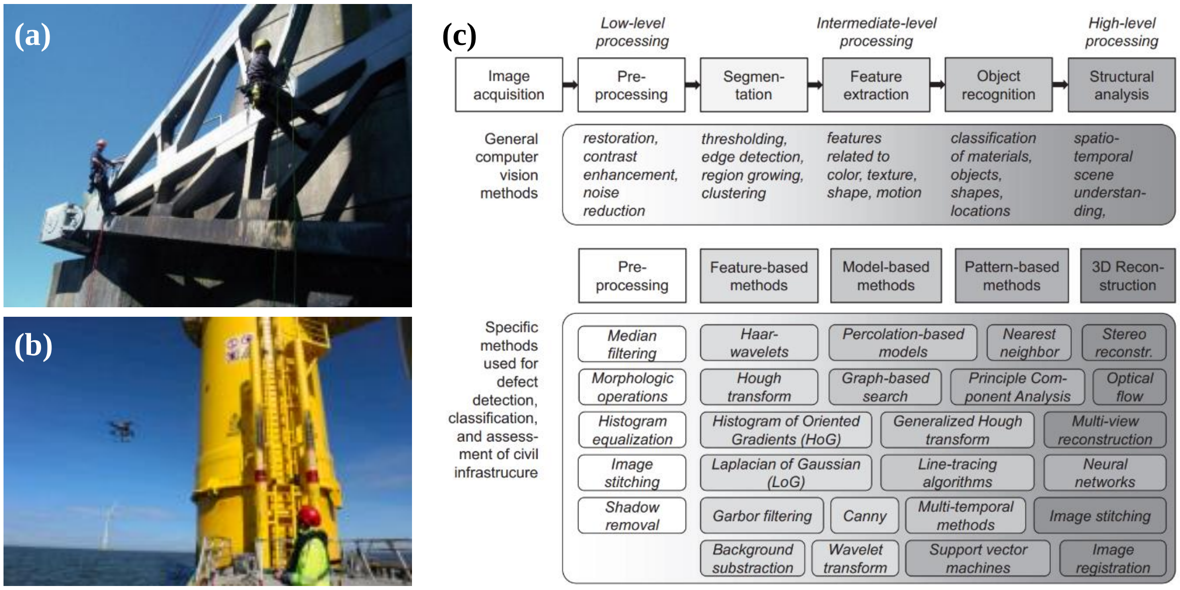

Figure 1. (a) conventional inspection [3], (b) drone-based inspection [4], (c) comparison of CV approaches and methods [1]

defect detection of bridges. Zhang et al. developed a pixel-level crack detection for concrete surfaces based on U-net architecture, albeit the model was dependent on the quality of images [6]. Some studies use the local directional evidence method to improve crack detection for low-contrast images [7]. Dung et al. used CV methods to detect gusset plates of steel bridge joists using convolutional neural networks (CNNs) [8].

Despite the advantages of CV-based techniques, significant challenges exist, including poor learning due to high data imbalance in training datasets (particularly for training sets focused on structural failure and damage) and excessive inference due to model complexity [9]. In addition, these models often ignore the "context" of the collected data, which results in limited applicability and generalizability. Here, context refers to any ancillary information (visual or non-visual) regarding a specific object of interest, or other objects present in the image [10]. Such contexts can play a critical role in CV models, as humans and machines process context differently. Figure 2 shows different categories of context applicable to CV models [10]. Previous research suggests that augmenting CV algorithms with contexts increases accuracy, albeit the gain depends on the context type [11].

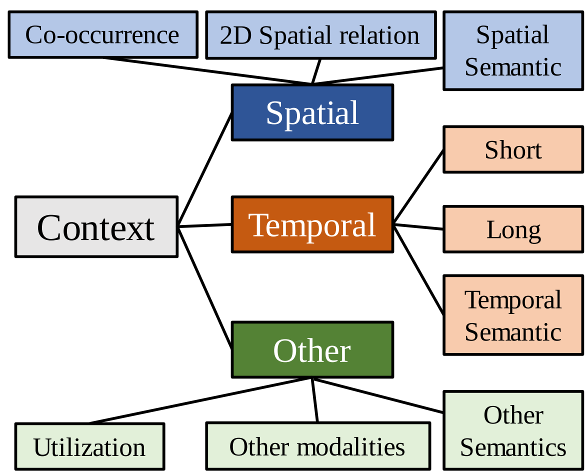

Figure 2. Categories of context and their sub-categories (inspired by [10])

This research aims to develop robust context-aware CV-based models with low inference time that can provide practice-oriented insights on the condition of monitored assets. Examples of such contexts include the co-occurrence of specific structural details of a bridge, which could help improve the accuracy of object detection tasks. Such context can also elevate regular CV tasks into inferences suitable for inspectors. Understanding the spatial relationship between the structural details of a bridge while identifying damage at a specific detail could be translated into scenarios for damage progression, and possible comprehensive remedies to prevent such mechanisms. As part of this project, various spatial and temporal contexts will be examined to understand the types of contexts that enhance the performance of CV models across different inspection tasks. In this initial phase of the study, the project will focus on steel bridges. However, the developed technology will be applicable to other transportation assets such as barriers, walls, and signs. These context-aware CV models will also create opportunities for integration with other technologies, such as geographic information systems or the Internet of Things, for holistic and integrated transportation asset management systems.

Research Objectives

This proposal aims to develop CV-based technologies to automate the inspection and monitoring of steel bridges. The developed models will be robust across different situations (e.g., lighting, orientation) and provide high-level, practice-oriented insight about assets. To this end, this project has three main objectives:

1. Investigate the relationship between different spatial and temporal contexts and the performance of CV tasks for asset condition assessment.
2. Develop a CV-based framework that can leverage the ancillary information of contexts to derive a more accurate and comprehensive description of the bridges condition.
3. Compare the performance of context-aware and conventional CV models across different tasks (e.g., object detection).

Research Methods

The research plan consists of four main tasks as follows:

Task 1. Literature review: An extensive literature review on applying CV in bridge condition assessment will be performed. This review aims to identify the state-of-the-art algorithms and methodologies for feature extraction, object detection, and image segmentation relevant to this domain of transportation data.

Task 2. Data collection and processing: An initial dataset will be compiled based on open datasets (e.g., COCO-Bridge [12]). Through active collaboration with UDOT, context-rich asset data (e.g., different lighting conditions and backgrounds, various structural details in the same image) will be collected and integrated into this initial set. The PI also has a different UDOT project on benchmarking CVs, which facilitates data sharing and coordination. As part of this task, undergraduate students will help the graduate student with image annotations, including contextual information. Pre-processing data is critical for the success of the CV models, and a wide range of techniques, such as noise reduction or contrast enhancement and database augmentation, will be used. In particular, generative adversarial networks will be examined for data augmentation.

Task 3. CV Model development: A CV workflow will be developed to extract features from processed and annotated asset data, train CV models, and tune their hyperparameters. A wide array of algorithms (identified through Task 1) will be evaluated. In particular, recurrent (for temporal contexts) and graph (for spatial contexts) neural networks will be examined. Additionally, CNN-based architectures such as U-Net and YOLO will be examined. The transfer learning concept will be employed to improve model performance by leveraging pre-trained models. The model's hyperparameters and architecture will be optimized to balance model complexity and accuracy.

Task 4. Integrating contextual information: A key aspect of this research is to fuse contextual information into computer vision models. Therefore, a methodology will be developed to integrate contextual information into the CV framework, including the importance of structural details, picture visibility, and spatial relationships between different structural details. Attention mechanisms (self and spatial) will be explored to understand the most effective approach for leveraging contexts. In addition, context fusion at different stages of model development (as part of model features or in combination with the model output) will be studied. This task will also focus on understanding how different types of contexts (Figure 2) can affect CV models' accuracy and insights. In particular, efforts will be directed at spatial contexts, such as spatial semantics and the relationship of different existing details across images of bridges. The context-aware models will be designed to be "adaptive", where updating techniques will be proposed to alter network features weight or threshold for defect detection with the supplemental information from contexts.

Task 5. Evaluation: The performance of context-aware CV models will be evaluated using various accuracy metrics (e.g., scalar metrics for classification, receiver operating characteristic curve) on unseen data using cross-validation. These data will be sampled from the collected data in a way that does not share information (i.e., data leak) with the training and testing sets. In addition, the models will be benchmarked with existing literature and conventional CV models.

Task 6. Final report: A final report will document the project outcome and methodology (Tasks 1-5) to disseminate the methodology to the research community and other transportation stakeholders.

Relevance to Strategic Goals

This project aims to develop technology-based solutions to improve inspection quality through automation. As such, this project supports two USDOT strategic goals as follows:

1. **Transformation (Primary goal):** The project develops transformative CV technologies that substantially advance the current state of automated inspections of transportation assets to extract better engineering-oriented inferences. Such a purpose-driven innovation will modernize current inspection methods and better meet future challenges.
2. **Safety (Secondary goal):** The developed technology aids with improved and automated understanding of bridge performance issues that could pose safety risks to asset users, allowing for better maintenance and repair strategies that subsequently make these assets safer for all people.

3. **Economic competitiveness:** The proposed research provides a low-cost alternative to traditional inspection methods, providing economic competitiveness for the next generation of inspection methods.

Educational Benefits

One graduate student in the PI's research group will conduct the proposed research. In addition, two undergraduate students will be involved in several parts of this project, particularly data collection and processing. The students will be mentored continuously and participate in national or regional conferences to present the project results. The PI will also attend USU summer programs to present the project outcome to engage and interest prospective high school students in pursuing a degree in civil engineering.

Outputs through Technology Transfer

The results of this research will develop an improved inspection technology, which will be published as one peer-reviewed research publication and conference presentation. All project data and models will also be published in open repositories (including USU digital commons) to allow other researchers, professionals, and practitioners to leverage the developed technology. The PI will seek opportunities to discuss the project outcome with UDOT (particularly the maintenance group) to seek opportunities for improving the current practices/methods on inspection methods that are not "boots on the ground".

Expected Outcomes and Impacts

This research develops an innovative CV-based framework that emulates human perception by understanding context in visual data, and providing actionable insights for decision-making. The proposed framework will provide real-time, context-sensitive information regarding the structural integrity and operability of bridges, suitable for risk-informed maintenance strategies. In the short term, the project will advance the current state of automated inspection frameworks, ensuring accuracy and scalability. In the long term, such frameworks can be integrated into broader technological systems for integrated maintenance scheduling. Additionally, all the developed data and models will be made public, which can be used by other researchers and DOT personnel across Region 8.

Work Plan

The proposed research plan will be carried out over 24 months as shown in Table 1.

Project Cost

Total Project Costs:	$196,083.61
CTIPS Funds Requested:	$ 98,041.61
Matching Funds:	$ 98,042.00
Source of Matching Funds:	Utah State University

Table 1 – Work plan

Task	**Duration** (months)	**Milestones** (outcomes/technology transfer)
1. Literature review	1.5	
2. Data collection and processing	4	
3. Model development	7	• Initial discussion with UDOT
4. Integrating contextual information	7	
5. Evaluation	4	• Open data/models published in open repository • Submission of journal article
6. Final report	1	• Technical report describing the developed technology and application • Conference presentation

References

[1] Koch C, Georgieva K, Kasireddy V, Akinci B, Fieguth P. A Review on Computer Vision Based Defect Detection and Condition Assessment of Concrete and Asphalt Civil Infrastructure. Advanced Engineering Informatics 2015;29:196–210. https://doi.org/10.1016/j.aei.2015.01.008

[2] Seo J, Duque L, Wacker J. Drone-Enabled Bridge Inspection Methodology and Application. Automation in Construction 2018;94:112–26. https://doi.org/10.1016/j.autcon.2018.06.006

[3] Spencer BF, Hoskere V, Narazaki Y. Advances in Computer Vision-Based Civil Infrastructure Inspection and Monitoring. Engineering 2019;5:199–222. https://doi.org/10.1016/j.eng.2018.11.030

[4] Yang C, Liu X, Zhou H, Ke Y, See J. Towards Accurate Image Stitching for Drone-Based Wind Turbine Blade Inspection. Renewable Energy 2023;203:267–79. https://doi.org/10.1016/j.renene.2022.12.063

[5] Paneru S, Jeelani I. Computer Vision Applications in Construction: Current State, Opportunities & Challenges. Automation in Construction 2021;132:103940. https://doi.org/10.1016/j.autcon.2021.103940

[6] Zhang L, Shen J, Zhu B. A Research on an Improved Unet-Based Concrete Crack Detection Algorithm. Structural Health Monitoring 2021;20:1864–79. https://doi.org/10.1177/1475921720940068

[7] Zhang H, Tan J, Liu L, Wu QMJ, Wang Y, Jie L. Automatic Crack Inspection for Concrete Bridge Bottom Surfaces Based on Machine Vision. Chinese Automation Congress 2017, p. 4938–43. https://doi.org/10.1109/CAC.2017.8243654

[8] Dung CV, Sekiya H, Hirano S, Okatani T, Miki C. A Vision-Based Method for Crack Detection in Gusset Plate Welded Joints of Steel Bridges Using Deep Convolutional Neural Networks. Automation in Construction 2019;102:217–29. https://doi.org/10.1016/j.autcon.2019.02.013

[9] Ai D, Jiang G, Lam S-K, He P, Li C. Computer Vision Framework for Crack Detection of Civil Infrastructure—A Review. Engineering Applications of Artificial Intelligence 2023;117. https://doi.org/10.1016/j.engappai.2022.105478

[10] Wang X, Zhu Z. Context Understanding in Computer Vision: A Survey. Computer Vision and Image Understanding 2023;229:103646. https://doi.org/10.1016/j.cviu.2023.103646

[11] Katti H, Peelen MV, Arun SP. Machine Vision Benefits from Human Contextual Expectations. Scientific Reports 2019;9:2112. https://doi.org/10.1038/s41598-018-38427-0

[12] Bianchi E, Abbott AL, Tokekar P, Hebdon M. COCO-Bridge: Structural Detail Data Set for Bridge Inspections. Journal of Computing in Civil Engineering 2021;35:04021003. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000949