Application Aware, Life-Cycle Oriented
Model-Hardware Co-Design Framework
for Sustainable, Energy Efficient ML Systems

Use case validations - Initial

Deliverable D6.1

WP6 - Application Oriented Validation

This project has received funding from the European Union's Horizon Europe research and innovation programme (HORIZON-CL4-2021-HUMAN-01) under grant agreement No 101070408

Project

Title:	SustainML: Application Aware, Life-Cycle Oriented Model-Hardware Co-Design Framework for Sustainable, Energy Efficient ML Systems
Acronym:	SustainML
Coordinator:	eProsima
Grant agreement ID:	101070408
Call:	HORIZON-CL4-2021-HUMAN-01
Program:	Horizon Europe
Start:	01 October 2022
Duration:	36 months
Website:	https://sustainml.eu
E-mail:	sustainml@eprosima.com
Consortium:	**eProsima (EPROS)**, Spain **DFKI**, Germany **TU Kaiserslautern (TUK)**, Germany **University of Copenhagen (KU)**, Denmark **National Institute for Research in Digital Science and Technology (INRIA)**, France **IBM Research GmbH**, Switzerland **UPMEM**, France

Deliverable

Number:	**D6.1**
Title:	**Use case validations - Initial**
Month:	24
Work Package:	WP6 - Application Oriented Validation
Work Package leader:	IBM
Deliverable leader:	IBM
Deliverable type:	R
Dissemination level:	Public (PU)
Date of submission:	30/09/2024
Version:	v1.0
Status:	Finished

Version history

Version	Date	Responsible	Author/Reviewer	Comments
v1.0	30/09/2024	IBM	IBM	

Executive summary

SustainML project aims to develop a design framework and an associated toolkit, so-called SustainML, that will foster energy efficiency throughout the whole life-cycle of Machine Learning (ML) applications: from the design and exploration phase that includes exploratory iterations of training, testing and optimizing different system versions through the final training of the production systems (which often involves huge amounts of data, computation and epochs) and (where appropriate) continuous online re-training during deployment for the inference process. The framework will optimize the ML solutions based on the application tasks, across levels from hardware to model architecture. It will also collect both previously scattered efficiency-oriented research, as well as novel Green-AI methods. Artificial Intelligence (AI) developers from all experience levels can make use of the framework through its emphasis on human-centric interactive transparent design and functional knowledge cores, instead of the common blackbox and fully automated optimization approaches.

This report corresponds to *Deliverable D6.1 - Use case validations - Initial*6.1 of the SustainML project. The deliverable focuses on validating the automated visual inspection use case in civil engineering, specifically targeting bridge structure maintenance. One of the key challenges in this domain is the scarcity of high-quality, fine-grained labeled image datasets for inspecting crucial bridge components and detecting structural defects. The limited availability of such datasets hinders the training of large neural network models typically required for this task.

In order to solve this task, we present IBM Inspecto—a collaborative effort between research and industry aimed at fully digitizing the visual inspection process. This initiative leverages foundation models and cutting-edge computer vision techniques to improve the speed and accuracy of essential inspection tasks.

Contents

1 Introduction

The maintenance and inspection of infrastructure elements such as bridges, tunnels, roads, runways, and dams are crucial for ensuring public safety and sustaining economic activity. Despite their significance, visual inspections are often performed manually, posing considerable challenges for infrastructure operators tasked with identifying potential issues before they escalate into costly repairs or dangerous conditions. This reliance on manual inspection not only lengthens maintenance cycles but also limits the thoroughness of evaluations, as inspectors may only report major detected problems and provide vague locational references. Consequently, this practice can lead to critical oversights, impeding the ability to verify historical inspections or pinpoint the emergence of new defects. The growing age of infrastructure exacerbates these challenges, heightening the risk of catastrophic failures, as exemplified by the tragic collapse of the Morandi Bridge in Italy in 2018.

In recent years, advancements in AI and ML, particularly in deep learning and Foundation Models (FMs), have transformed various sectors, including computer vision and natural language processing. While the civil engineering field has begun to embrace these technologies—evidenced by their application in condition assessments, novelty detection, and structural health monitoring—the integration of FMs into operational workflows remains limited. This gap between research advancements and practical applications stems from challenges in translating model outputs into actionable insights within existing inspection processes.

To address these challenges, our project leverages knowledge recycling techniques, particularly data augmentation and transfer learning, to develop data-efficient neural networks capable of fine-tuning for detailed segmentation using minimal curated annotated datasets. The validation scenario focuses on automating visual inspection within civil engineering, where the scarcity of labeled image datasets presents significant obstacles. In light of this, we introduce IBM Inspecto, a research-industry collaboration designed to fully digitalize the visual inspection process, harnessing FMs and advanced computer vision methods to enhance both the speed and accuracy of critical inspection tasks. We perform data collection thanks to our collaborating partner Sund & Bælt and in particular we use the data of the Storebælt Bridge.

This deliverable is structured as follows: Section 2 explores the advantages of AI foundation models for visual inspections in civil engineering, while Section 3 details the IBM Inspecto platform and how it facilitates the practical application of AI insights for domain experts. Finally, Section 4 outlines future directions and concludes our discussion on the implications of this work for the civil engineering industry.

2 Foundation Models for Visual Inspection

Deep learning computer vision models excel at analyzing large image datasets, particularly those collected via drones, to detect high-risk features such as structural cracks or visible corrosion in bridge steel reinforcement bars. The effectiveness of these models is closely linked to the availability of substantial volumes of annotated data. However, obtaining high-quality labels for rare occurrences, such as structural defects, poses significant challenges and costs. This issue is further complicated in visual inspection contexts, where the infrequency of defects and the complexity of tasks like object detection and instance segmentation elevate the difficulty and expense of labeling.

Recently, Self-Supervised Learning (SSL) has emerged as a crucial technique in computer vision, especially for pre-training FMs on large datasets that lack labels. SSL is gaining traction in civil engineering applications, particularly for visual inspection tasks. The automation of data collection methods, such as drone imaging over bridges, has led to the availability of large unlabeled datasets, making them prime candidates for SSL. In this approach, models learn to label data themselves through predefined pretext

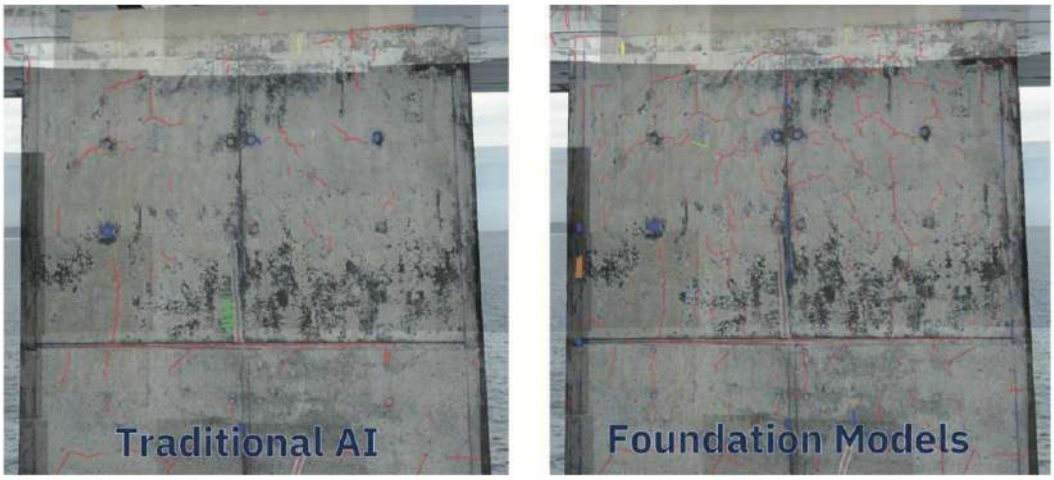

Figure 1: Comparison of AI outcomes on the south face of Storebaelt Pillar 3. On the left: results from traditional supervised learning. On the right: results from a foundation model (FM)

tasks. SSL techniques have evolved over the years, incorporating methods like denoising autoencoders and Siamese networks. More recent advancements include self-supervised pre-training methods such as PIRL [1], BYOL [2], MOCO [3], and SimCLR [4], further enhanced by vision transformer architectures exemplified by models like Vision Transformer (ViT) [5], DINO [6], MSN [7], and Masked Auto Encoder (MAE) [8]. The importance of SSL pre-training is twofold: it improves model performance and enables effective learning from limited examples in subsequent tasks.

In the realm of bridge defect detection, our experiments have concentrated on leveraging SSL pre-training to develop FMs. Specifically, we implemented the MAE technique, which involves masking a significant portion of an image and training the model to reconstruct the unmasked sections. This methodology enhances the model's understanding of the nuances of concrete bridge structures while improving its capacity to identify defects. Results indicate that an FM pre-trained on images of concrete bridges significantly enhances defect detection performance. This improvement is visually represented in qualitative assessment like the one in Figure 1, where FMs demonstrate a markedly higher recall than traditional models. Additionally, the use of large ViT models preserves image resolution more effectively than convolutional-based methods, allowing for better defect detection with reduced background noise. This refinement leads to more accurate measurements and prioritization of detected defects.

The training methodology we employed consists of a hierarchical, four-stage process. Initially, the model undergoes pre-training through SSL on a diverse collection of common object datasets, enabling it to abstract relevant information from a wide range of images, regardless of their direct relevance to civil infrastructure. In the second stage, the model continues SSL training, focusing on a "source dataset" that includes extensive unlabeled imagery predominantly featuring concrete structures, particularly bridges. This stage aims to refine the model's ability to extract salient features from images of concrete structures.

The third stage narrows the focus exclusively to the Storebælt Bridge, utilizing a specialized dataset composed entirely of images from this structure. This targeted approach is crucial for fine-tuning the model's understanding of the specific characteristics of Storebælt Bridge imagery. A subset of this dataset, consisting of approximately 1,000 images annotated with around 8,000 identified defects, is employed in the final training stage.

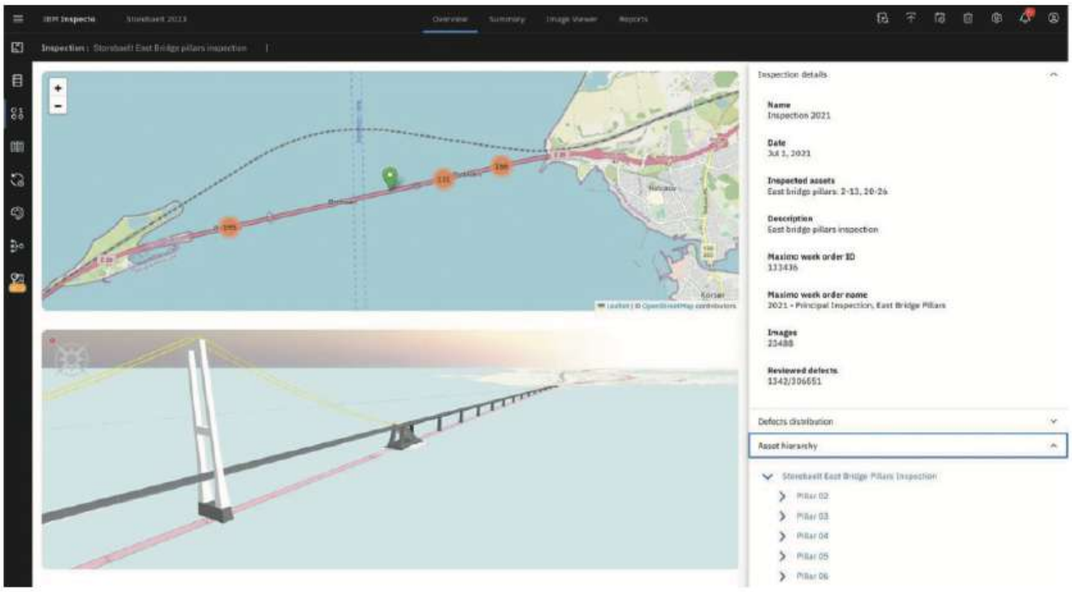

Figure 2: IBM Inspecto webpage

Figure 3: High-resolution analysis of the Storebælt Pillar 3, south face, bottom region, performed by IBM Inspecto. This view consists of 30 high-resolution images seamlessly stitched together by our algorithm. Our finely-tuned AI foundation model identifies various defects: cracks (red), net-cracks (orange), spalling (blue), rust (yellow), cracks with precipitation (pink), and algae (green).

Figure 4: High-resolution analysis of the Storebælt Pillar 3, south face, bottom region, by IBM Inspecto. Several zoomed-in areas are highlighted to demonstrate the accuracy of the masks generated by our AI foundation model, showcasing its precision in detecting and outlining defects.

In this last phase, the model transitions to supervised learning, using a cascade Mask R-CNN [9] head on top of the pre-trained backbone. This adaptation allows the model to perform both object detection—identifying bounding boxes around objects of interest—and instance segmentation—delineating the pixels within each box that correspond to defects. This comprehensive training approach ensures that the model is not only proficient in general image comprehension but also finely tuned to the intricacies of concrete structure imagery, particularly that of the Storebælt Bridge, enhancing its effectiveness in defect detection and segmentation tasks.

3 IBM Inspecto

IBM Inspecto is an innovative industry research platform designed for enterprise visual inspection, developed by IBM Research in collaboration with out partner civil engineering experts. This platform streamlines the visual inspection process by guiding engineers and domain specialists through every phase of engineering inspection. It addresses long-term data management, providing intuitive navigation through complex assets via advanced image stitching technology, mapping, hierarchical views, and 3D models. The system employs high-resolution defect detection capabilities powered by AI foundation models trained on the world's largest curated infrastructure image dataset. It also automates defect measurements, supports an assisted engineering review process, and allows for customizable report generation.

The platform features a comprehensive image viewer that enables structured access to large volumes of inspection images, facilitating easy organization. A prediction viewer displays results generated by AI models and extracts relevant attributes, such as crack length, to provide pixel-accurate segmentation masks of detected defects. Additionally, an overview viewer allows users to visualize extensive sections

Figure 5: IBM Inspecto results on the Stenungsund Bridge: the image highlights significant instances of spalling with corroded rebars, accurately detected by our model.

of infrastructure and navigate in real-time, giving context to defect locations. Merged predictions enable users to comprehend the relationships and spatial arrangements of defects while consolidating multiple detections across different images into a singular prediction. A statistical viewer aggregates data on detected defects, offering insights for individual images as well as for the overall merged view. Finally, the report generator creates customized reports based on the captured inspection data, detailing each defect along with essential attributes and direct links to an enlarged view of visualized defects.

To illustrate the capabilities of IBM Inspecto, we focus on its application in inspecting the Storebælt Bridge. Once the inspection data is uploaded and analyzed by our fine-tuned foundation models, engineers can initiate the visual inspection process from the overview page, which summarizes the inspection results, including a map of drone capture points and access to the bridge's 3D model and hierarchical structure. This overview enables engineers to navigate the entire infrastructure efficiently and gain insights into the inspected surfaces and the number of defects detected by AI.

From the overview page (Figure 2), civil engineers can delve into specific elements and closely examine the concrete surfaces. For instance, the high-resolution analysis of pillar 3 highlighted in Figure 3 of the Storebælt Bridge showcases the power of IBM Inspecto. Utilizing a grid mode with the DJI Matrice 300 drone, images are stitched together to provide a zoomable view of the structure. The AI foundation model detects various types of defects, including cracks, net-cracks, spalling, rust, and algae, marked in distinct colors for clarity as shown in Figure 4.

After reviewing the bridge surface, engineers can generate customizable reports that encompass all necessary information regarding the inspection. These reports are available in both HTML and PDF formats and include detailed summaries of detected defects, such as the one displayed for pillar 3, which highlights the most severe issues and provides links for further inspection.

Beyond the Storebælt Bridge, IBM Inspecto has successfully been applied to various other structures, such as the Stenungsund Bridge in Sweden and the Dübendorf Airbase runway. The Stenungsund Bridge (Figure 5), which reflects the condition of many aging bridges across Europe and the United States,

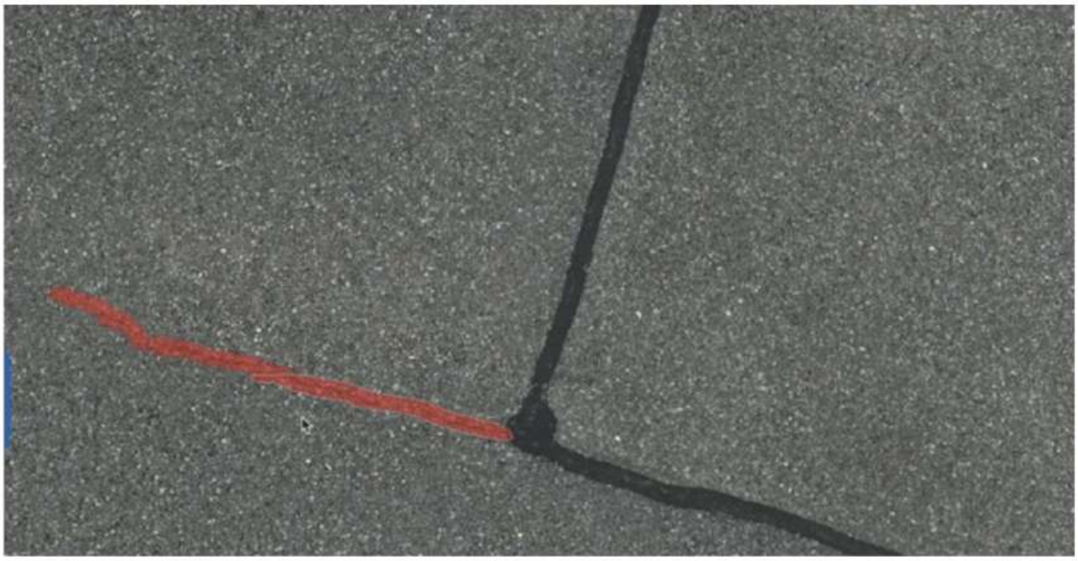

Figure 6: A close-up view of a crack at the Dubendorf Airbase that was previously repaired but has since continued to expand.

exhibits significant deterioration, including spalling and corroded rebars. By fine-tuning the model with approximately 500 images, we achieved high detection accuracy for these issues.

Similarly, the Dübendorf Airbase runway presented a different challenge compared to concrete bridge surfaces. Nevertheless, our algorithm effectively reconstructed a comprehensive stitched image of the runway, identifying even the smallest cracks without additional fine-tuning. An example of this is a crack that had expanded post-repair depicted in Figure 6, demonstrating the model's ability to detect new growth adjacent to previously treated areas precisely. Through these applications, IBM Inspecto showcases its versatility and efficacy in enhancing visual inspection processes across diverse infrastructure types.

4 Conclusions

IBM Inspecto is a cutting-edge research-industry platform developed by IBM Research that leverages advanced foundation models. Through our collaboration with Sund & Bælt, we have showcased the platform's application to the Storebælt Bridge and various other infrastructures. Key features include intuitive navigation, precise defect detection in high-resolution images, and automated generation of customizable reports. IBM Inspecto is designed to streamline the visual inspection of complex assets, significantly reducing both time and costs while fully digitizing the inspection process. Its versatility makes it suitable for a broad range of infrastructures, including bridges, buildings, roads, and runways.

Looking ahead, we are enhancing IBM Inspecto with additional AI functionalities aimed at further accelerating the inspection process. One of our major initiatives is the development of a defect review assistant powered by multi-modal foundation models. This assistant will learn from a carefully curated set of domain experts who manually review defects, enabling it to automate the review process for thousands of defects identified by our AI models. We anticipate this advancement will accelerate defect review by one to two orders of magnitude.

Lastly, we plan to incorporate features that allow users to compare inspected surfaces over time. Utilizing our stitched images, IBM Inspecto will conveniently highlight changes between inspections without necessitating re-annotation or reassessment of previously identified defects. This capability will further streamline future inspections, making the process even more efficient and user-friendly.

References

[1] Ishan Misra and Laurens van der Maaten. *Self-Supervised Learning of Pretext-Invariant Representations*. 2019. arXiv: 1912.01991 [cs.CV]. URL: https://arxiv.org/abs/1912.01991.

[2] Jean-Bastien Grill et al. *Bootstrap your own latent: A new approach to self-supervised Learning*. 2020. arXiv: 2006.07733 [cs.LG]. URL: https://arxiv.org/abs/2006.07733.

[3] Kaiming He et al. *Momentum Contrast for Unsupervised Visual Representation Learning*. 2020. arXiv: 1911.05722 [cs.CV]. URL: https://arxiv.org/abs/1911.05722.

[4] Ting Chen et al. *A Simple Framework for Contrastive Learning of Visual Representations*. 2020. arXiv: 2002.05709 [cs.LG]. URL: https://arxiv.org/abs/2002.05709.

[5] Alexey Dosovitskiy et al. *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. 2021. arXiv: 2010.11929 [cs.CV]. URL: https://arxiv.org/abs/2010.11929.

[6] Mathilde Caron et al. *Emerging Properties in Self-Supervised Vision Transformers*. 2021. arXiv: 2104.14294 [cs.CV]. URL: https://arxiv.org/abs/2104.14294.

[7] Mahmoud Assran et al. *Masked Siamese Networks for Label-Efficient Learning*. 2022. arXiv: 2204.07141 [cs.LG]. URL: https://arxiv.org/abs/2204.07141.

[8] Kaiming He et al. *Masked Autoencoders Are Scalable Vision Learners*. 2021. arXiv: 2111.06377 [cs.CV]. URL: https://arxiv.org/abs/2111.06377.

[9] Kaiming He et al. *Mask R-CNN*. 2018. arXiv: 1703.06870 [cs.CV]. URL: https://arxiv.org/abs/1703.06870.