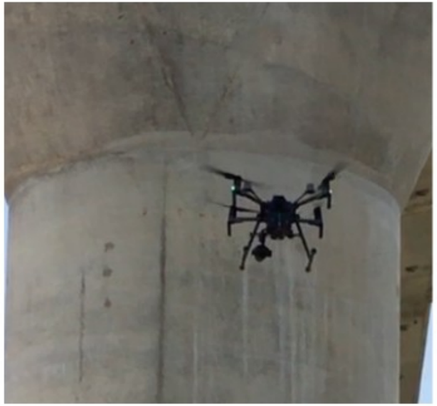
Towards Automated Visual Inspection of Civil Engineering Structures



Mattia Rigotti
IBM Research Zurich
mrg@zurich.ibm.com

IEEE MEDITERRANEAN ELECTROTECHNICAL CONFERENCE
Palermo - Italy / June 14-16, 2022

Notices and disclaimers

-
- **U.S. Government Users Restricted Rights — use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM.**
- This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.
- Information in these presentations (including information relating to products that have not yet been announced by IBM) has been reviewed for accuracy as of the date of initial publication and could include unintentional technical or typographical errors. IBM shall have no responsibility to update this information. **This document is distributed "as is" without any warranty, either express or implied. In no event, shall IBM be liable for any damage arising from the use of this information, including but not limited to, loss of data, business interruption, loss of profit or loss of opportunity.** IBM products and services are warranted per the terms and conditions of the agreements under which they are provided. The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions.
- IBM products are manufactured from new parts or new and used parts.
  In some cases, a product may not be new and may have been previously installed. Regardless, our warranty terms apply."
- **Any statements regarding IBM's future direction, intent or product plans are subject to change or withdrawal without notice.**
- Performance data contained herein was generally obtained in a controlled, isolated environments. Customer examples are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual performance, cost, savings or other results in other operating environments may vary.
- References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business.
- Workshops, sessions and associated materials may have been prepared by independent session speakers, and do not necessarily reflect the views of IBM. All materials and discussions are provided for informational purposes only, and are neither intended to, nor shall constitute legal or other guidance or advice to any individual participant or their specific situation.

Notices and disclaimers continued

- It is the customer's responsibility to ensure its own compliance with legal requirements and to obtain advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulatory requirements that may affect the customer's business and any actions the customer may need to take to comply with such laws. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the customer follows any law.

- Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products about this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products. IBM does not warrant the quality of any third-party products, or the ability of any such third-party products to interoperate with IBM's products. **IBM expressly disclaims all warranties, expressed or implied, including but not limited to, the implied warranties of merchantability and fitness for a purpose.**

- The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents, copyrights, trademarks or other intellectual property right.

Team

IBM Research


Florian Scheidegger

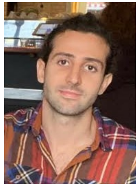

Thomas Frick


Filip Janicki

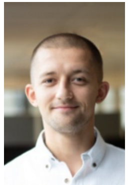

Ioana Giurgiu


Mattia Rigotti

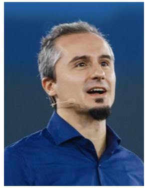

Andrea Bartezzaghi


Roy Assaf


Panagiotis Chatzidoukas


Cezary Skura


Chiara Marchiori

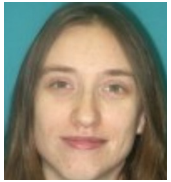

Cristiano Malossi


Samuel Kallner


Dmitri Pikus


Gil Sharon


Maya Barnea


Ira Rosen


Benny Mandler


Natti Eder


Nili Guy

Civil Engineering Partners

Sund≈Bælt

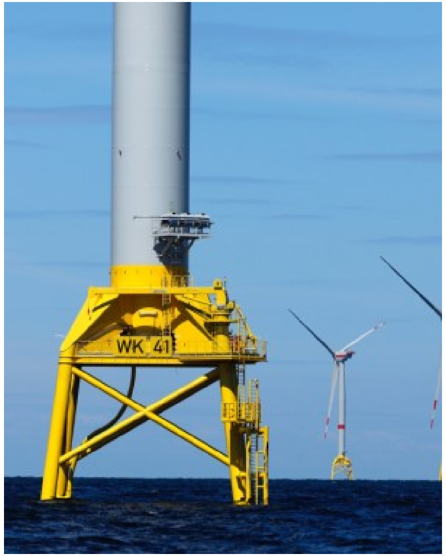
SACERTIS

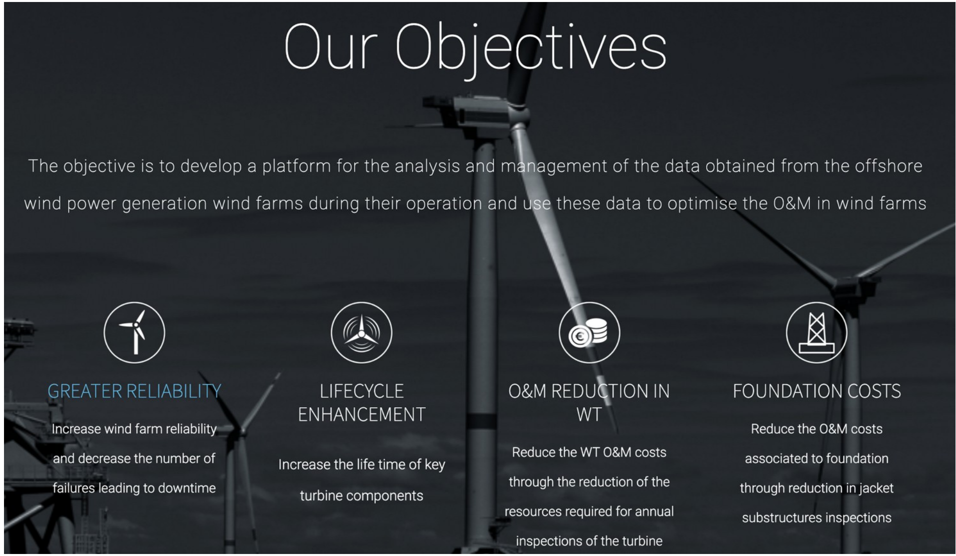
Innovation in EU key projects with civil engineering partners

ROMEO

IBM Research develop ML models to detect and predict Wind Turbine component failures based on a rich set of historic data provided by partners and design a reference Cloud implementation of the end-to-end data flow

www.romeoproject.eu

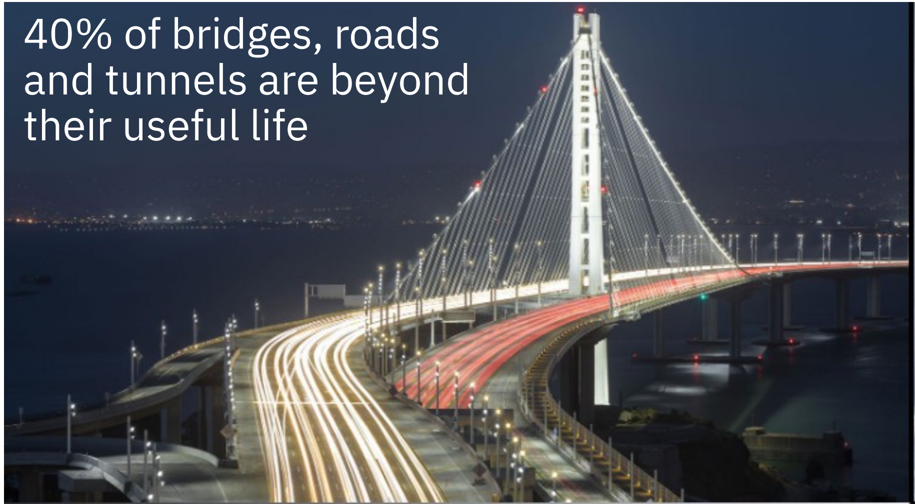




IM-SAFE.EU

IBM Research defines the digitalization avenue for future of civil infrastructures monitoring and inspection, leveraging AI, cloud, and automation

https://im-safe-project.eu

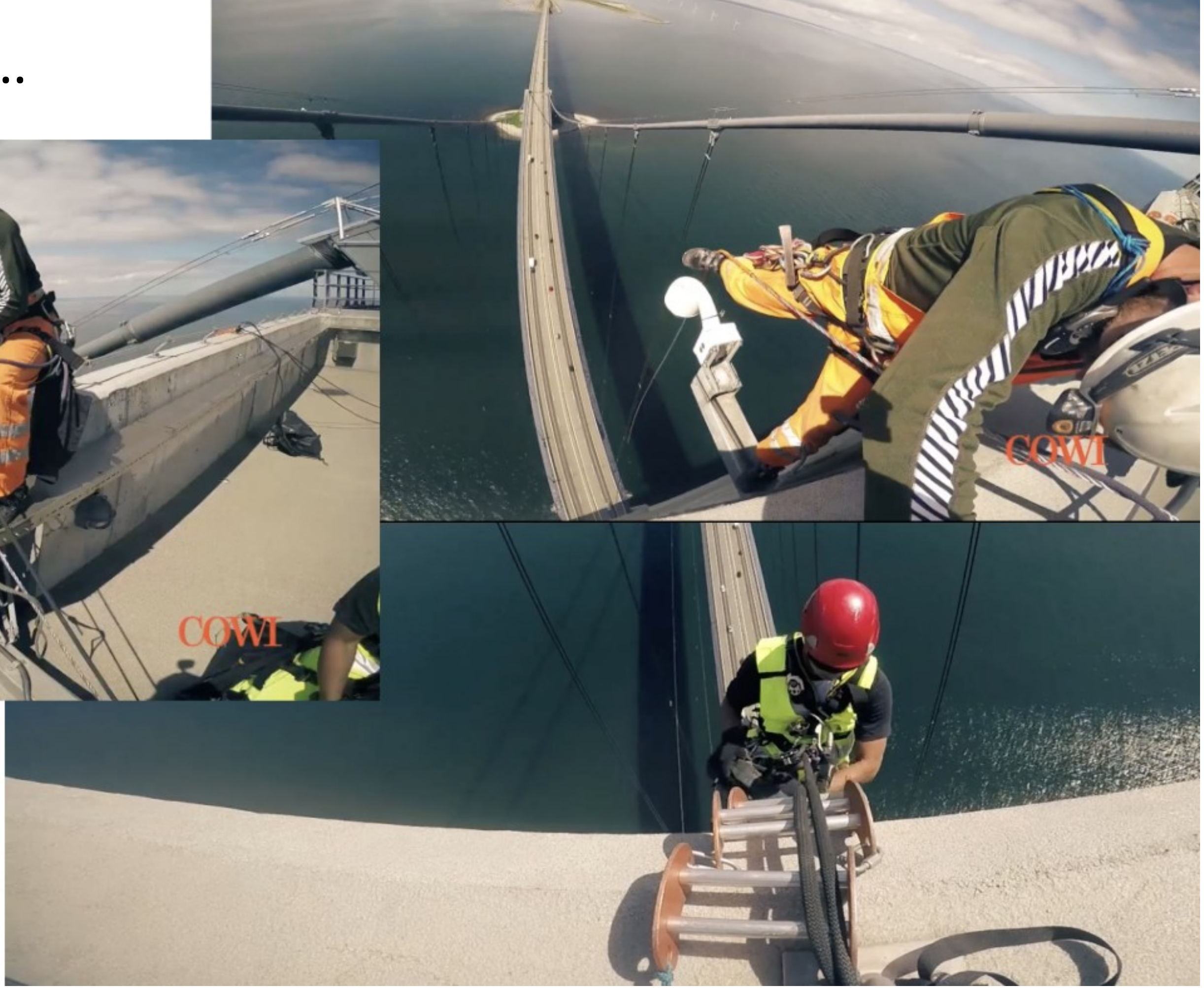

40% of bridges, roads and tunnels are beyond their useful life


$50 Billion and 2 Billion man-hours spent on manual inspection of bridges


Bridges and Tunnels are complex structures that move and change constantly


One failure can shutdown critical arteries

Inspections in the past.....



Inspections today

Drone/Robots visual inspection pipeline for civil infrastructures

Visual inspection pipeline

High-quality data acquisition

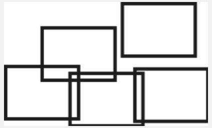
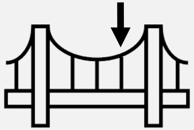


Photo stiching

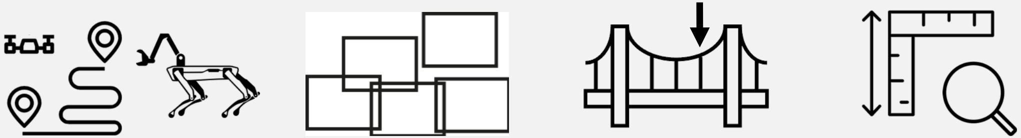


High-resolution detection of defects

Defect measuring

Drone/Robots visual inspection pipeline for civil infrastructures

Visual inspection pipeline



High-quality data acquisition

Photo stiching

High-resolution detection of defects

Defect measuring

Automated acquisition of thousands high-resolution images with repeatable drones and robots missions

Drone/Robots visual inspection pipeline for civil infrastructures

Visual inspection pipeline

High-quality data acquisition

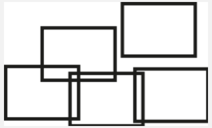
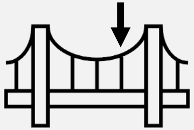


Photo stiching

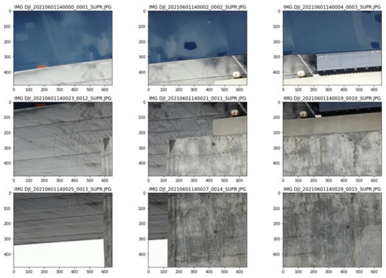


High-resolution detection of defects

Defect measuring

Automated acquisition of thousands high-resolution images with repeatable drones and robots missions

Automated 2D rendering of large elements (pillars, towers, etc.) and defects disambiguation and localization in global coordinate system



Drone/Robots visual inspection pipeline for civil infrastructures

Visual inspection pipeline

High-quality data acquisition

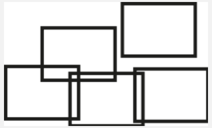
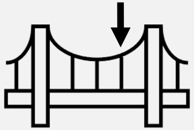


Photo stiching

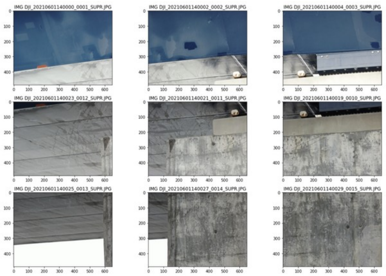


High-resolution detection of defects

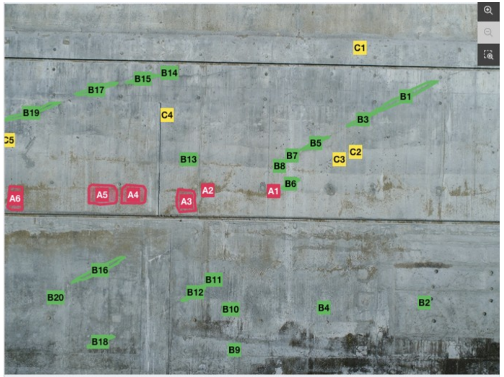
Defect measuring

Automated acquisition of thousands high-resolution images with repeatable drones and robots missions

Automated 2D rendering of large elements (pillars, towers, etc.) and defects disambiguation and localization in global coordinate system



Deep learning based computer vision methods to detect tiny and rare defects on each high-res image



Drone/Robots visual inspection pipeline for civil infrastructures

Visual inspection pipeline

High-quality data acquisition

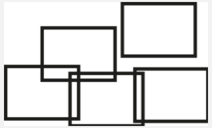
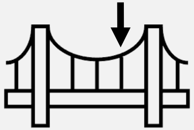


Photo stiching

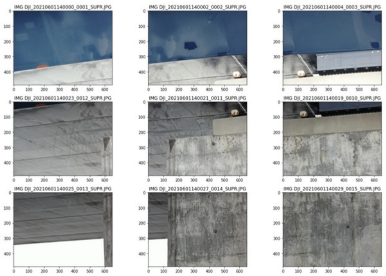


High-resolution detection of defects

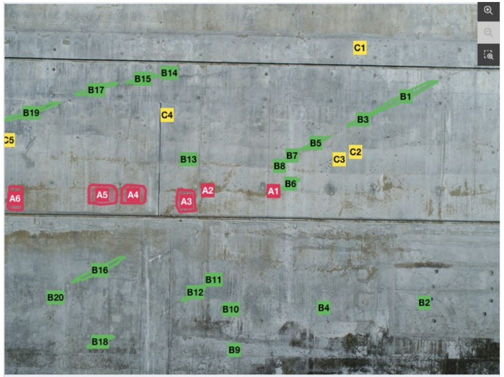
Defect measuring

Automated acquisition of thousands high-resolution images with repeatable drones and robots missions

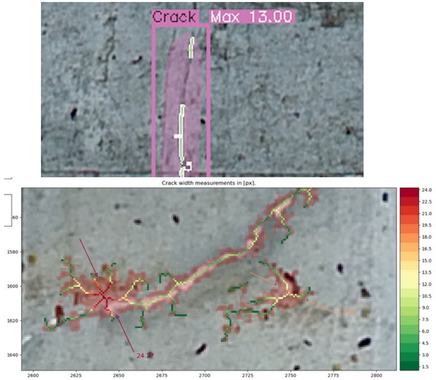
Automated 2D rendering of large elements (pillars, towers, etc.) and defects disambiguation and localization in global coordinate system

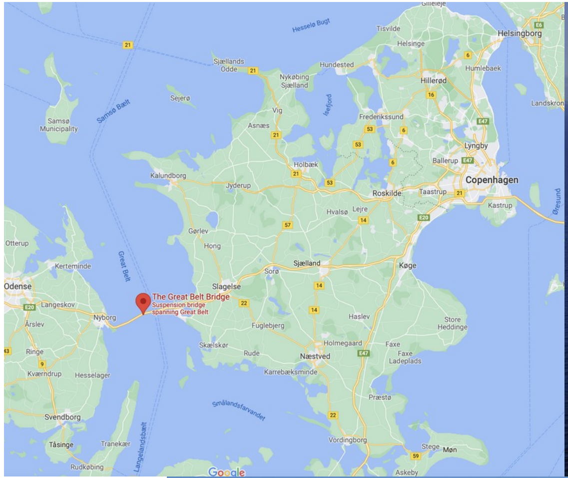


Deep learning based computer vision methods to detect tiny and rare defects on each high-res image



Future: explainable suggestions for maintenance decisions





The Storebaelt Link

Total length: 17.5 km

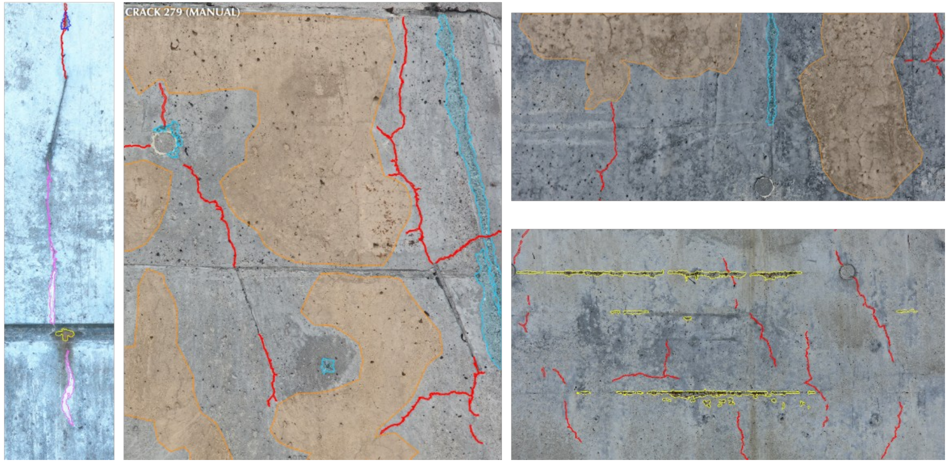
West Bridge
1 rail and
1 road bridge
Total length: 6’611m
Construction period: 1989-1994

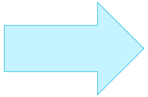
East Bridge
Total length: 6’790m
Construction period: 1991-1998
Suspension bridge: 2’700 m
Main span 1’624 m
Two side spans of 535 m
Navigation clearance is 65 m
Height of pylons: 254 m

June 2021 Inspection

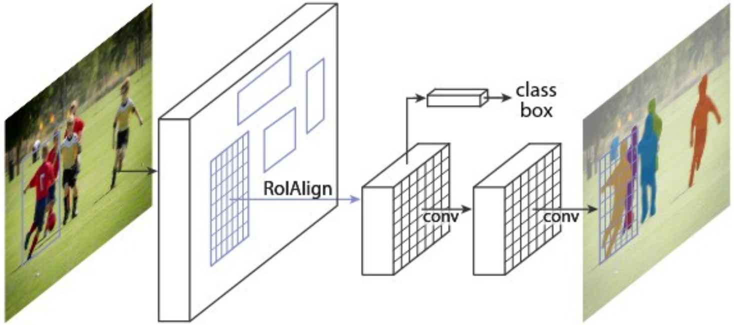
- Latest drone technology
  - Matrix DJI 300 RTK
  - GSD calibration
- 22 pillars inspected
- Images generated
  - High quality/high resolution
  - > 23k
  - ~ 200 GB
- Each image > 6k x 4k pixels (10 MB)

Manual annotations and model inference





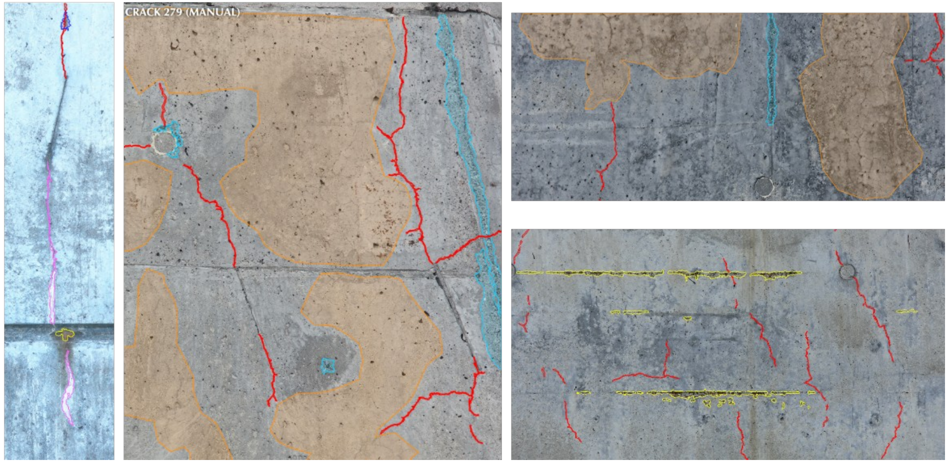
Mask R-CNN [He et al. 2017]



- Defect annotation guidelines
- Expert manual annotations
- Test set
- Train set

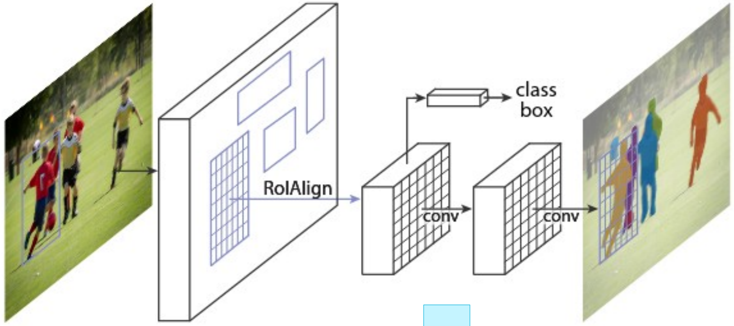
- Training of Mask R-CNN
- Detectron2 implementation
- Customized with pyramid and tiling

Manual annotations and model inference

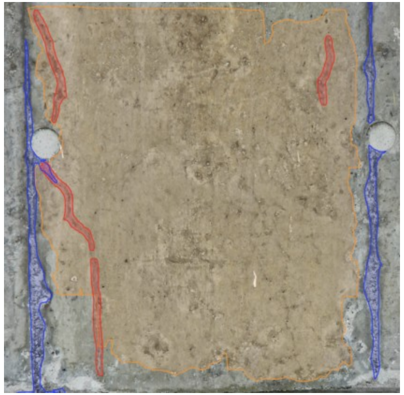


- Defect annotation guidelines
- Expert manual annotations
- Test set
- Train set

Mask R-CNN [He et al. 2017]

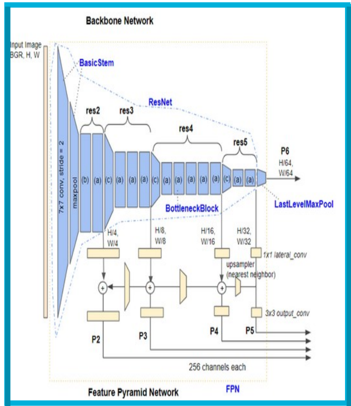


- Output: instance segmentation masks
- 6 types of defects



One Click Learning – IBM Research Platform

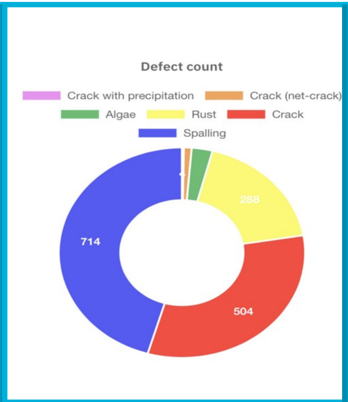
A scalable, cloud based platform to:



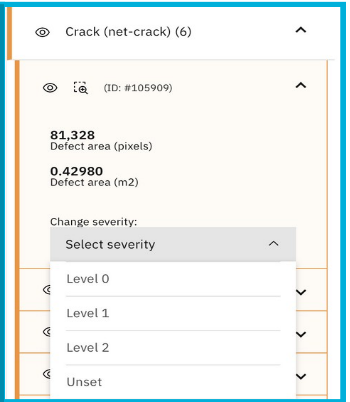
Empower Engineers and Infrastructure Managers to use pretrained AI models for everyday inspections

Detect and caracterize damage size and measure other important attributes

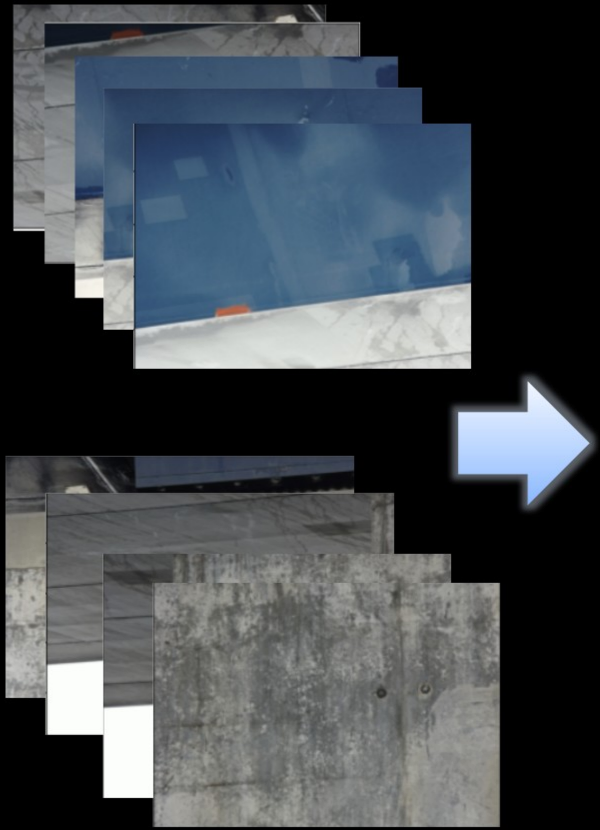
Reconstruct infrastructure elements and locate damages in scene with automated image stitching

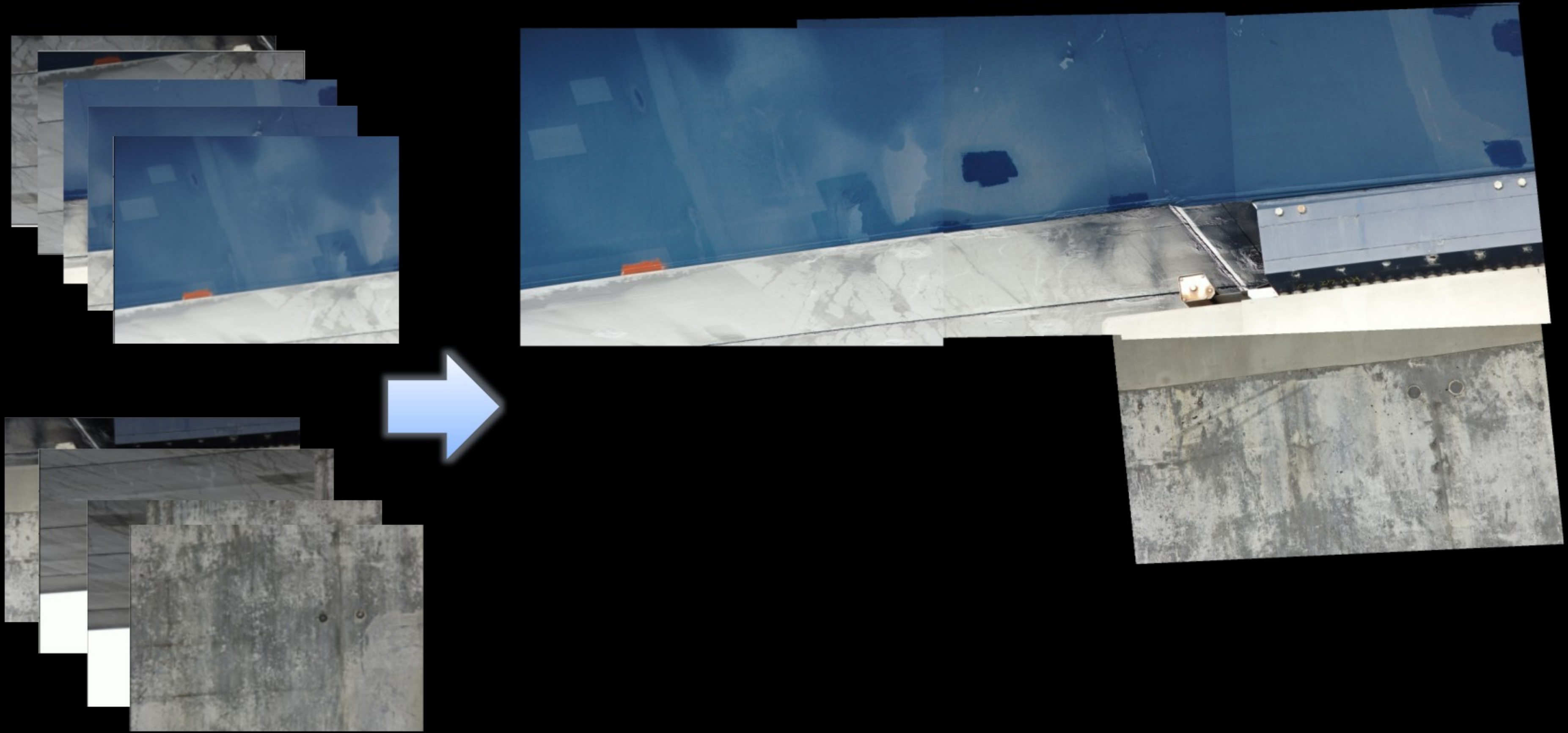


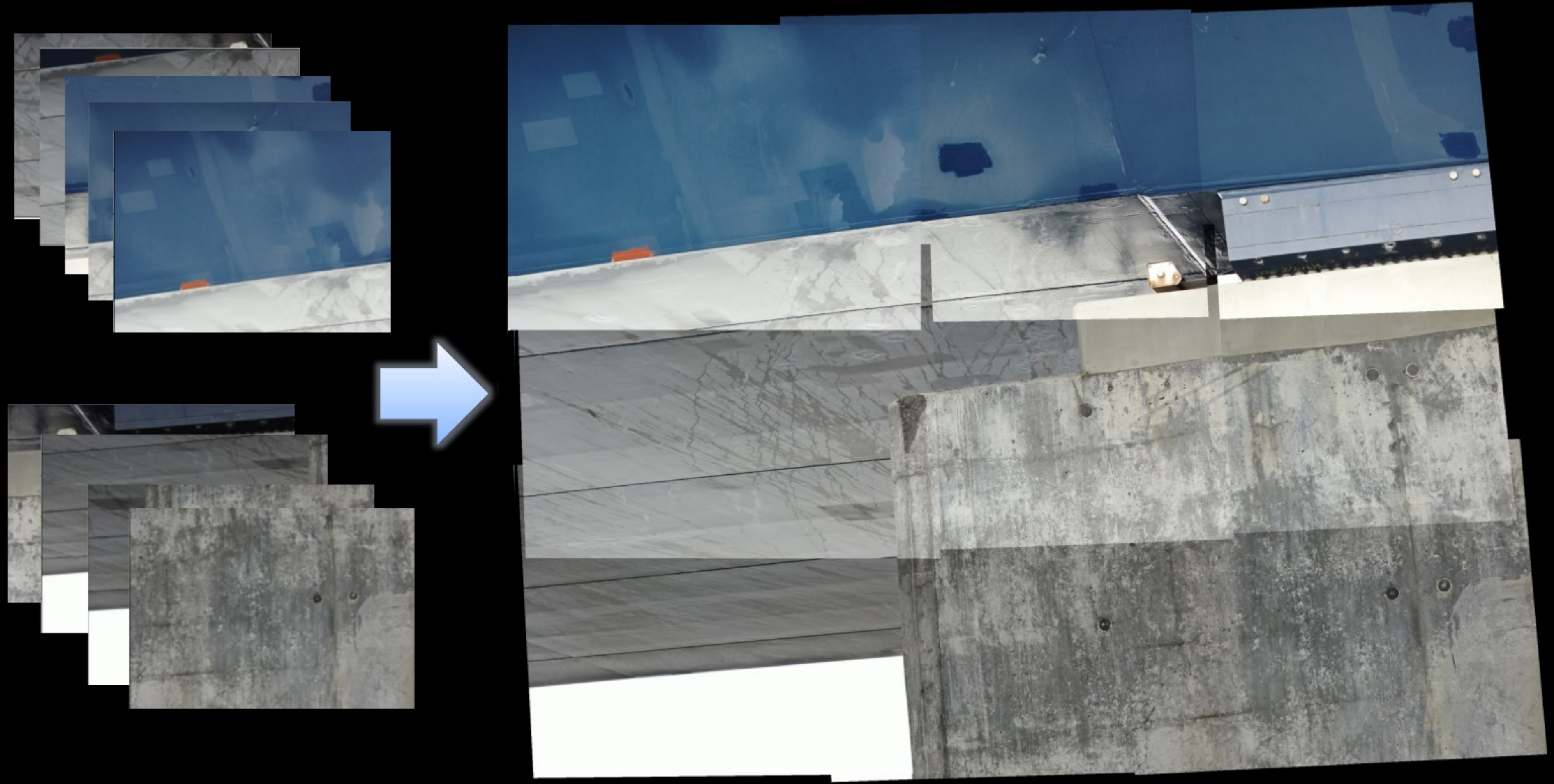
Powerful data exploration of large datasets and large images with corresponding detected damage



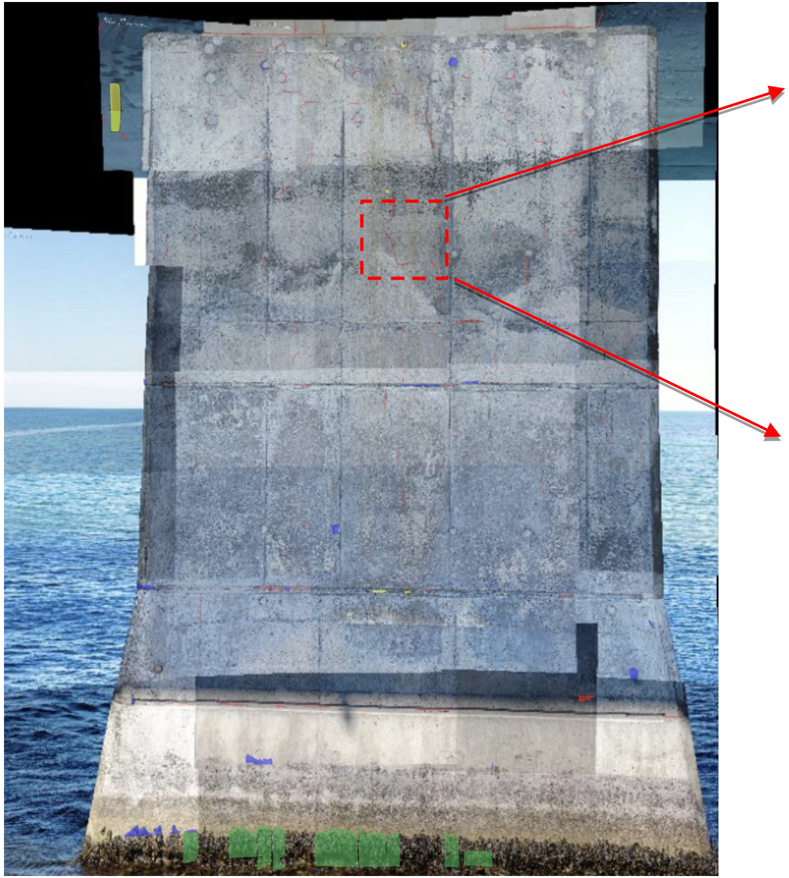
Extraction and aggregation of analysis results



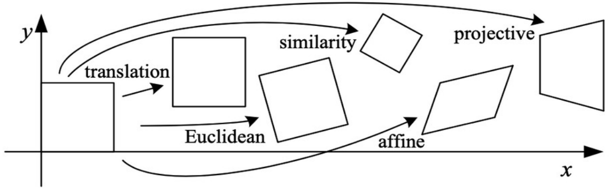


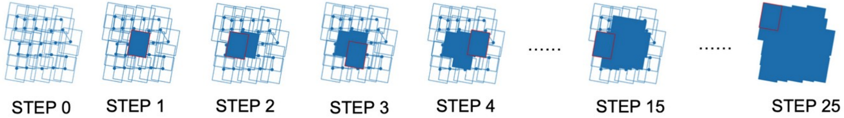


Automated stitching algorithm

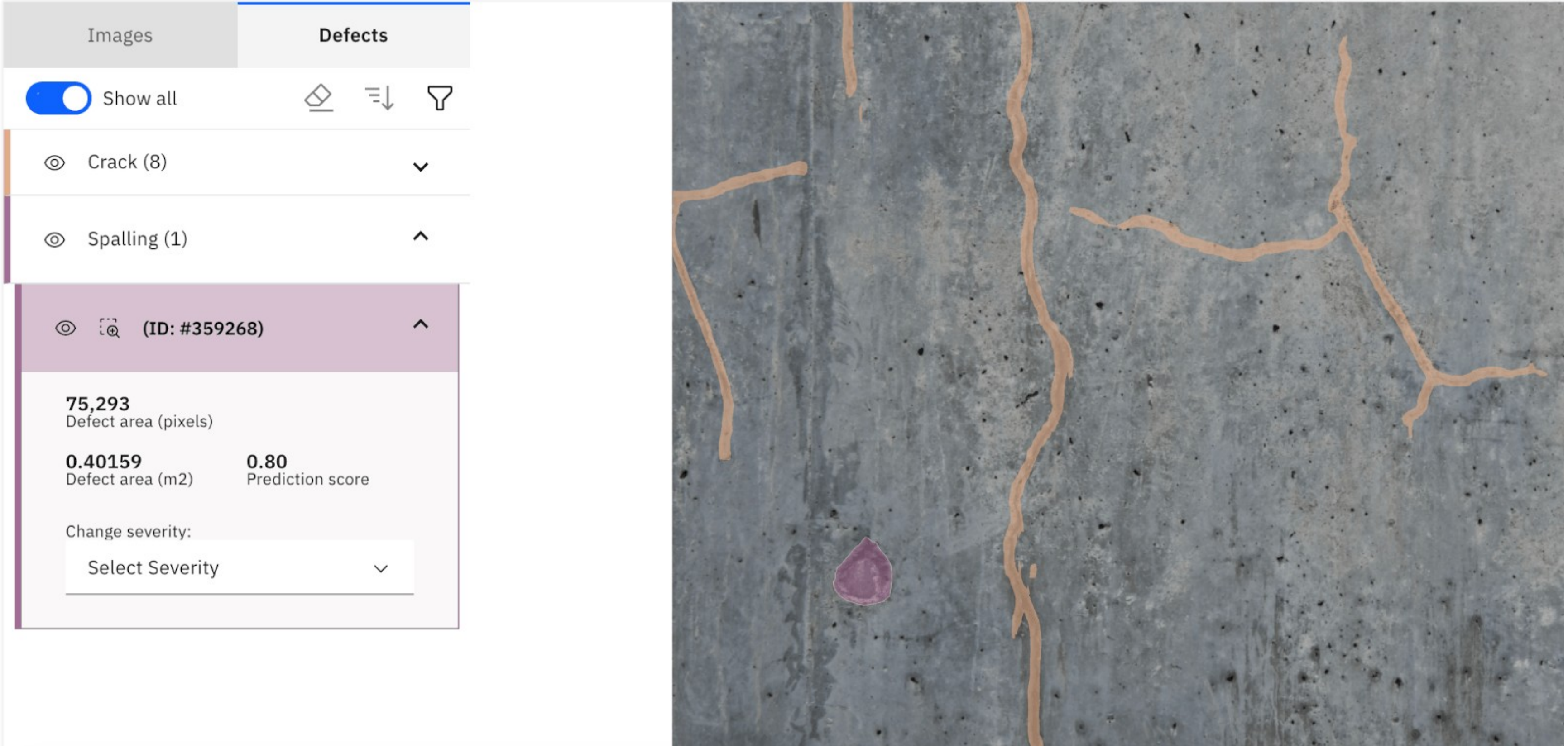


COS enables smooth navigation of large images by streaming multi-resolution pyramid of tiles on demand

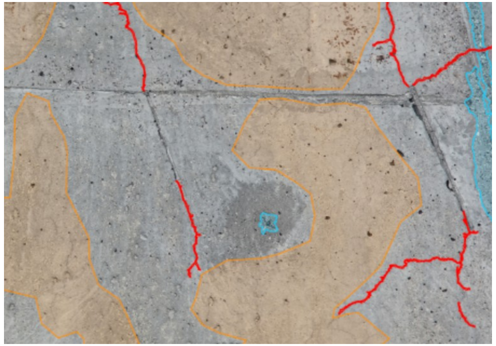




Defect segmentation and characterization



AI aiding in maintenance decisions



Defects detection

Structural engineer

Maintenance decision

Maintenance decisions are taken by engineers based on *sequential* virtual inspections of structure

AI *taking* maintenance decisions

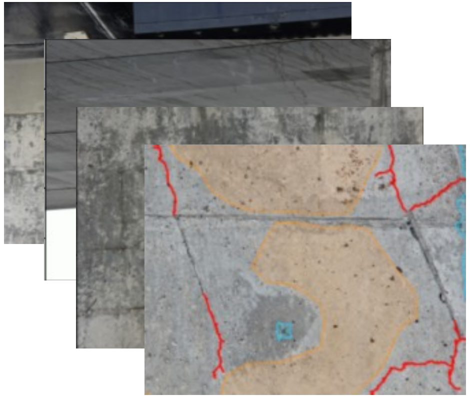
Defects detection

Structural engineer

Maintenance decision

Maintenance decisions are taken by engineers based on *sequential* virtual inspections of structure

How about automating also inspection and maintenance decision-making?



Machine Learning Model

Automated maintenance (used as suggestions)

Structural engineer

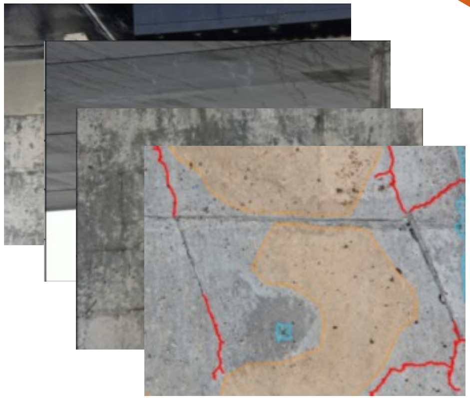
Takes final decision based on suggestion

Trusted AI decisions

ML prediction and decisions need to be **explainable** so that domain experts know whether they are **trustworthy**

Maintenance decisions are taken by engineers based on *sequential* virtual inspections of structure

How about automating also inspection and maintenance decision-making?



Machine Learning Model

Automated maintenance (used as suggestions)

Structural engineer

Takes final decision based on suggestion

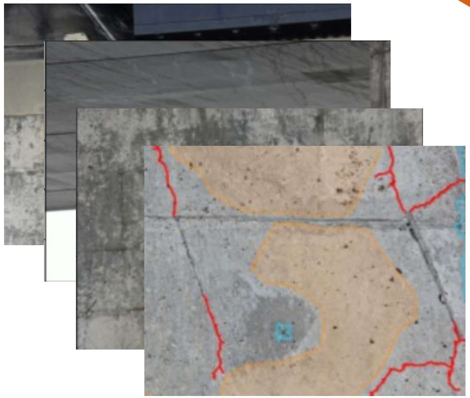
Trusted AI decisions

Maintenance decisions are taken by engineers based on *sequential* virtual inspections of structure

How about automating also inspection and maintenance decision-making?

ML prediction and decisions need to be **explainable** so that domain experts know whether they are **trustworthy**

Explainable AI (XAI)



Machine Learning Model

Automated maintenance (used as suggestions)

Structural engineer

Takes final decision based on suggestion

Trusted AI decisions

IBM www.ibm.biz/trusted-ai

Overview

← Teams

Trusted AI

Our trust in technology relies on understanding how it works. We need to understand why AI makes the decisions it does. We're developing tools to make AI more explainable, fair, robust, private, and transparent.

Explore our teams ↓

AI Testing →	Adversarial Robustness and Privacy →
Explainable AI →	Fairness, Accountability, Transparency →
Trustworthy Generation →	Uncertainty Quantification →

Trusted AI decisions

IBM www.ibm.biz/trusted-ai

Overview

← Teams

Trusted AI

Our trust in technology relies on understanding how it works. We need to understand why AI makes the decisions it does. We're developing tools to make AI more explainable, fair, robust, private, and transparent.

Explore our teams ↓

AI Testing →	Adversarial Robustness and Privacy →
Explainable AI We're creating tools to help AI systems explain why they made the decisions they did. →	Fairness, Accountability, Transparency →
Trustworthy Generation →	Uncertainty Quantification →

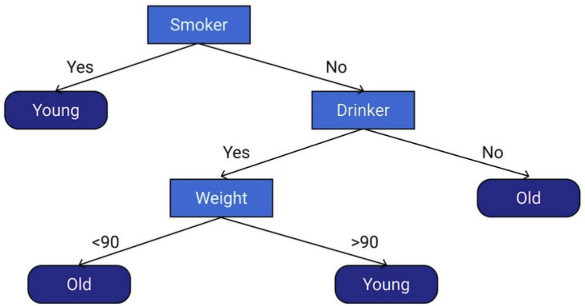
AI Explainability: Challenges of model interpretability

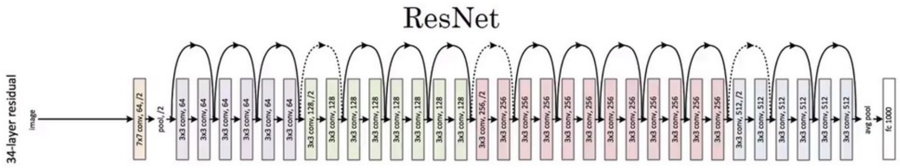
Interpretable “white box models”

- e.g. decision trees, linear regression
- *Inherently interpretable*: decisions based on human-understandable categories (concepts) grounded in domain expertise
- **BUT** limited expressivity and accuracy

“Black box models”

- e.g. Deep neural networks
- Very expressive and accurate
- **BUT** opaque and uninterpretable

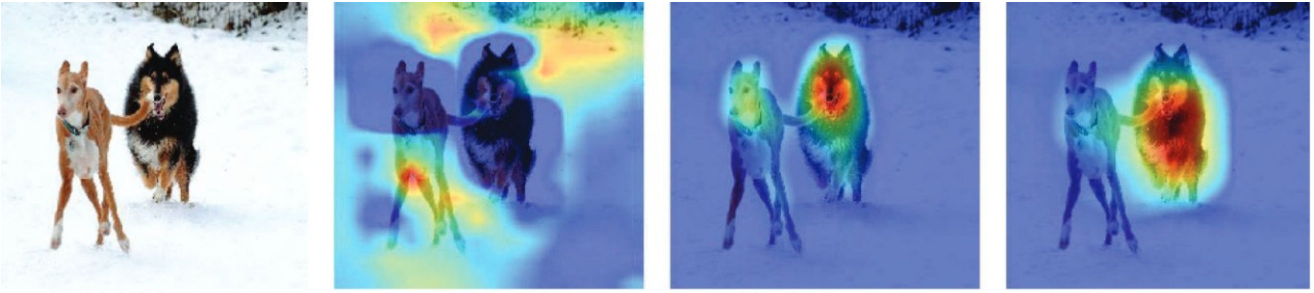




Failures of Black Box Explainability

Post-hoc explainability (e.g. Grad-CAM "heatmaps")

- Operate on ***low-level features*** (e.g. pixel values), combined unintelligibly
  → not human understandable ***concepts***
- Explanations not guaranteed to be ***faithful*** explanations
  (i.e. decisions not causally linked to explanations)



(a) Original Image (b) Grad-CAM (c) CNN-fixation (d) Eigen-CAM

Failures of Black Box Explainability

Post-hoc explainability (e.g. Grad-CAM “heatmaps”)

- Operate on ***low-level features*** (e.g. pixel values), combined unintelligibly
  → not human understandable ***concepts***
- Explanations not guaranteed to be ***faithful*** explanations
  (i.e. decisions not causally linked to explanations)

Attention mechanism

- Attention weights offer intuitive “explanations” of predictions
- However, “Attention is not explanation”, i.e. it is not a “fail-safe indicator” [Serrano & Smith, 2019], especially in neural networks

SST

Original: reggio falls victim to relying on the very digital technology that he fervently scorns creating a meandering inarticulate and ultimately disappointing film

Adversarial: reggio falls victim to relying on the very digital technology that he fervently scorns creating a meandering inarticulate and ultimately disappointing film $\Delta\hat{y}$: *0.005*

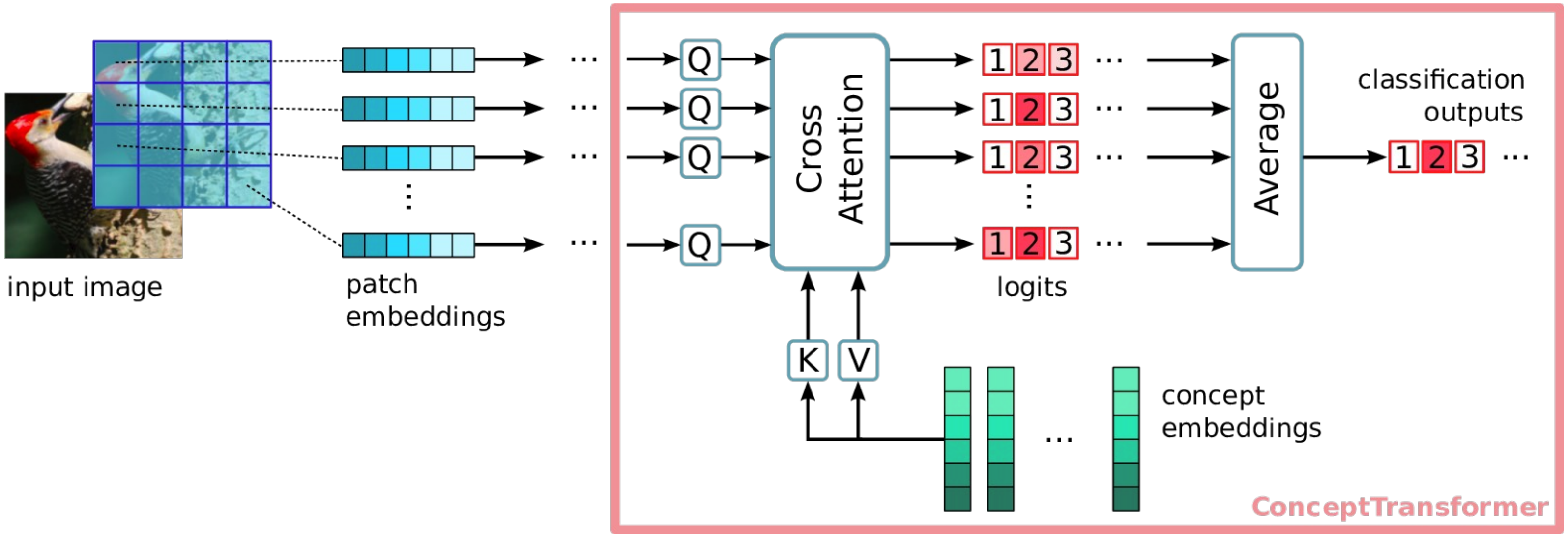
Sentiment recognition model on adversarial example:
attention is allocated to a token ("that")
→ not a good explanation for the sentence being classified as negative [Jain & Wallace, 2019]

AI Explainability

Desiderata

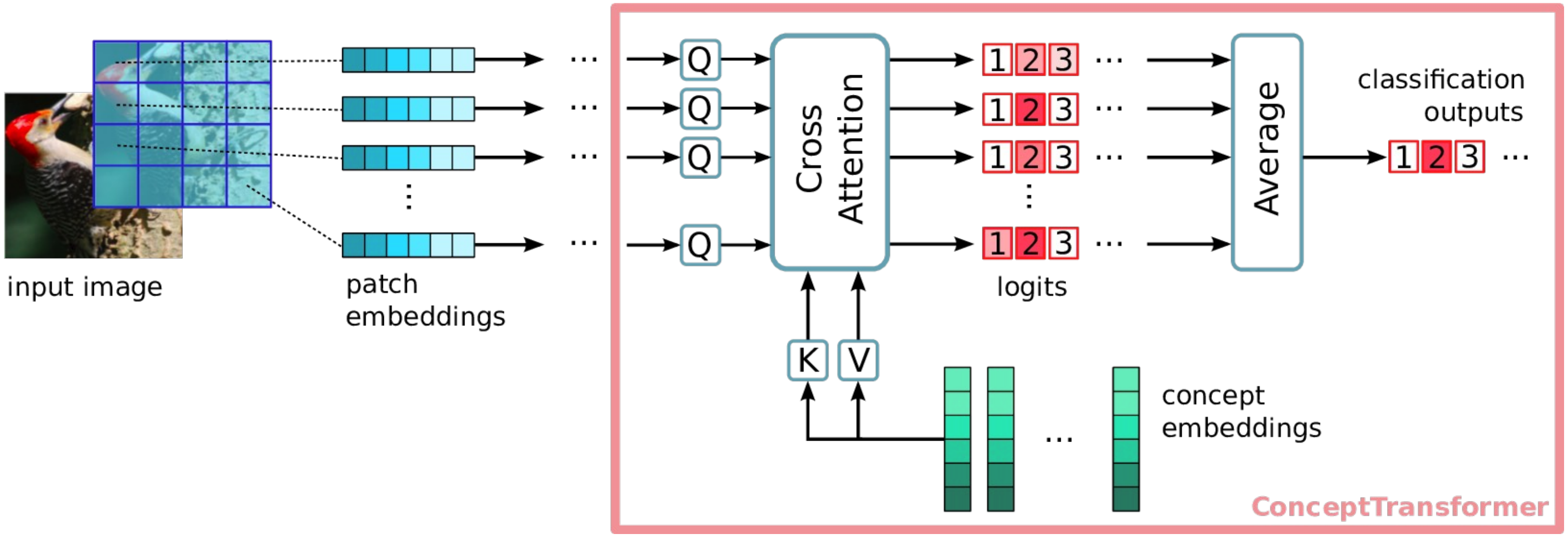
- ***Flexibility***: explanations for arbitrary deep learning model
- ***Plausibility***: explanations grounded in domain at hand and convincing to domain experts
- ***Faithfulness***: explanation should reflect the decision of the model, guaranteed to actually explain the model's operation

Attention-based interpretability with **ConceptTransformers**



- Transformer-based *drop-in* replacement module as classier head
- Classifier output depends on pre-determined *domain-relevant concepts*

Attention-based interpretability with **ConceptTransformers**

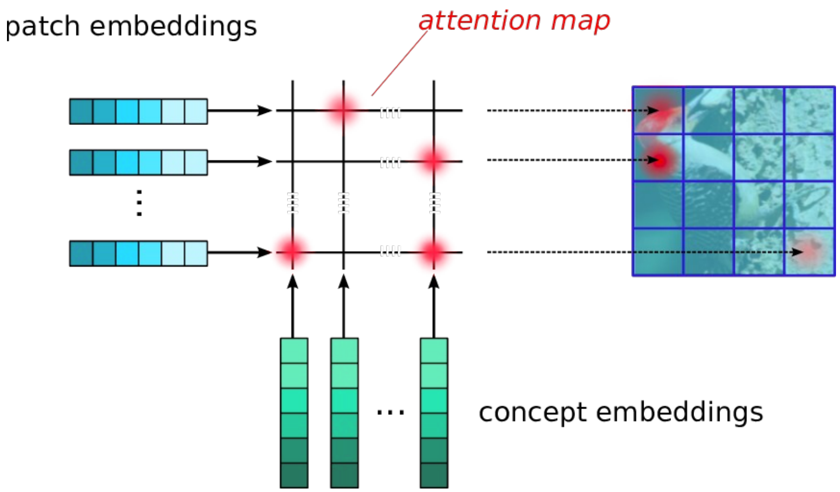


1) **Vision backbone + tokenizer**: standard vision backbone like ResNet50 feature extractor, followed by a "tokenizer" into patches, or ViT
2) **Cross-Attention between patches and concepts**: cross-attention layer computes attention weights from image patches onto concepts

Attention-based interpretability with **ConceptTransformers**

2) **Cross-Attention between patches and concepts**: cross-attention layer computes attention weights from image patches onto concepts

Cross-attention

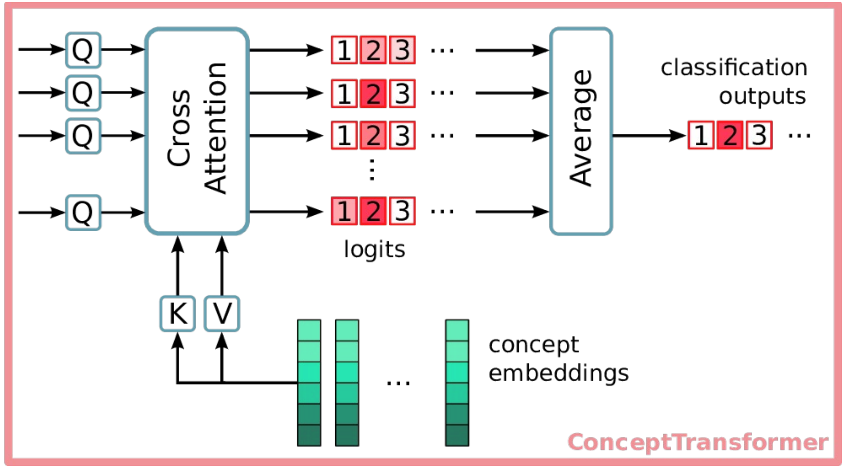


- Concepts are determined a priori based on domain knowledge
- Patch trained to **distribute attention weights on available concepts** consistently with domain knowledge
- Final decision based on the attention contributions of available concepts

→ **explanations are PLAUSIBLE by construction**

ConceptTransformers: interpretability by design

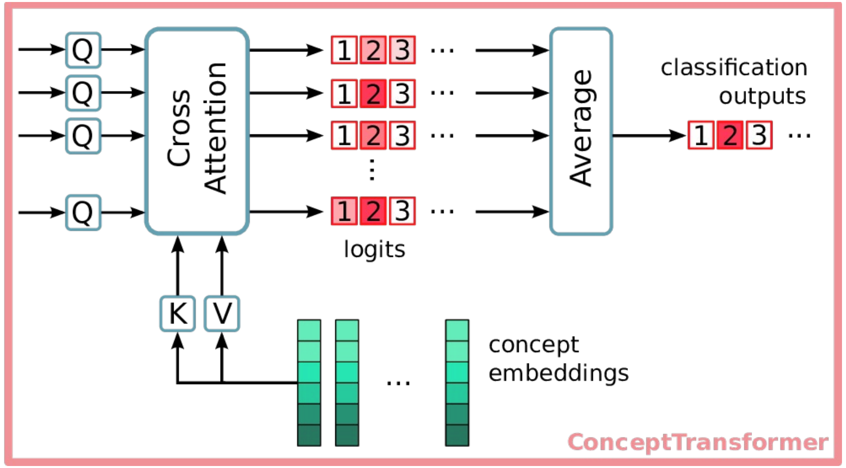
- The attention weights computed by ConceptTransformer are concept-based explanations that are **faithful by design**



Outputs: Probability of class i given input image x:
$Pr(i|x) = \mathrm{softmax}_i \left(\sum_{c=1}^{C} \beta_c \; \gamma_c(x) \right)$ (Equation 2)

ConceptTransformers: interpretability by design

- The attention weights computed by ConceptTransformer are concept-based explanations that are **faithful by design**



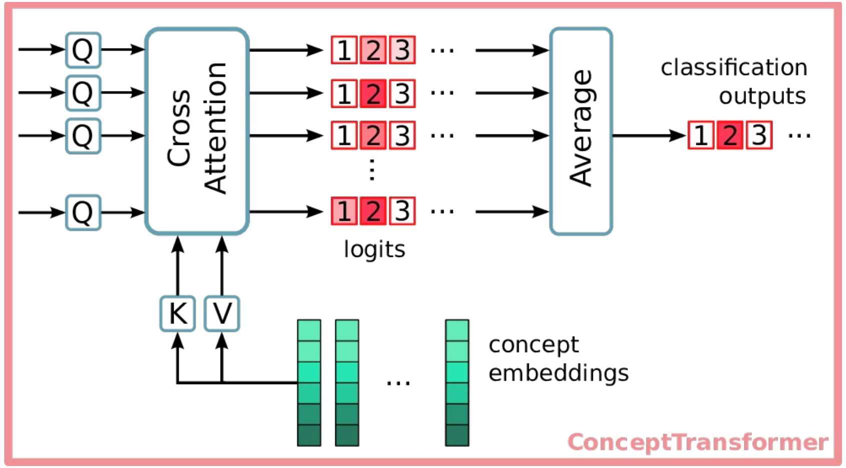
Outputs: Probability of class i given input image x:
$Pr(i|x) = \mathrm{softmax}_i \left(\sum_{c=1}^{C} \beta_c \, \gamma_c(x) \right)$ (Equation 2)

- with $(\beta_c)_i = [VO]_{ci}$ vectors of weights that are fixed after training,
- and $\gamma_c(x) = \frac{1}{P} \sum_{p=1}^{P} \alpha_{pc}$,
- where α_{pc} is attention weight from patch p to concept c

$$\alpha_{pc} = \mathrm{softmax} \left(\frac{1}{\sqrt{d_m}} Q \, K^\top \right)_{pc}$$

ConceptTransformers: interpretability by design

- The attention weights computed by ConceptTransformer are concept-based explanations that are **faithful by design**



Outputs: Probability of class i given input image x:

$Pr(i|x) = \mathrm{softmax}_i \left(\sum_{c=1}^{C} \beta_c \, \gamma_c(x) \right)$ (Equation 2)

- with $(\beta_c)_i = [VO]_{ci}$ vectors of weights that are fixed after training,
- and $\gamma_c(x) = \frac{1}{P} \sum_{p=1}^{P} \alpha_{pc}$,
- where α_{pc} is attention weight from patch p to concept c

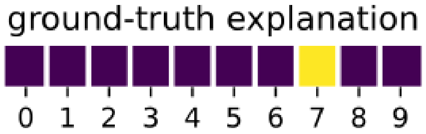
$$\alpha_{pc} = \mathrm{softmax} \left(\frac{1}{\sqrt{d_m}} Q \, K^\top \right)_{pc}$$

Proposition 1 *Each concept relevance score $\gamma_c(x)$ in Equation 2 is a* faithful *explanation of the output. More specifically, the probability of choosing the preferred output $i^c = \arg\max_i (\beta_c)_i$ of concept c (assuming it's unique) is guaranteed to decrease if $\gamma_c(x)$ is externally set to zero. Moreover, the correlation between $\gamma_c(x)$ and $Pr(i^c|x)$ is strictly positive.*

$\longrightarrow$ Decreasing $\gamma_c(x)$, changes the prob of a decision in a predictable way

ConceptTransformers: MNIST even/odd toy example

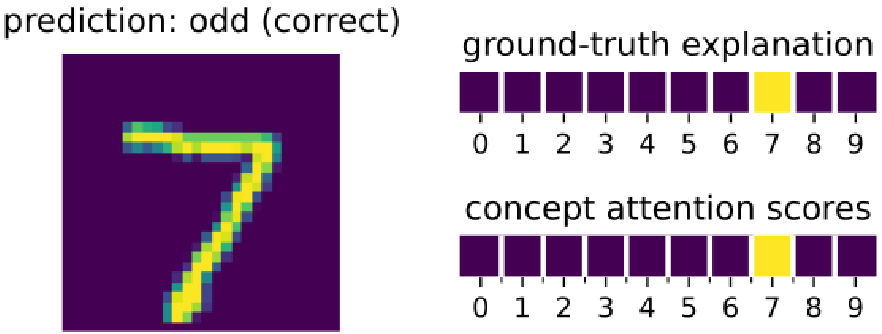
- **Task**: Classify MNIST images of digits as *even* or *odd* (binary task)
- **Explanatory concepts:** Digit from ‘0’ to ‘9’
- **Architecture**: CT with only one patch (since it’s not a spatial task)



“it’s odd because it’s a ‘7’”

ConceptTransformers: MNIST even/odd toy example

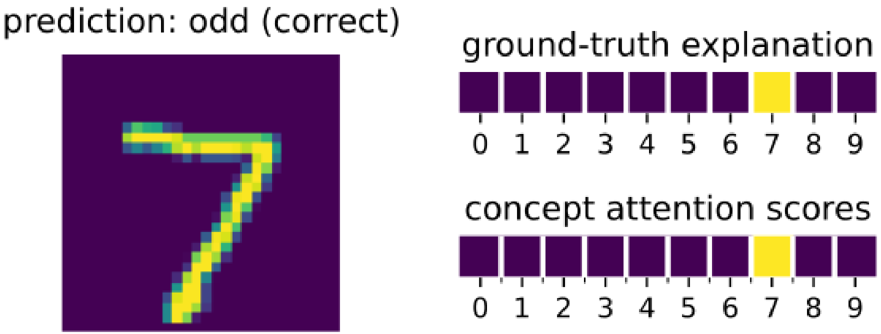
- **Task**: Classify MNIST images of digits as *even* or *odd* (binary task)
- **Explanatory concepts:** Digit from ‘0’ to ‘9’
- **Architecture**: CT with only one patch (since it’s not a spatial task)



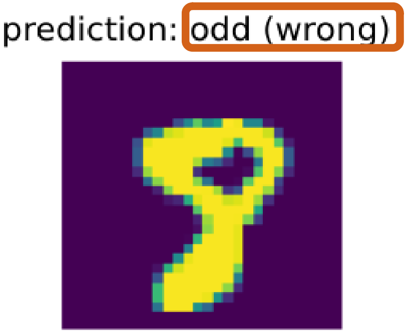
“it’s odd because it’s a ‘7’”

ConceptTransformers: MNIST even/odd toy example

- **Task**: Classify MNIST images of digits as *even* or *odd* (binary task)
- **Explanatory concepts:** Digit from ‘0’ to ‘9’
- **Architecture**: CT with only one patch (since it’s not a spatial task)



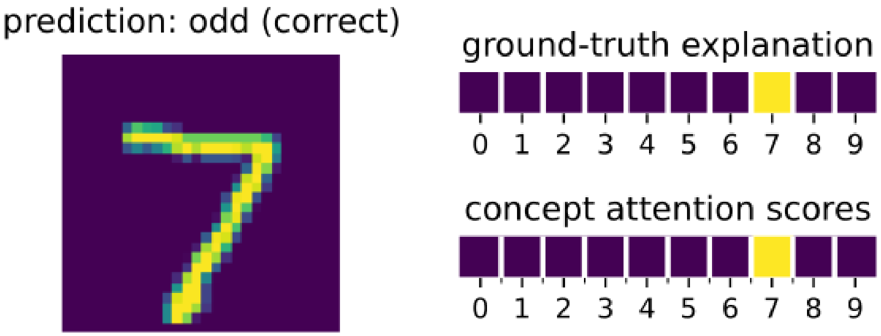
“it’s odd because it’s a ‘7’”



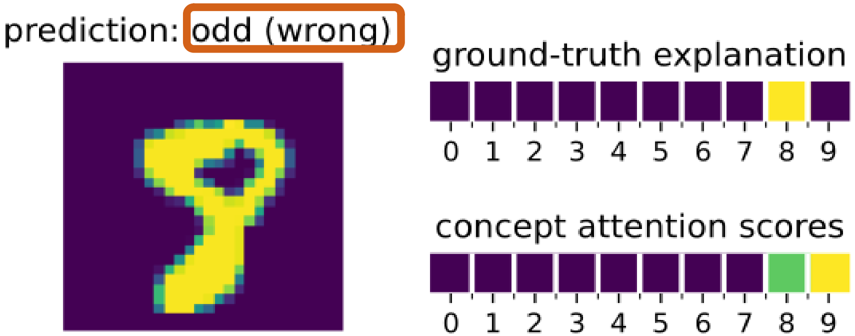
What is the reason for the mistake?

ConceptTransformers: MNIST even/odd toy example

- **Task**: Classify MNIST images of digits as *even* or *odd* (binary task)
- **Explanatory concepts:** Digit from ‘0’ to ‘9’
- **Architecture**: CT with only one patch (since it’s not a spatial task)



“it’s odd because it’s a ‘7’”

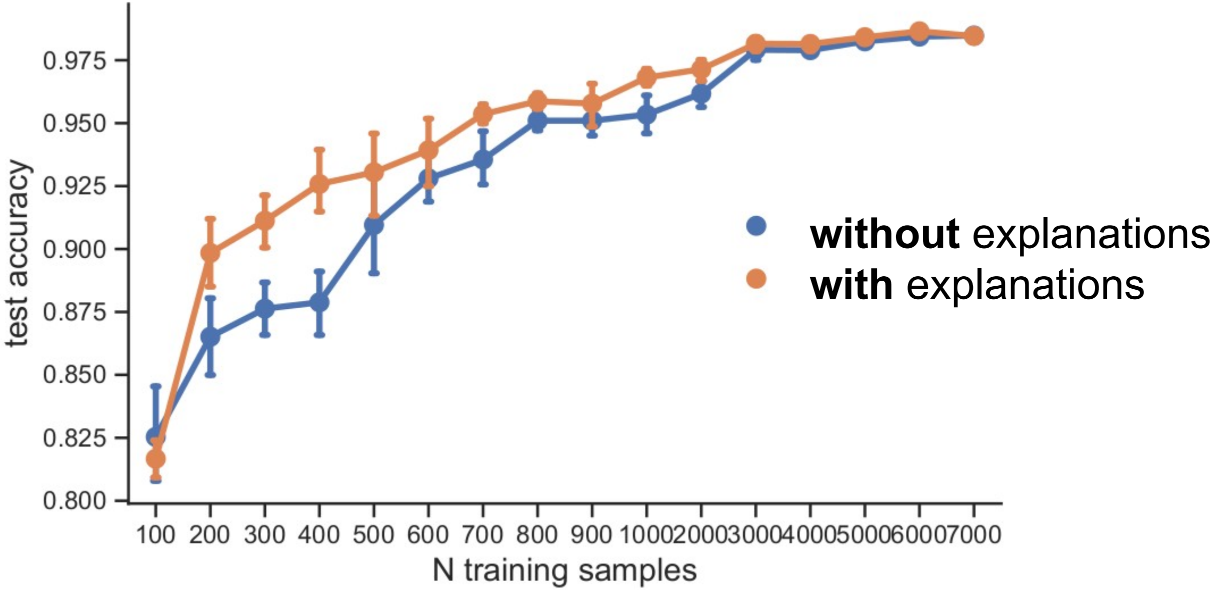


What is the reason for the mistake?

***Faithfulness in practice**: mistakes can be inspected and understood*

ConceptTransformers: MNIST even/odd toy example

- **Task**: Classify MNIST images of digits as *even* or *odd* (binary task)
- **Explanatory concepts:** Digit from ‘0’ to ‘9’
- **Architecture**: CT with only one patch (since it’s not a spatial task)



- Training with concept-based explanation is more training efficient
- → higher performance with less data

ConceptTransformers on CUB-200-2011 Birds Classification task

- **Task**: Classify Bird among 200 species
- **Explanatory concepts**: Anatomical parts and corresponding attributes
- **Architecture**: VisionTransformer with CT classification head

Prediction: Black_footed_Albatross (*correct*)

```
Spatial explanations:
  - has_eye_color::black
Global explanations:
  - has_primary_color::brown
```

Prediction: Groove_billed_Ani (*correct*)

```
Spatial explanations:
  - has_eye_color::black
  - has_forehead_color::black
  - has_leg_color::grey
  - has_leg_color::black
Global explanations:
  - has_primary_color::black
  - has_size::small_(5_-_9_in)
```

Prediction: Indigo_Bunting (*correct*)

```
Spatial explanations:
  - has_breast_pattern::solid
  - has_eye_color::black
  - has_forehead_color::blue
  - has_belly_pattern::solid
  - has_leg_color::black
Global explanations:
  - has_primary_color::blue
```

Prediction: Black_billed_Cuckoo (*correct*)

```
Spatial explanations:
  - has_eye_color::black
  - has_forehead_color::brown
  - has_leg_color::grey
  - has_leg_color::black
Global explanations:
  - has_primary_color::brown
  - has_size::small_(5_-_9_in)
```

ConceptTransformers on CUB-200-2011 Birds Classification task

- **Task**: Classify Bird among 200 species
- **Explanatory concepts**: Anatomical parts and corresponding attributes
- **Architecture**: VisionTransformer with CT classification head

Training	Accuracy [%]			
Multi-stage	Part R-CNN: 76.4	PS-CNN: 76.2	**PN-CNN: 85.4**	SPDA-CNN: 85.1
	PA-CNN: 82.8	MG-CNN: 83.0	2-level attn.: 77.9	FCAN: 82.0
	Neural const.: 81.0	ProtoPNet: 84.8		
End-to-end	B-CNN: 85.1	CAM: 70.5	DeepLAC: 80.3	ST-CNN: 84.1
	MA-CNN: 86.5	RA-CNN: 85.3	CT [w/o]: 76.9$\pm$3	**CT: 88.0$\pm$0.4**

We obtain state-of-the-art classification performance on CUB-2011
(with a simpler training procedure than the second best)

ConceptTransformers on CUB-200-2011 Birds Classification task

- **Task**: Classify Bird among 200 species
- **Explanatory concepts**: Anatomical parts and corresponding attributes
- **Architecture**: VisionTransformer with CT classification head

Published as a conference paper at ICLR 2022

ATTENTION-BASED INTERPRETABILITY WITH CONCEPT TRANSFORMERS

Mattia Rigotti, Christoph Miksovic, Ioana Giurgiu, Thomas Gschwind & Paolo Scotton
IBM Research
Zurich, Switzerland
{mrg,cmi,igi,thg,psc}@zurich.ibm.com

Drone/Robots visual inspection pipeline for civil infrastructures

Visual inspection pipeline

High-quality data acquisition

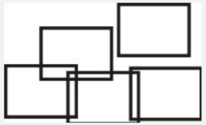
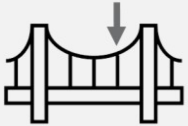


Photo stiching

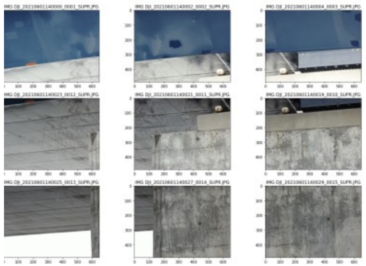


High-resolution detection of defects

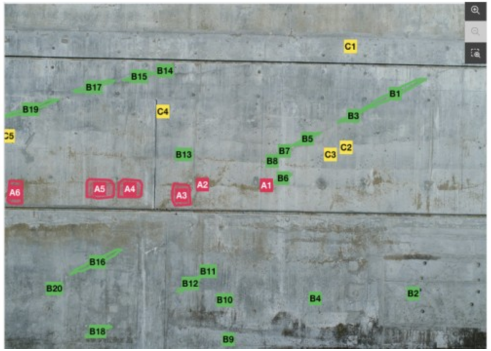
Defect measuring

Automated acquisition of thousands high-resolution images with repeatable drones and robots missions

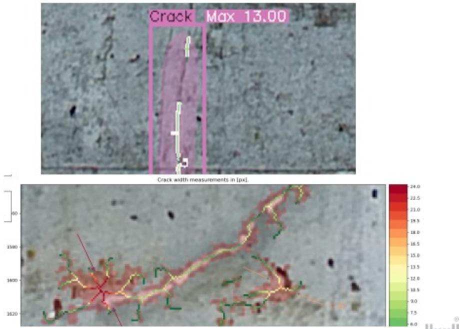
Automated 2D rendering of large elements (pillars, towers, etc.) and defects disambiguation and localization in global coordinate system



Deep learning based computer vision methods to detect tiny and rare defects on each high-res image



Future: explainable suggestions for maintenance decisions

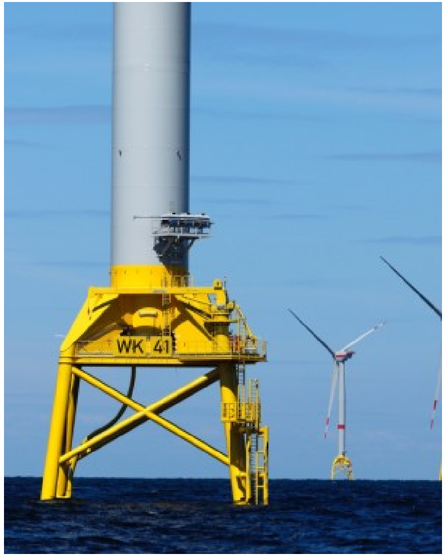


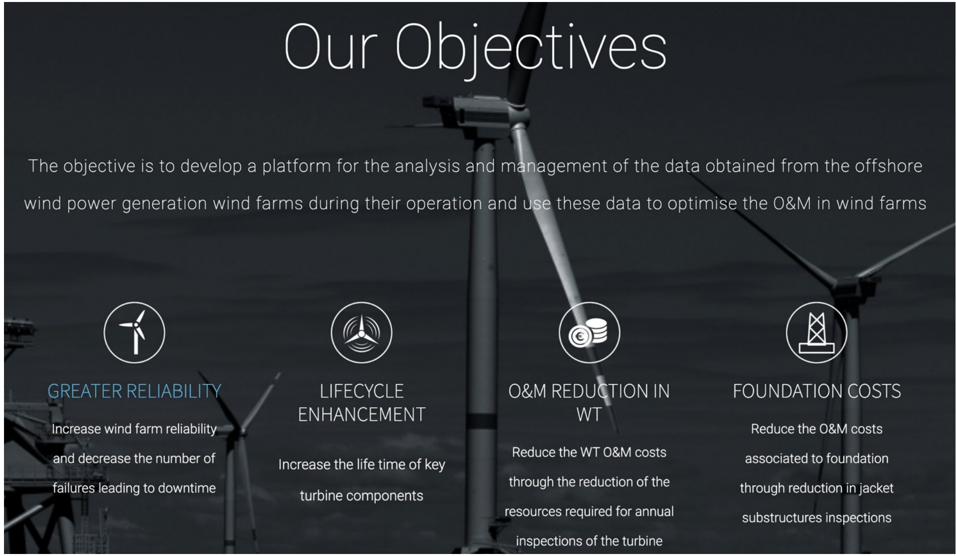
Innovation in EU key projects with civil engineering partners

ROMEO

IBM Research develop ML models to detect and predict Wind Turbine component failures based on a rich set of historic data provided by partners and design a reference Cloud implementation of the end-to-end data flow

www.romeoproject.eu





IM-SAFE.EU

IBM Research defines the digitalization avenue for future of civil infrastructures monitoring and inspection, leveraging AI, cloud, and automation

https://im-safe-project.eu

THANK YOU!

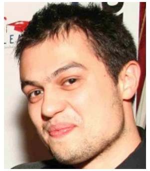
IBM Research


Florian Scheidegger

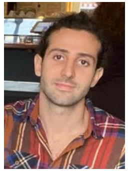

Thomas Frick


Filip Janicki

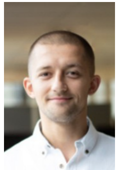

Ioana Giurgiu


Mattia Rigotti

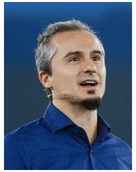

Andrea Bartezzaghi


Roy Assaf


Panagiotis Chatzidoukas

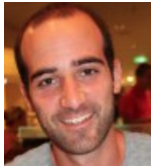

Cezary Skura


Chiara Marchiori

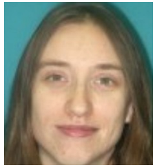

Cristiano Malossi


Samuel Kallner

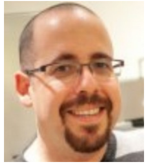

Dmitri Pikus


Gil Sharon


Maya Barnea


Ira Rosen


Benny Mandler


Natti Eder


Nili Guy

Civil Engineering Partners

Sund≈Bælt

SACERTIS