# Segmenting without Annotating: Crack Segmentation and Monitoring via Post-hoc Classifier Explanations

IMC 2023

**Florent Forest**[1], Hugo Porta[2], Devis Tuia[2], Olga Fink[1]

[1] Intelligent Maintenance and Operations Systems (IMOS), EPFL, Switzerland

[2] Environmental Computational Science and Earth Observation (ECEO), EPFL, Switzerland

first.last@epfl.ch

September 12, 2023

# Overview

# Introduction

# Automatic visual inspection for infrastructure condition monitoring




**Problem statement**

Detection and monitoring of **surface cracks** in infrastructure elements.

# Automatic visual inspection for infrastructure condition monitoring

**Problem statement**

Detection and monitoring of **surface cracks** in infrastructure elements.

Manual visual inspection:

- Limited availability
- Inspector subjectivity
- Service interruptions
- Hard-to-access or hazardous locations

# Automatic visual inspection for infrastructure condition monitoring

**Problem statement**

Detection and monitoring of **surface cracks** in infrastructure elements.

Manual visual inspection:

- Limited availability
- Inspector subjectivity
- Service interruptions
- Hard-to-access or hazardous locations

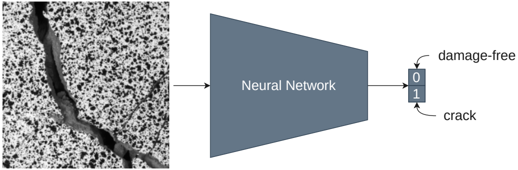
→ Automatic visual inspection

# Machine learning for image-based crack detection

Data-driven approaches based on supervised deep learning have demonstrated excellent performance in detection of cracks in images, but they require **large annotated datasets** for training.

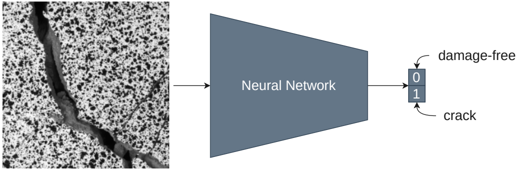
Data-driven approaches based on supervised deep learning have demonstrated excellent performance in detection of cracks in images, but they require **large annotated datasets** for training.

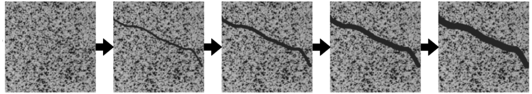
**Classification task:**



Data-driven approaches based on supervised deep learning have demonstrated excellent performance in detection of cracks in images, but they require **large annotated datasets** for training.

**Classification task:**



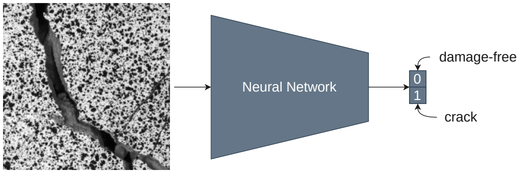



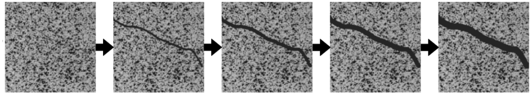
Severity quantification and monitoring is crucial for timely decision-making.

# Machine learning for image-based crack detection

Data-driven approaches based on supervised deep learning have demonstrated excellent performance in detection of cracks in images, but they require **large annotated datasets** for training.

**Classification task:**



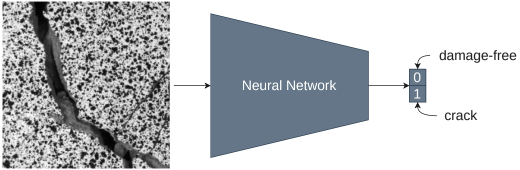
Severity quantification and monitoring is crucial for timely decision-making.

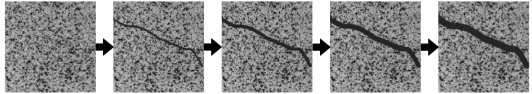
What we need: 

# Machine learning for image-based crack detection

Data-driven approaches based on supervised deep learning have demonstrated excellent performance in detection of cracks in images, but they require **large annotated datasets** for training.

**Classification task:**



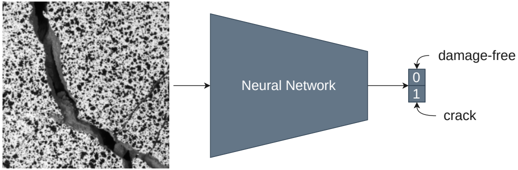



Severity quantification and monitoring is crucial for timely decision-making.

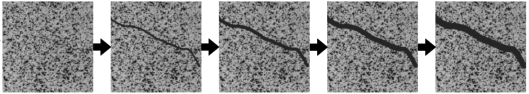
What we need:  → Severity metrics

Data-driven approaches based on supervised deep learning have demonstrated excellent performance in detection of cracks in images, but they require **large annotated datasets** for training.

**Classification task:**




- Does not allow severity quantification
- Fast and easy image-level annotation (1 bit)

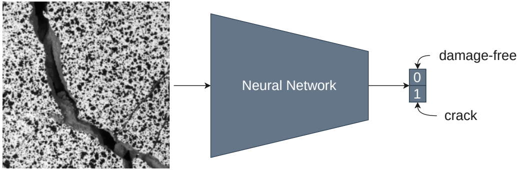


Severity quantification and monitoring is crucial for timely decision-making.

What we need:  → Severity metrics

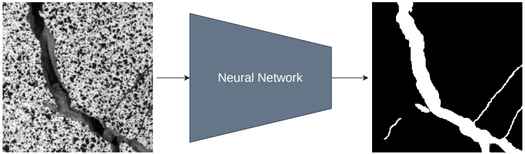
Data-driven approaches based on supervised deep learning have demonstrated excellent performance in detection of cracks in images, but they require **large annotated datasets** for training.

**Classification task:**




- Does not allow severity quantification
- Fast and easy image-level annotation (1 bit)

**Semantic segmentation task:**




- Allows severity quantification and monitoring
- Tedious and costly pixel-level annotation ($256 \times 256 \rightarrow 2^{16} = 64$ Kb)

# Segmenting without annotating using explainable AI

Segmentation algorithms are data-hungry, and pixel-level labeling is tedious and costly.
$\rightarrow$ Barrier to the deployment of automated crack segmentation systems.

**Research question**

Can we obtain image segmentations while avoiding pixel-level annotation?

# Segmenting without annotating using explainable AI

Segmentation algorithms are data-hungry, and pixel-level labeling is tedious and costly.
$\rightarrow$ Barrier to the deployment of automated crack segmentation systems.

**Research question**

Can we obtain image segmentations while avoiding pixel-level annotation?

**Weakly-supervised segmentation with explainable AI (XAI)**

# Segmenting without annotating using explainable AI

Segmentation algorithms are data-hungry, and pixel-level labeling is tedious and costly.
→ Barrier to the deployment of automated crack segmentation systems.

**Research question**

Can we obtain image segmentations while avoiding pixel-level annotation?

**Weakly-supervised segmentation with explainable AI (XAI)**

1. Train a classifier to discriminate between damage-free and cracked samples
   - weakly-supervised (image-level labels)

# Segmenting without annotating using explainable AI

Segmentation algorithms are data-hungry, and pixel-level labeling is tedious and costly.
→ Barrier to the deployment of automated crack segmentation systems.

**Research question**

Can we obtain image segmentations while avoiding pixel-level annotation?

**Weakly-supervised segmentation with explainable AI (XAI)**

1. Train a classifier to discriminate between damage-free and cracked samples
2. Find which pixels are contributing to the crack class (**attribution maps**)

- weakly-supervised (image-level labels)
- post-hoc XAI techniques [1]

[1] A. B. Arrieta *et al.*, "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI" in *Information fusion*, 2019.

Segmentation algorithms are data-hungry, and pixel-level labeling is tedious and costly.
$\rightarrow$ Barrier to the deployment of automated crack segmentation systems.

**Research question**

Can we obtain image segmentations while avoiding pixel-level annotation?

**Weakly-supervised segmentation with explainable AI (XAI)**

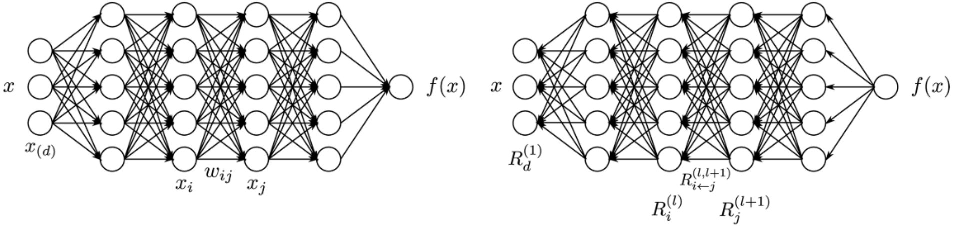
1. Train a classifier to discriminate between damage-free and cracked samples
2. Find which pixels are contributing to the crack class (**attribution maps**)
3. Extract approximate segmentation masks

- weakly-supervised (image-level labels)
- post-hoc XAI techniques [1]
- expected match between attributions and segmentation

Previous work applied Layer-wise Relevance Propagation (LRP) for damage segmentation [2], but comparison between $\neq$ XAI methods and severity quantification is lacking.

[1] A. B. Arrieta *et al.*, "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI" in *Information fusion*, 2019.
[2] C. Seibold *et al.*, "From Explanations to Segmentation: Using Explainable AI for Image Segmentation" in *17th International Conference on Computer Vision Theory and Applications (VISAPP)*, 2022.

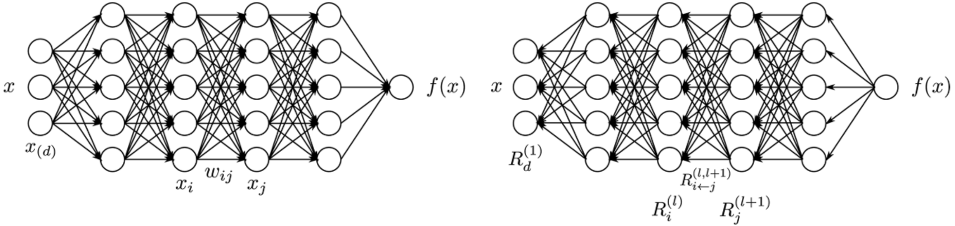
# One example: Layer-wise Relevance Propagation




Layer-wise Relevance Propagation (LRP) propagates **relevance scores** from layer $l+1$ to $l$ in a backward pass, using *messages* $R_{i\leftarrow j}^{(l,l+1)}$ and *propagation rules* [3].

[3] G. Montavon *et al.*, "Layer-Wise Relevance Propagation: An Overview" in *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, 2019.

# One example: Layer-wise Relevance Propagation




Layer-wise Relevance Propagation (LRP) propagates **relevance scores** from layer $l+1$ to $l$ in a backward pass, using *messages* $R_{i\leftarrow j}^{(l,l+1)}$ and *propagation rules* [3].

- Conservation property: $\sum_i R_{i\leftarrow j}^{(l,l+1)} = R_j^{(l+1)}$
- Easy for linear networks $x_j = \sum_i x_i w_{ij}$: $R_{i\leftarrow j} = x_i w_{ij}$
- For non-linear networks $x_j = g(\sum_i x_i w_{ij} + b_j)$, we only have rules with **approximate conservation**.

- LRP-$\epsilon$: $R_i = \sum_j \frac{x_i w_{ij}}{\epsilon + \sum_{\mathbf{0},i} x_i w_{ij}} R_j$
- LRP-$\gamma$: $R_i = \sum_j \frac{x_i (w_{ij} + \gamma w_{ij}^+)}{\sum_{\mathbf{0},i} x_i (w_{ij} + \gamma w_{ij}^+)} R_j$
- LRP-$\alpha\beta$: $R_i = \sum_j \left( \alpha \frac{(x_i w_{ij})^+}{\sum_{\mathbf{0},i} (x_i w_{ij})^+} + \beta \frac{(x_i w_{ij})^-}{\sum_{\mathbf{0},i} (x_i w_{ij})^-} \right) R_j$
- $z^{\mathcal{B}}$-rule: $R_i = \sum_j \frac{x_i w_{ij} - l_i w_{ij}^+ - h_i w_{ij}^-)}{\sum_i x_i w_{ij} - l_i w_{ij}^+ - h_i w_{ij}^-)} R_j$

[3] G. Montavon *et al.*, “Layer-Wise Relevance Propagation: An Overview” in *Explainable AI: Interpreting, Explaining and*

## Methodology

### Main contributions

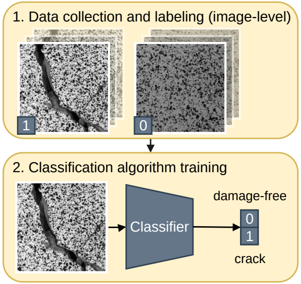
- We evaluate and compare several post-hoc XAI methods.
- We investigate damage severity quantification and growth monitoring.

# Methodology

## Main contributions

- We evaluate and compare several post-hoc XAI methods.
- We investigate damage severity quantification and growth monitoring.

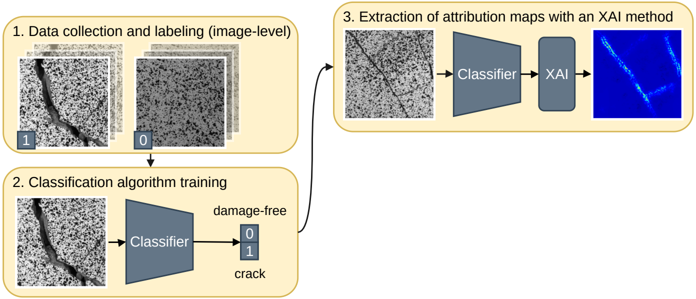
**Proposed methodology:**



## Main contributions

- We evaluate and compare several post-hoc XAI methods.
- We investigate damage severity quantification and growth monitoring.

## Proposed methodology:

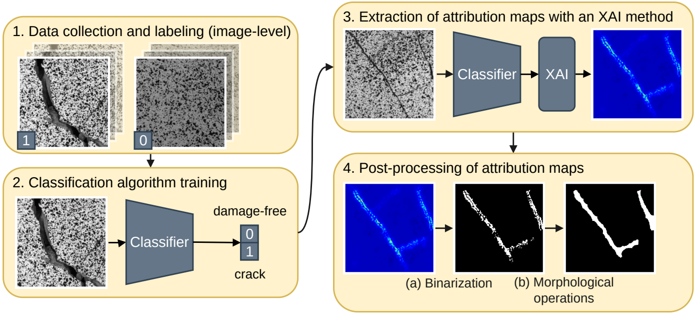


# Methodology

## Main contributions

- We evaluate and compare several post-hoc XAI methods.
- We investigate damage severity quantification and growth monitoring.

**Proposed methodology:**

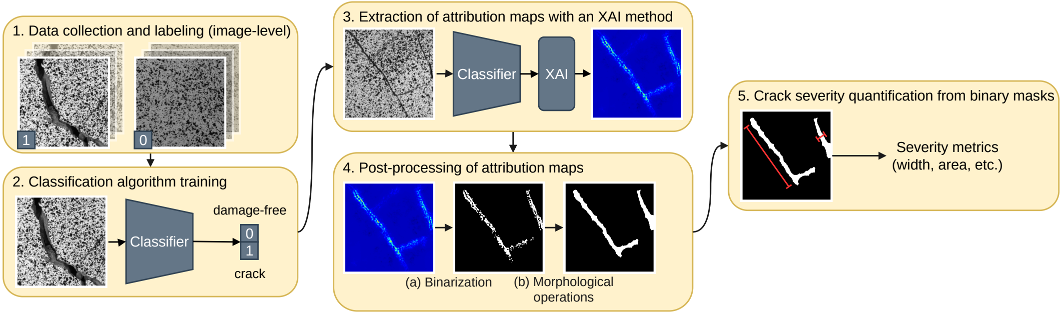


# Methodology

## Main contributions

- We evaluate and compare several post-hoc XAI methods.
- We investigate damage severity quantification and growth monitoring.

## Proposed methodology:



# Experimental settings

# Compared methods

XAI methods (weakly-supervised)

- Input×Gradient [4]
- Integrated Gradients [5]
- DeepLift [6]
- DeepLiftShap, GradientShap [7]
- Layer-wise Relevance Propagation [8]

Unsupervised methods

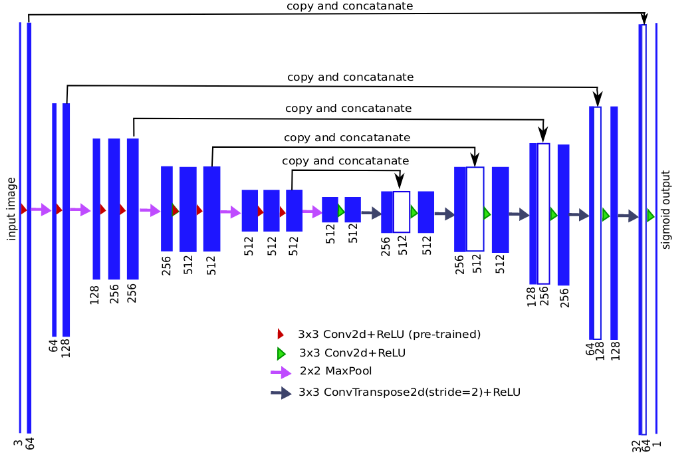
- Raw image pixels
- Convolutional Autoencoder (CAE) residuals

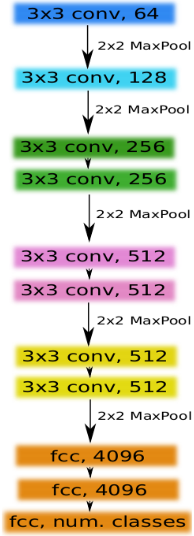
Supervised method

- U-Net (oracle trained on pixel-level labels)

[4] D. Baehrens *et al.*, "How to Explain Individual Classification Decisions" in *Journal of Machine Learning Research*, 2010.
[5] M. Sundararajan *et al.*, *Axiomatic Attribution for Deep Networks* in, 2017.
[6] A. Shrikumar *et al.*, *Learning Important Features Through Propagating Activation Differences* in, 2019.
[7] S. M. Lundberg *et al.*, "A Unified Approach to Interpreting Model Predictions" in *Advances in Neural Information Processing Systems*, 2017.
[8] S. Bach *et al.*, "On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation" in *PloS One*, 2015.

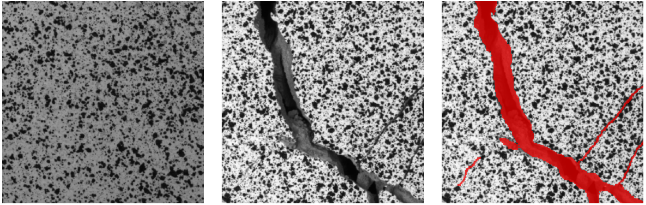
- **Classifier**: VGG11-128 (VGG11 with 128 neurons in FC layers)
- **CAE**: VGG11 encoder and symmetrical decoder
- **U-Net**: U-Net11 (VGG11 encoder)



# Dataset

- Experimental DIC cracks dataset [9], 256×256 image patches from stone masonry walls damaged in a shear-compression experiment conducted at the EESD EPFL laboratory.
- Annotated segmentation masks for the cracked image patches (used for evaluation only). To perform binary classification, we added 874 negative patches coming from the same walls.



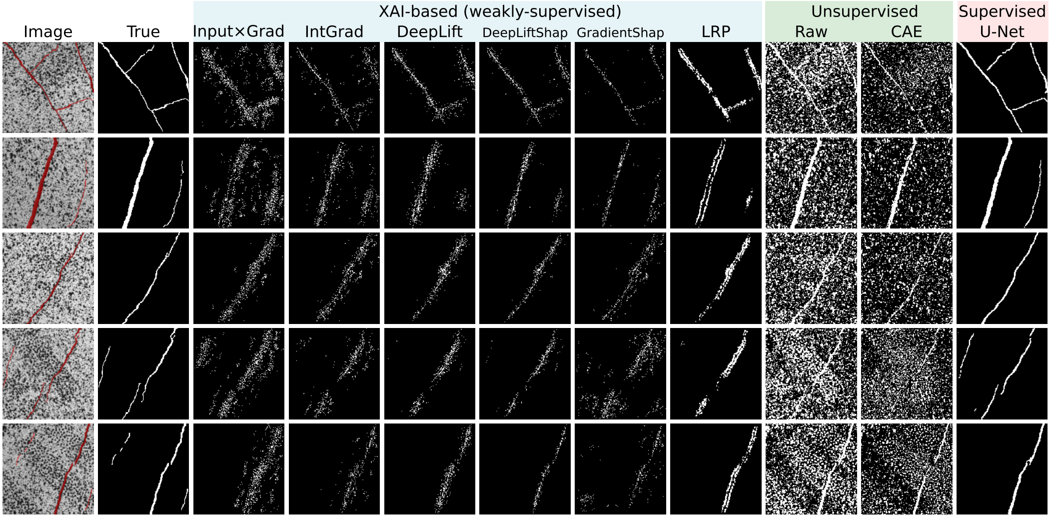
[9] A. Rezaie *et al.*, "Comparison of crack segmentation using digital image correlation measurements and deep learning" in *Construction and Building Materials*, 2020.

# Experimental results

# Experimental results

## Segmentation

# Qualitative results | Visualization after binarization



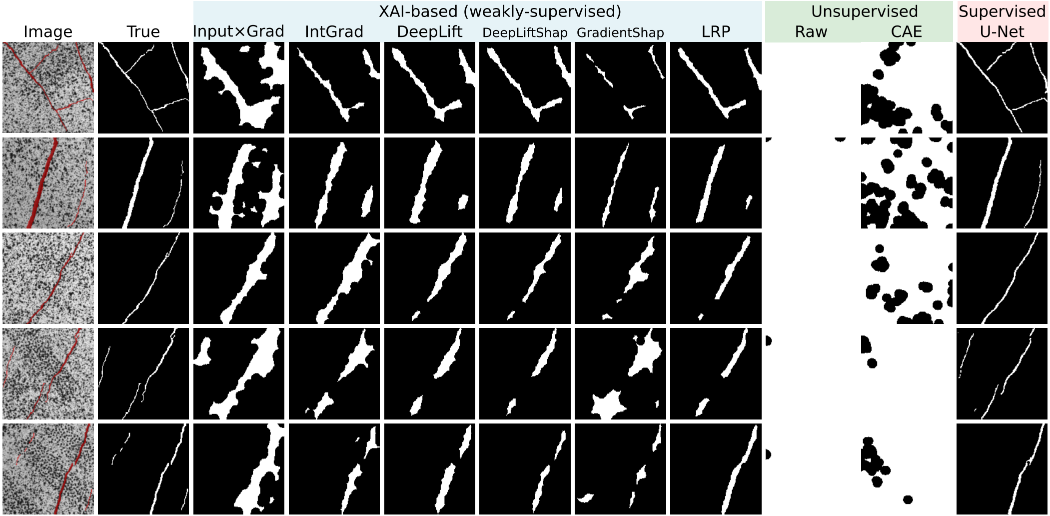

XAI-based (weakly-supervised)
Unsupervised
Supervised
Image
True
Input×Grad
IntGrad
DeepLift
DeepLiftShap
GradientShap
LRP
Raw
CAE
U-Net

**Table 1:** Crack segmentation quality evaluation (values in %).

| Method | F1 | Precision | Recall | IoU |
|---|---|---|---|---|
| Input×Gradient | 23.30 | 14.37 | 61.55 | 13.19 |
| IntGrad | 27.74 | 20.81 | 41.56 | 16.10 |
| DeepLift | 34.44 | 28.75 | 42.96 | 20.81 |
| DeepLiftShap | **38.19** | 36.37 | 40.21 | **23.60** |
| GradientShap | 20.61 | 14.09 | 38.38 | 11.49 |
| LRP | **37.43** | 35.06 | 40.16 | **23.03** |
| Raw pixels | 4.73 | 2.42 | 100.0 | 2.42 |
| CAE | 5.93 | 3.07 | 90.09 | 3.06 |
| U-Net | 83.67 | 82.22 | 85.17 | 71.93 |

XAI-based (weakly-supervised) | Unsupervised | Fully supervised

# Experimental results

## Severity quantification

**Severity metrics:** number of cracks per patch (CPP) [10], total crack area, maximum crack width [11].

**Table 2:** Crack severity quantification evaluation.

| Method | CPP MAE | Area MAPE | Width MAPE |
|---|---|---|---|
| Input×Gradient | 1.13 | 448.1 | 358.8 |
| IntGad | 0.94 | 271.3 | 268.9 |
| DeepLift | **0.81** | 146.0 | 264.6 |
| DeepliftShap | **<u>0.78</u>** | **103.6** | 189.2 |
| GradientShap | 1.76 | 338.8 | 295.5 |
| LRP | 0.90 | **<u>91.0</u>** | **<u>163.1</u>** |
| U-Net | 0.74 | 20.1 | 20.8 |

XAI-based (weakly-supervised) Fully supervised

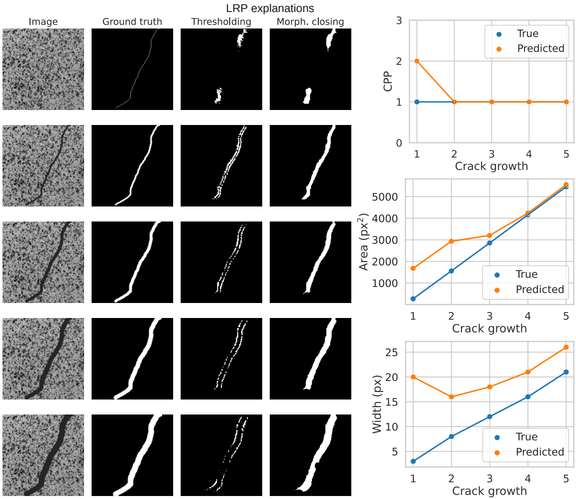
[10] B. G. Pantoja-Rosero *et al.*, "TOPO-Loss for continuity-preserving crack detection using deep learning" in *Construction and Building Materials*, 2022.
[11] M. Carrasco *et al.*, "Image-Based Automated Width Measurement of Surface Cracking" in *Sensors*, 2021.

# Experimental results

## Growth monitoring

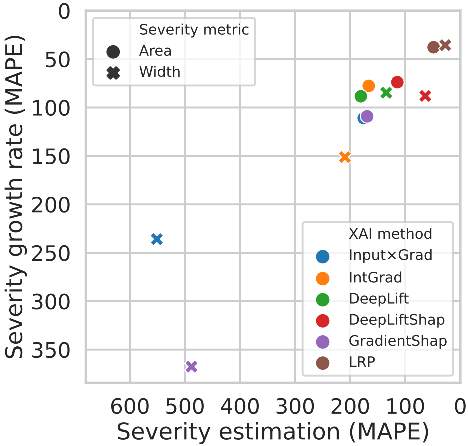
# Growth monitoring experiment

Simulation of 100 **artificial linear growth trajectories** of cracks.



# Growth monitoring experiment

| Method | Area growth | |
|---|---|---|
| | Average *r* | Slope MAPE |
| Input×Gradient | -0.32 | 235.9 |
| IntGrad | 0.77 | 151.3 |
| DeepLift | 0.44 | **84.7** |
| DeepLiftShap | **0.90** | 88.0 |
| GradientShap | 0.33 | 367.7 |
| LRP | **0.84** | **35.6** |
| **Method** | **Width growth** | |
| | Average *r* | Slope MAPE |
| Input×Gradient | -0.03 | 110.1 |
| IntGrad | 0.22 | 77.7 |
| DeepLift | 0.18 | 88.4 |
| DeepLiftShap | **0.71** | **73.8** |
| GradientShap | 0.07 | 109.1 |
| LRP | **0.80** | **37.8** |



# Conclusions and Future work

- Approximate segmentation masks can be obtained from the post-hoc explanations of a classifier using XAI methods.
- We evaluated the performance of 6 XAI methods in terms of segmentation quality, severity quantification and growth monitoring abilities.
- While quality is lower than supervised segmentation approaches, the labeling cost is significantly lower.
- The best-performing methods are LRP and DeepLift(Shap). By taking into account computational runtime, LRP offers the best solution.

# Conclusions and Future work

- Approximate segmentation masks can be obtained from the post-hoc explanations of a classifier using XAI methods.
- We evaluated the performance of 6 XAI methods in terms of segmentation quality, severity quantification and growth monitoring abilities.
- While quality is lower than supervised segmentation approaches, the labeling cost is significantly lower.
- The best-performing methods are LRP and DeepLift(Shap). By taking into account computational runtime, LRP offers the best solution.

Future work:

- Apply the methodology to different types of defects and infrastructures.
- Evaluate the approach using real crack growth data.
- Use approximate segmentations as coarse labels for supervised or semi-supervised segmentation.
- Investigate other families of explainable AI methods

Thanks for listening! Questions?