*Bridge Maintenance, Safety, Management, Digitalization and Sustainability –
Jensen, Frangopol & Schmidt (eds)*


# AI for bridge inspection with IBM Inspecto

A.C.I. Malossi, R. Assaf, N. Avogaro, A. Bartezzaghi, B. Ebouky, T. Frick, I. Giurgiu, F.M. Janicki, P. Kluska, M. Rigotti, F. Scheidegger, A. Simeski & C. Skura
*IBM Research Europe – Zurich Laboratory, Rüschlikon, Switzerland*

D. Caraballo & Y. Cinar
*IBM Research Europe – Paris Laboratory, Orsay, France*

F. Bormlund & S. Gjerding
*Sund & Bælt Holding A/S, Korsør, Denmark*

ABSTRACT: We present our research innovation in enterprise visual inspection for civil infrastructures. We train and adapt AI foundation models to work on high-resolution images collected by drones for detecting and segmenting very small defects in large areas and with high precision. Further, we provide support to reliability engineers by measuring defects with computer vision and by providing additional information that is required to assess damage severity and plan maintenance interventions. We provide access to our research innovation through our SaaS IBM Inspecto. IBM Inspecto is a research-industry platform developed by IBM Research that leverages foundation models and advanced computer vision methods to fully digitalize the visual inspection process, accelerating and improving accuracy in critical inspection tasks. In this work, we present visual inspection results on the Storebælt Bridge in collaboration with Sund & Bælt, as well as on other applications.

## 1 INTRODUCTION

The maintenance and inspection of infrastructure elements, such as bridges, tunnels, roads, runways, and dams, is of immense significance for public safety and for maintaining economic activity. Despite that, most visual inspections are carried out manually, and infrastructure operators are faced with the challenging task of detecting every problem before it gets expensive to be repaired, or even too dangerous for civilians. Manual visual inspections make maintenance cycles longer than they should be, as they require that a person physically inspects the entire surface of an infrastructure. Civil engineers not only need to detect problems but also to review them and decide what action must be taken. As a result, our global network of infrastructures is inspected less frequently than it should be and, in some cases, inspection job plans are only partially executed. For instance, in today's inspection reports, only major detected problems are included, and their location is only roughly indicated. This may cause further issues downstream of the process when selected defects are repaired with imprecise information. Moreover, if something is not included in such reports, it is practically impossible to verify in the future if that was an actual mistake or if no problem was visible at that time. With the infrastructure network becoming older and older every year, all these challenges risk causing major fatalities, as we have already seen in examples such as the collapse of the Morandi Bridge in Italy in 2018[1].

The last decade has marked incredible technical progress in AI and Machine Learning (ML) with remarkable success stories in the fields of computer vision, natural language pro-

1. https://en.wikipedia.org/wiki/Ponte_Morandi

DOI: 10.1201/9781003483755-421

cessing, and other domains. This wave of success and excitement in AI has been in large part driven by Deep Learning (LeCun 2015), and recently by Foundation Models (FMs), i.e., deep neural networks that are trained on large broad unlabeled datasets and that can be deployed on a wide range of downstream tasks.

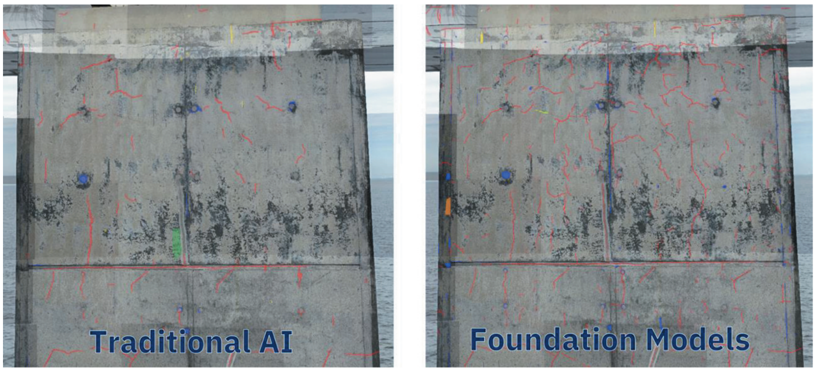


Figure 1. Comparison of AI results on Storebaelt Pillar 3 south face. Left: traditional supervised learning. Right: FM.

While ML has seen increased adoption in the civil engineering domain in recent years, for instance in applications like condition assessment (Bianchi 2022), novelty detection for structural health monitoring (Manzini 2022), or other structural engineering tasks (Salehi 2018), the impact of FMs specifically on civil engineering is however still in its infancy, owing to a gap between research development and application use cases. Part of this gap is due to the non-straightforward adoption of model results into existing operational flows.

To close this gap, we present IBM Inspecto, a research-industry platform developed by IBM Research in collaboration with our engineering partner Sund & Bælt, that leverages FMs and advanced computer vision methods to fully digitalize the visual inspection process, accelerating and improving accuracy in critical inspection tasks. This paper is organized as follows: Section 2 introduces AI FMs and discusses their benefit for visual inspection of civil infrastructures, Section 3 presents IBM Inspecto and describes how domain experts can consume results of AI FMs in a convenient way. Finally, Section 4 describes future directions and conclusions of this work.

## 2 FOUNDATION MODELS FOR VISUAL INSPECTION

Deep learning computer vision models are adept at sifting through large image datasets typical acquired by use of drones to identify high-risk elements like structural cracks or visible corrosion in steel reinforcement bars in bridges. The accuracy and efficacy of these models hinge on access to a substantial volume of annotated data. However, obtaining high-quality labels, especially for rare occurrences like structural defects, is both challenging and expensive. This issue is accentuated in visual inspection scenarios, where the rarity of defects and the complexity of tasks like object detection and instance segmentation compound the difficulty and cost of labeling.

Recently, Self-supervised learning (SSL) has emerged as a pivotal approach in the realm of computer vision, especially for pre-training FMs on large amounts of unlabeled data. SSL is also becoming a cornerstone in civil engineering applications, particularly for the critical task of visual inspection. It is now common that large datasets are becoming available thanks to the automation of data collection methods, such as drone data acquisitions over bridges. This abundance of unlabeled data is an ideal candidate for

SSL, where models are trained on data that they label themselves through predefined pretext tasks. SSL's history in computer vision spans several years, featuring techniques like denoising autoencoders and Siamese networks. More recent innovations include self-supervised pretraining approaches like PIRL (Mirsa, 2019), BYOL (Grill, 2020), MOCO (He, 2020), and SimCLR (Chen 2020), further advanced by integrating vision transformer architectures, as seen in Vision Transformer (ViT) (Dosovitskiy, 2019), DINO (Zhang, 2022), MSN (Assran, 2022), and Masked Auto Encoder (MAE) (He, 2021). The significance of SSL pretraining is twofold: it not only improves the overall performance of models but also makes them adept at learning from a few examples in downstream tasks.

In the context of bridge defect detection, our experiments have focused on SSL pretraining to develop FMs. We utilized the MAE technique, where a significant portion of an image is masked, and the model is trained to reconstruct the unmasked image. This approach not only aids the model in understanding the intricacies of concrete bridge structures but also enhances its ability to pinpoint defects.

Results demonstrate that an FM pre-trained on images of concrete bridges boosts performance in defect detection tasks. This is further highlighted in Figure 1, where given a qualitative assessment we can see that FMs have a much higher recall over traditional models. Furthermore, due to the use of large ViT models, the image resolution is kept more intact when compared to convolutional-based methods. This further refines the detection and allows the model to identify the defects with less background noise, this can then lead to better measurements and therefore ranking of the defects.

### 2.1 *Detailing the training method*

In the preceding section, we mention that we developed an FM employing the MAE technique. This is executed through a hierarchical, four-stage training process. Initially, the model undergoes pre-training via SSL, devoid of any labeling, on a diverse collection of common object datasets. This phase is instrumental in enabling the model to hone its capability for abstracting pertinent information from a broad spectrum of images, irrespective of their relevance to civil infrastructure or bridges specifically.

Progressing to the second stage, the model's training continues under the SSL paradigm. It is during this phase that the model is exposed to a 'source dataset', comprising an extensive collection of unlabeled imagery predominantly featuring concrete structures, with a significant focus on bridges. This dataset encompasses a variety of structures, including the Storebælt Bridge, but extends to incorporate bridges from multiple geographical regions and other concrete edifices. The primary objective of this stage is to refine the model's proficiency in extracting salient features specifically from images of concrete structures.

The third stage of the model's training, adhering once again to SSL principles, narrows the focus exclusively to the Storebælt Bridge. The dataset in this stage is composed solely of images captured of this bridge, creating a specialized data domain that the model learns to navigate. This targeted approach is pivotal in fine-tuning the model's ability to interpret the specific characteristics inherent in the Storebælt Bridge imagery. Notably, a subset of this dataset, encompassing approximately 1000 images, is annotated with roughly 8000 identified defects, and is used in the fourth stage of training.

In the fourth and final stage, the model is trained using supervised learning. This phase employs a cascade Mask R-CNN head that sits on top of the pre-trained model backbone which was pre-trained in the previous steps. This adapts the model to i) object detection, i.e., identifying bounding boxes around objects of interest, and ii) instance segmentation, i.e., identifying the pixels within each box corresponding to defects. That multifaceted training approach ensures that the model is not only versed in general image comprehension but is also finely attuned to the nuances of concrete structure imagery, particularly that of the Storebælt Bridge, thereby enhancing its efficacy in defect detection and segmentation tasks.

## 3 IBM INSPECTO

IBM Inspecto is an industry research platform for enterprise visual inspection developed by IBM Research. With IBM Inspecto, engineers and domain experts are guided through the visual inspection process with support for all the phases of engineering inspection, including long-term management of large amounts of data, intuitive navigation of complex assets with unique image stitching technology, maps, hierarchy, and 3D models, high-resolution defect detection powered with AI FMs trained on the world largest curated infrastructure image dataset, automatic measurement of detected defects, assisted engineering review process and customizable report generation.

In more detail, IBM Inspecto features the following components:

- Image viewer: enable ordered access to large amounts of inspection images and in structured folders.
- Prediction viewer: displays the results of the AI models and automatically extracts associated attributes, such as crack length to represent pixel-accurate segmentation masks of defects.
- Overview viewer: allows the user to view large sections of the infrastructure and navigate in real-time. It also enables understanding of the context in which a defect was detected.
- Merged predictions and overview: enables the user to understand the relationships and positions of all defects and consolidates multiple detections of a defect in different images into a single prediction.
- Statistical viewer: aggregated statistics are provided on all detected defects, both for individual images and for the entire merged overview.
- Report generator: enables the generation of customized for all data captured in the inspection, with detailed views of each defect, important attributes, and direct links to OCL for an enlarged view of the visualized defect. Georeferencing allows runway maintenance staff to manually check defects on the runway and repair damages with a sealant.

### 3.1 *Great Belt bridge inspection*

In this section, we demonstrate IBM Inspecto on the Storebælt[2] bridge. After the data has been uploaded to the service and analyzed by our fine-tuned FMs, the engineer can start the

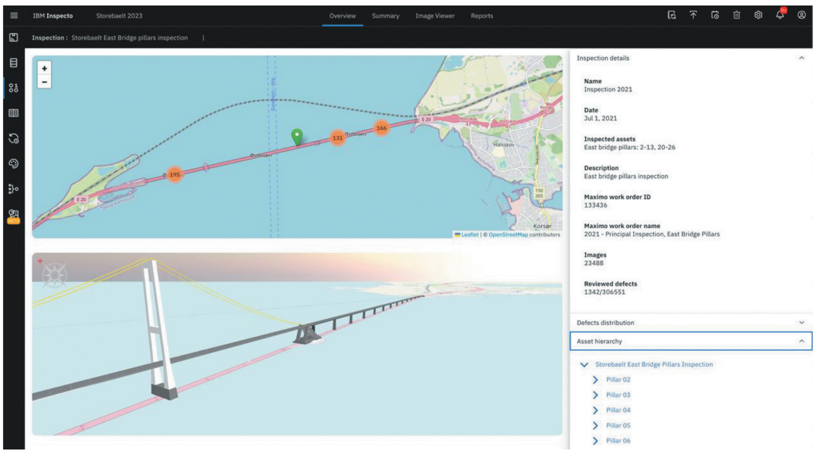


Figure 2. IBM Inspecto overview page. This page contains a summary of the performed visual inspection, including a map with the location of drone capturing points, the 3D model, as well as access to the bridge hierarchy and corresponding IBM Maximo work orders.

2. https://storebaelt.dk/en/

visual inspection process from the overview page in Figure 2: such a view conveniently allows the user to navigate around the entire infrastructure and have a first overview of the number of surfaces that have been inspected as well as of the number of defects detected by AI.

From the overview page, the civil engineer can access any element and inspect the concrete surface in detail. In Figures 3 and 4 we show an example view of one face of pillar 3 on the Storebælt. The stitched view is zoomable up to native high-resolution of the individual images, that are collected with the DJI Matrice 300 drone, using high-resolution grid mode.

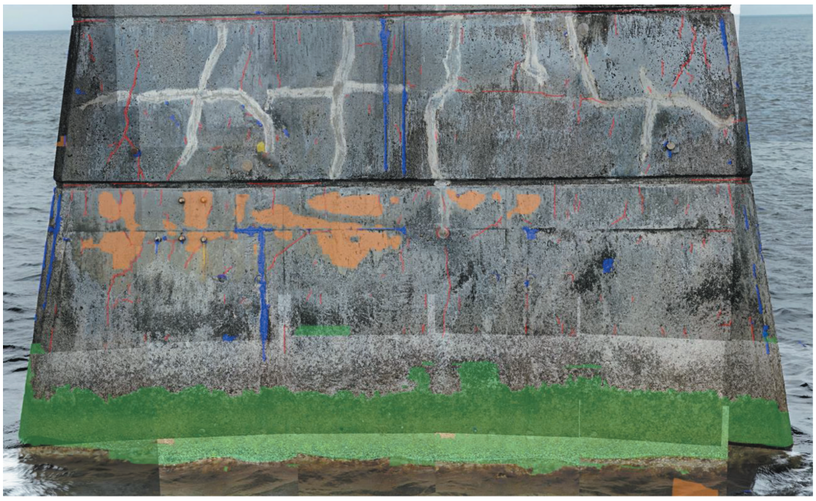

Figure 3. IBM Inspecto high-resolution analysis of the Storebælt pillar 3, south face, bottom region. The view is made of 30 high-resolution images automatically stitched together by our algorithm. Our fine-tuned AI FM detects cracks (red) net-cracks (orange), spalling (blue), rust (yellow), crack with precipitation (pink), as well as algae (green).

Figure 4. IBM Inspecto high-resolution analysis of the Storebælt pillar 3, south face, bottom region. A few selected zoomed areas to show the precision of masks generated by our AI FM.

List of all detected findings

| Defect ID | Defect type | Main measure | Condition Rating | Reviewed | Review comment | Element function fullfilled | Consequence for elements | Drone position and altitude | Link | File name |
|---|---|---|---|---|---|---|---|---|---|---|
| Pillar 03 - South face | | | | | | | | | | |
| 303 | Crack | 98.65 cm | 3 | Yes | Long vertical crack, to be repaired before winter. | Yes | No | (55.348754°, 11.09026°) 44.949m | Link | DJI_20210604103310_0006_SUPR |
| 148 | Crack | 92.41 cm | 3 | Yes | Long vertical crack, to be repaired before winter. | Yes | No | (55.348746°, 11.090254°) 43.76m | Link | DJI_20210604103501_0007_SUPR |
| 257 | Spalling | 0.053 m2 | 3 | Yes | Extended spalling, to be repaired before winter. | Yes | No | (55.348746°, 11.090254°) 43.76m | Link | DJI_20210604103501_0007_SUPR |
| 247 | Spalling | 0.0339 m2 | 3 | Yes | Extended spalling, to be repaired before winter. | Yes | No | (55.348746°, 11.090254°) 43.76m | Link | DJI_20210604103501_0007_SUPR |
| 446 | Spalling | 0.0334 m2 | 3 | Yes | Extended spalling, to be repaired before winter. | Yes | No | (55.348754°, 11.09026°) 44.949m | Link | DJI_20210604103310_0006_SUPR |
| 448 | Spalling | 0.0233 m2 | 3 | Yes | Extended spalling, to be repaired before winter. | Yes | No | (55.348754°, 11.09026°) 44.949m | Link | DJI_20210604103310_0006_SUPR |
| 246 | Spalling | 0.0224 m2 | 3 | Yes | Extended spalling, to be repaired before winter. | Yes | No | (55.348746°, 11.090254°) 43.76m | Link | DJI_20210604103501_0007_SUPR |
| 96036 | Spalling | N/A | 3 | Yes | Extended spalling, to be repaired before winter. | Yes | No | (55.348754°, 11.09026°) 44.949m | Link | DJI_20210604103422_0034_SUPR.JPG |
| 96037 | Spalling | N/A | 3 | Yes | Extended spalling, to be repaired before winter. | Yes | No | (55.348754°, 11.09026°) 44.949m | Link | DJI_20210604103422_0034_SUPR.JPG |
| 96198 | Crack | N/A | 3 | Yes | Long vertical crack, to be repaired before winter. | Yes | No | (55.348754°, 11.09026°) 44.949m | Link | DJI_20210604103405_0026_SUPR.JPG |
| 96398 | Crack | N/A | 3 | Yes | Long vertical crack, to be repaired before winter. | Yes | No | (55.348754°, 11.09026°) 44.949m | Link | DJI_20210604103403_0025_SUPR.JPG |
| 96445 | Crack | N/A | 3 | Yes | Long vertical crack, to be repaired before winter. | Yes | No | (55.348754°, 11.09026°) 44.949m | Link | DJI_20210604103425_0035_SUPR.JPG |
| 96470 | Crack | N/A | 3 | Yes | Long vertical crack, to be repaired before winter. | Yes | No | (55.348754°, 11.09026°) 44.949m | Link | DJI_20210604103427_0036_SUPR.JPG |

Figure 5. IBM Inspecto summary table to be included in the asset inspection report. Note that the indicated severity scores and comments scores are randomly generated for demonstration purposes and do not represent the actual condition of the bridge.

Once the civil engineer has terminated reviewing the bridge surface, a customizable report can be generated including all necessary information. IBM Inspecto reports are available both as HTML as well as PDF. In Figure 5, an example of a summary table with all the detected defects on pillar 3 south face, is provided. The table shows on the top the most severe defects and includes links to view them in the service.

## 3.2 *Other examples of applications*

We successfully applied IBM Inspecto to a variety of other applications, from older bridges like the Stenungsund in Sweden (see Figure 6) to the Dübendorf Airbase Runway (see Figures 7 and 8). These two applications show how our AI FMs can adapt easily to predict defects on very different types of concrete surfaces. The Stenungsund bridge is very representative of most bridges in Europe and the US, which were built more than 50 years ago and have major signs of deterioration, such as areas with spalling with corroded rebars. We could reach very precise detection accuracy by fine-tuning the model with roughly 500 images.

Figure 6. IBM Inspecto results on the Stenungsund bridge: the image shows large instances of spalling with corroded rebars that our model detected correctly.

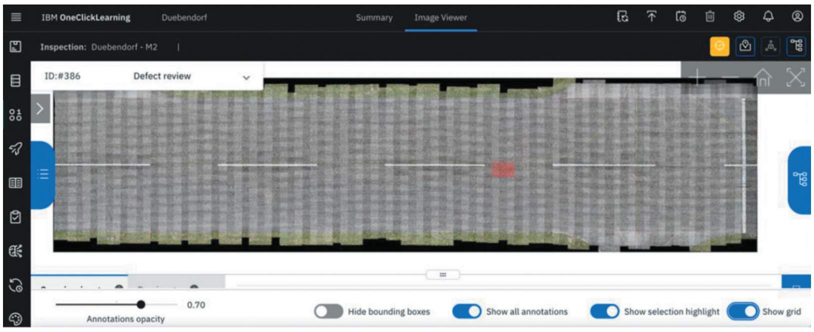


Figure 7. IBM Inspecto results on the Dubendorf Airbase: we inspected 250m of the runway with more than 800 high-resolution images automatically stitched together.

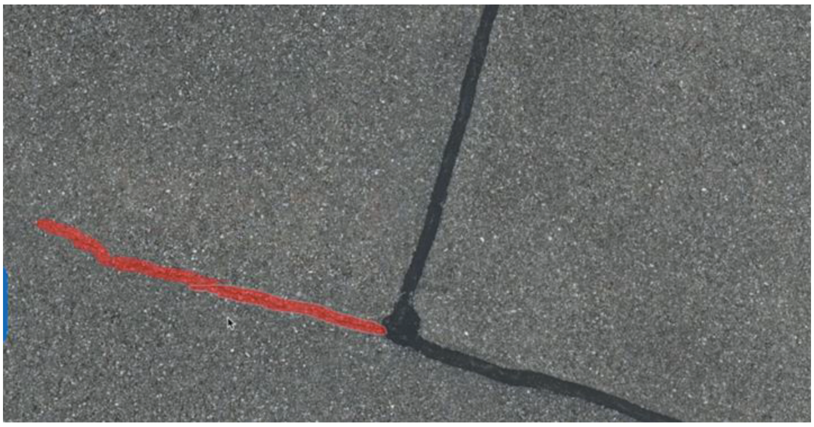

Figure 8. A detail of a crack that was previously repaired on the Dubendorf Airbase and that continued to grow later.

The Dübendorf Airbase runway surface is also quite different than concrete bridge walls. Nevertheless, our algorithm has been able to completely reconstruct the global overview stitched image and analyze it detecting even the smallest crack without any fine-tuning. A nice example in Figure 8 shows a crack that continued to grow after repair, with our model being able to detect only the new branch precisely next to the repaired patch.

## 4 CONCLUSIONS AND FUTURE WORKS

In this paper, we presented IBM Inspecto, a research-industry platform powered by IBM Research foundation models. We apply IBM Inspecto to the Storebælt Bridge in collaboration with Sund & Bælt, as well as on other applications. We demonstrate key features such as the intuitive navigation, the accurate detection of defects in high-resolution images, and the automated generation of customizable reports. IBM Inspecto aims to reduce the time and cost of visual inspection of complex assets, while digitalizing completely the entire inspection process and can be applied to a wide variety of infrastructures, from bridges to buildings, roads, and runways.

We are currently extending IBM Inspecto with additional AI features that will further accelerate the inspection process. First, we are developing a defect review assistant, powered by multi-modal FMs. Our review assistant will learn from a small, selected number of domain experts who manually reviewed defects, to then perform an automated review of all the other thousands of defects detected by our AI FMs. We envision that this process will accelerate the review of the detected defects by at least one if not two orders of magnitude. To allow users to extend our pre-trained models and detect other types of defects that are present on their infrastructures, we are developing a Visual Prompting Lab service. There, engineers will just need to provide a few examples of images and swiftly scribble on top of the new type of defects to teach the algorithm what they would like to be detected. The process will be naturally intuitive and require minimal data and human labor. Finally, we will enable the users to compare inspected surfaces over time. Leveraging our stitched images, IBM Inspecto will conveniently show what has changed over each surface from the previous inspection, without asking to the user to re-annotate and re-assess previously assessed defects, thus further reducing the work to be done in every future inspection.

## ACKNOWLEDGMENTS

This work was supported by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number **22.00295**.

## REFERENCES

Assran M., Caron M. et al. Masked Siamese Networks for Label-Efficient Learning, arXiv 2022.

Bianchi S, Biondini F. Bridge Condition Assessment Using Supervised Decision Trees. Proceedings of the 1st Conference of the European Association on Quality Control of Bridges and Structures: EUROSTRUCT 2021, Springer, 2022; 1108–1116.

Chen T., Kornblith S., Norouzi M., Hinton G. A Simple Framework for Contrastive Learning of Visual Representations, arXiv 2020.

Dosovitskiy A., Beyer L. et al.An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, arXiv 2019.

LeCun Y, Bengio Y, Hinton G. Deep Learning. Nature 2015 May 521(7553):436–444.

Grill J., Strub F., et al.. Bootstrap your own latent: A new approach to self-supervised Learning arXiv 2020.

He K., Chen X., et al. Masked Autoencoders Are Scalable Vision Learners, arXiv. 2021.

He K. and Fan H., Wu Y., Xie S., Girshick R. Momentum Contrast for Unsupervised Visual Representation Learning, arXiv 2020.

Manzini N, Mar N, Schmidt F, Bercher JF, Orcesi A, Marchand P, et al. An Automated Machine Learning-Based Approach for Structural Novelty Detection Based on SHM. In: Proceedings of the 1st Conference of the European Association on Quality Control of Bridges and Structures: EUROSTRUCT 2021 Springer, 2022;1180–1189.

Misra I., van der Maaten L., Self-Supervised Learning of Pretext-Invariant Representations, arXiv 2019.

Salehi H, Burgueño R. Emerging Artificial Intelligence Methods in Structural Engineering. Engineering Structures, 2018; 171:170–189.

Zhang H., Li F., et al. DINO: DETR with Improved DeNoising Anchor Boxes for End-to-End Object Detection, arXiv 2022.