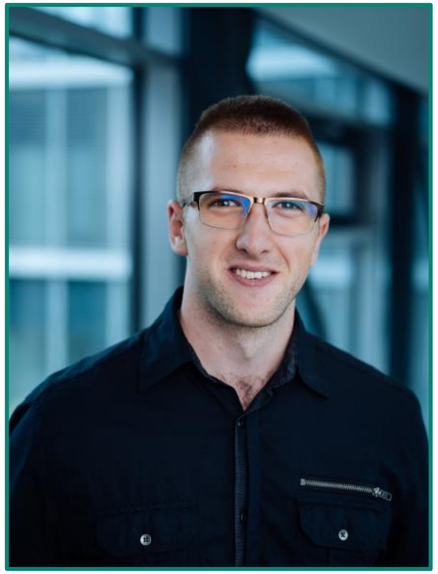
# Towards Explaining SEM Defect Image Classification

Vahidin Hasić, IPCEI PhD Student
Corinna Kofler, SFM KAI DSC
04.06.2025

# Academic & industrial collaborators



**Vahidin Hasić**

PhD Student

**Senka Krivić**

Academic Supervisor

**University of Sarajevo**
Faculty of Electrical Engineering, Computer Science & Informatics

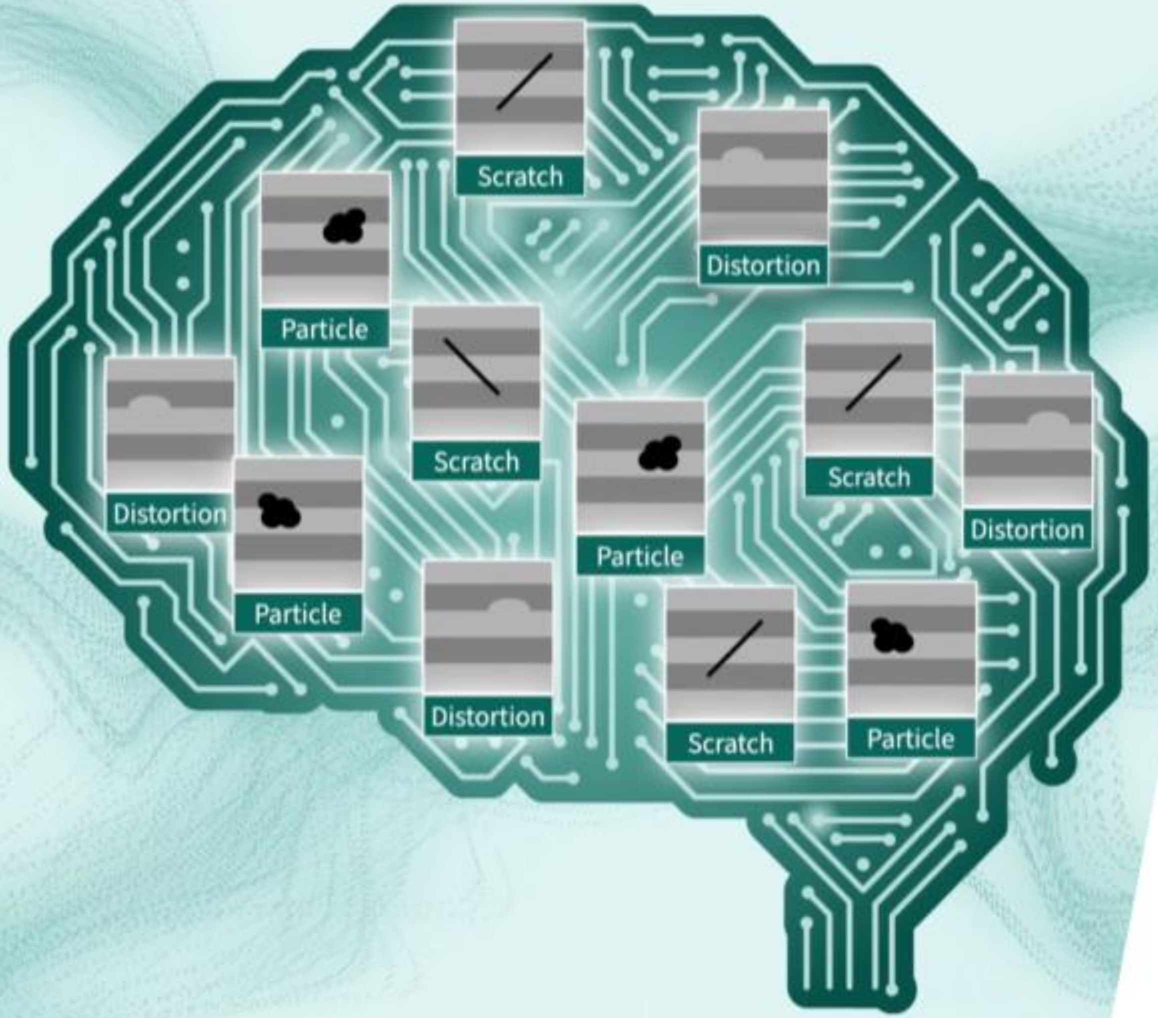
**Corinna Kofler**

Industrial Supervisor

**Olivia Pfeiler**

Head of KAI Data Science

**KAI GmbH**
Data Science, Computer Vision




# CoDIAC: AI-Powered Defect Image Classification

Objective classification

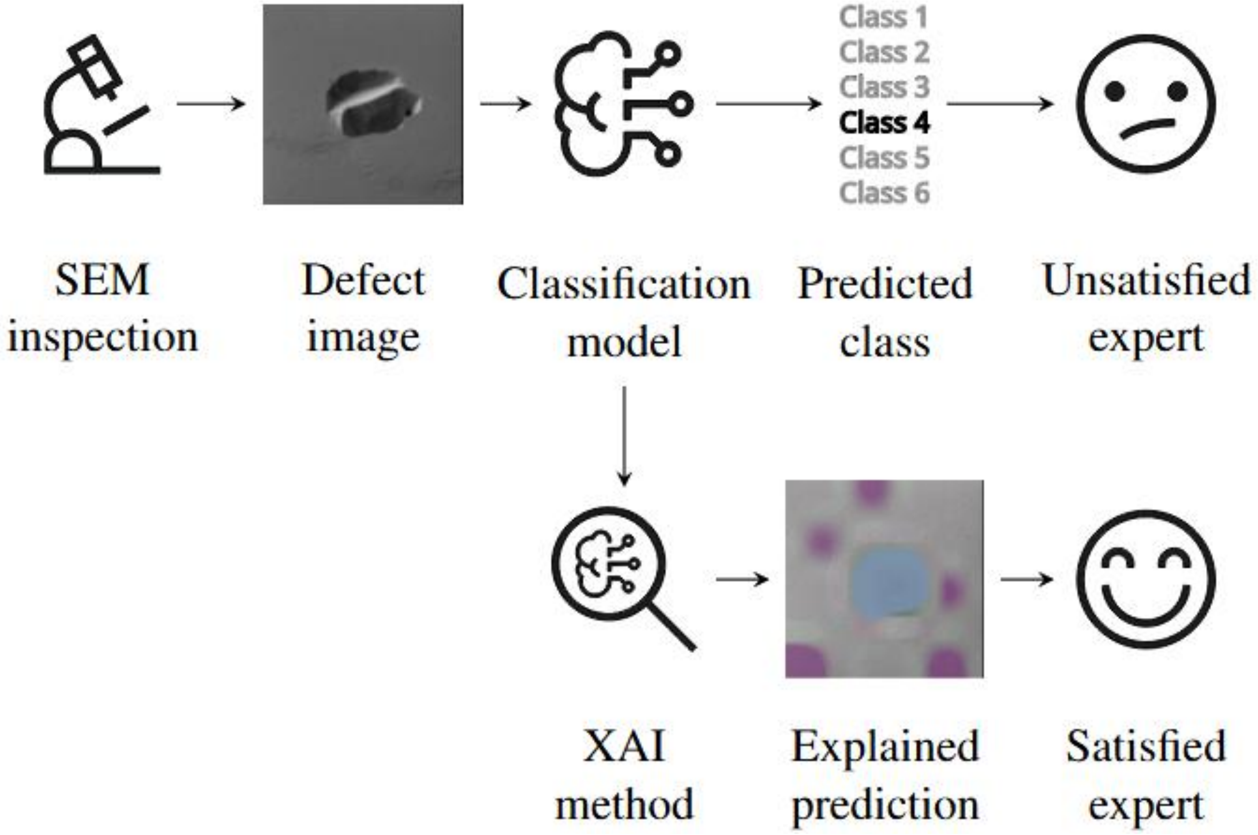
Time efficiency

Resource optimization

# Motivation

## Why explainable AI?

- Increased trust in AI models
- Improved debugging of AI models
- Increased transparency of AI models
- Regulatory and ethical concerns




Application paper under review at the European Conference on Artificial Intelligence (ECAI 2025)

# Dataset

## We used the public Carinthia dataset for our research

### Images

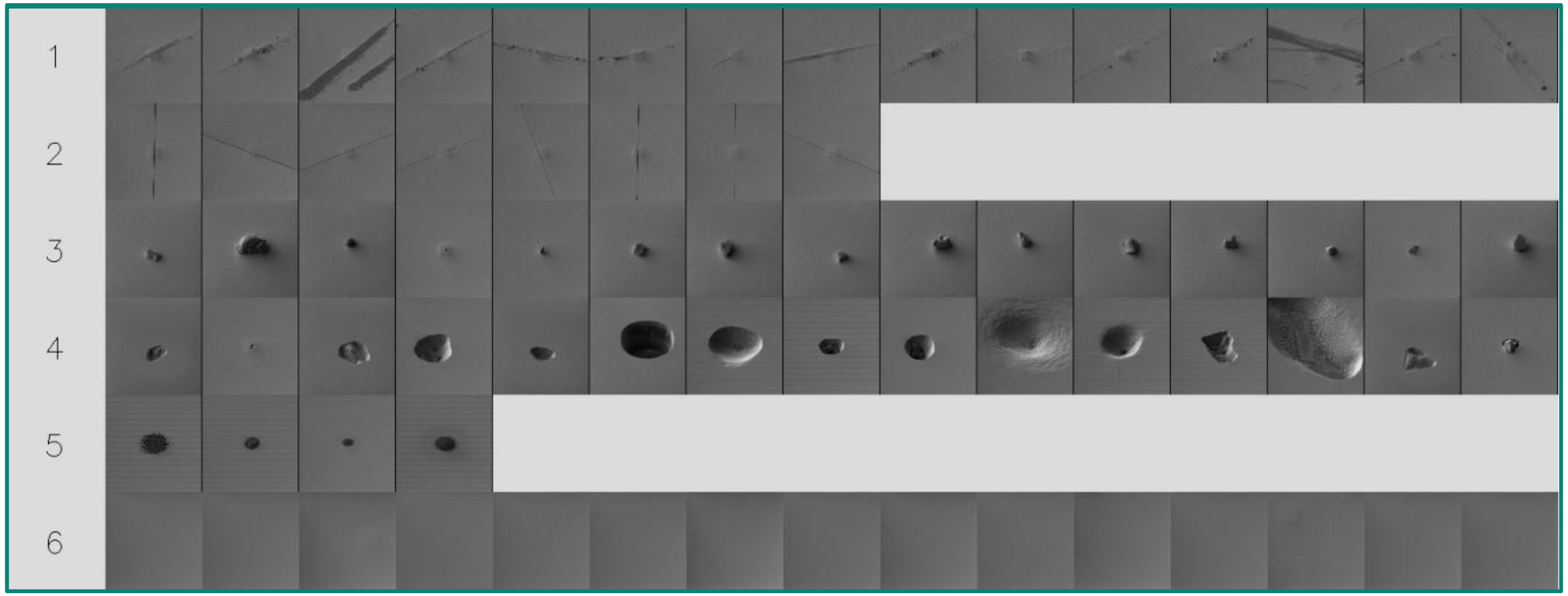
– Historical images from the production database

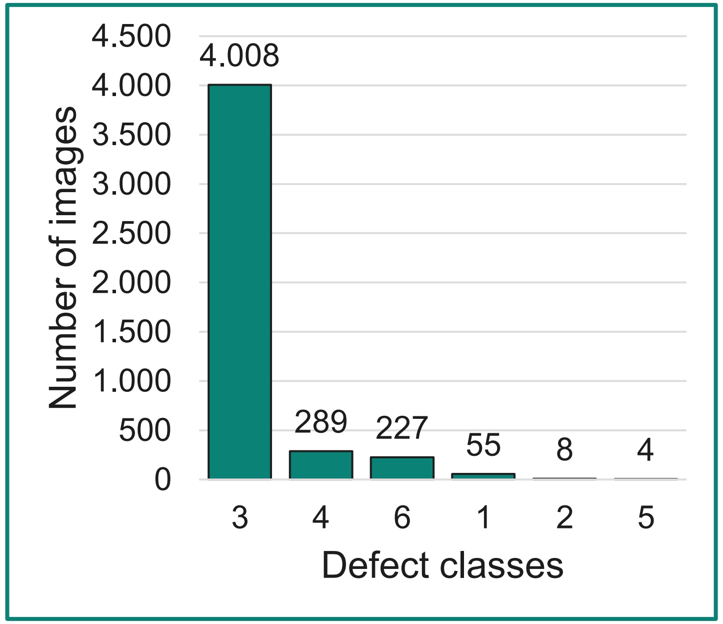
### Labeling

– Defined 6 expert defect classes
– Expert labeled ~4.600 images

### Published

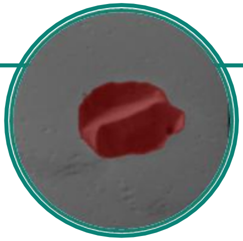
– Carinthia dataset publicly available on Zenodo



# Contributions

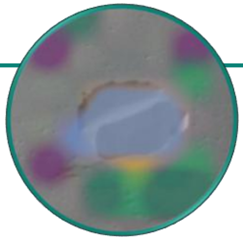
## Automatic mask generation



SAM2[1]

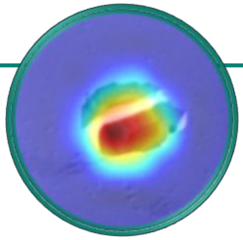
- Automatic mask generation using SAM2
- Expert validation of ground truth masks
- Carinthia-S dataset with ground truth segmentation masks

## Evaluation of XAI methods on the Carinthia-S dataset



CRAFT[2]

- Concept Recursive Activation FacTorization
- Concept-based explanations
- Provide explanations in form of human-understandable concepts



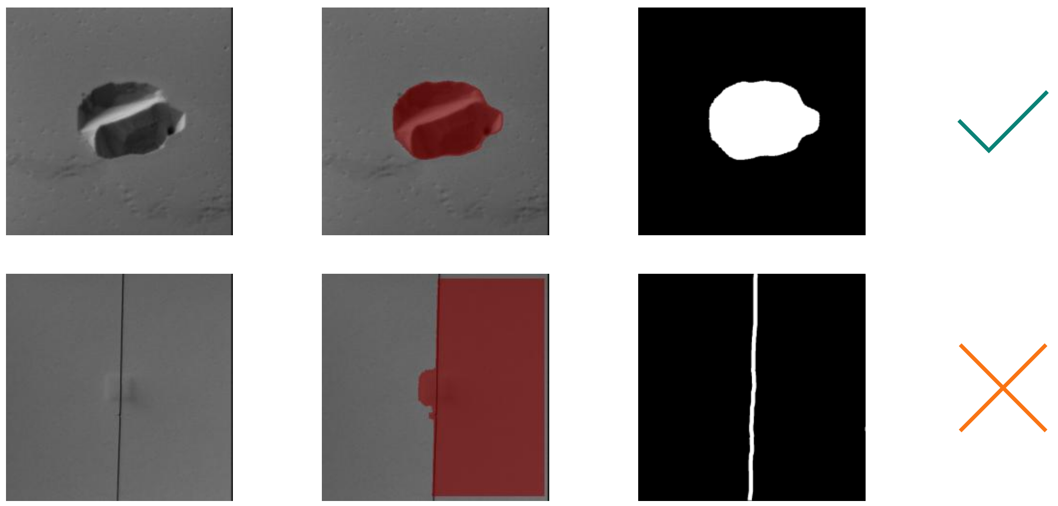
GradCAM[3]

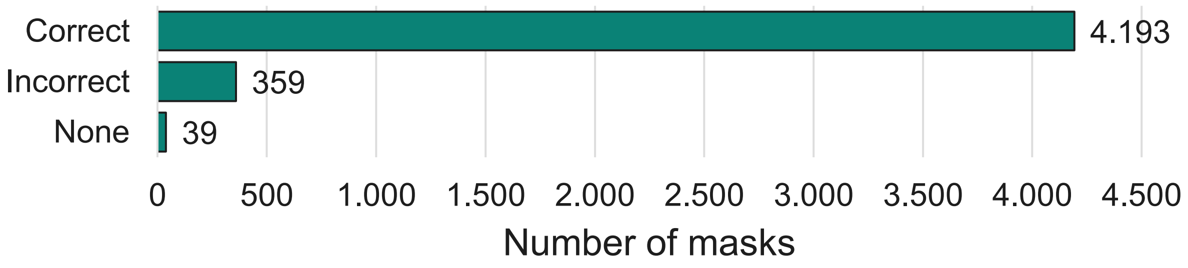
- Gradient-weighted class activation
- Feature-based explanations
- Provide explanations in form of saliency maps

1. Ravi N, Gabeur V, Hu YT, Hu R, Ryali C, Ma T, Khedr H, Rädle R, Rolland C, Gustafson L, Mintun E. Sam 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714. 2024 Aug 1.
2. Fel, Thomas, et al. "Craft: Concept recursive activation factorization for explainability." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.
3. Selvaraju, Ramprasaath R., et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization." Proceedings of the IEEE international conference on computer vision. 2017.

# Contributions

## Generating ground truth masks with SAM2






**Algorithm 1** Ground Truth Generation

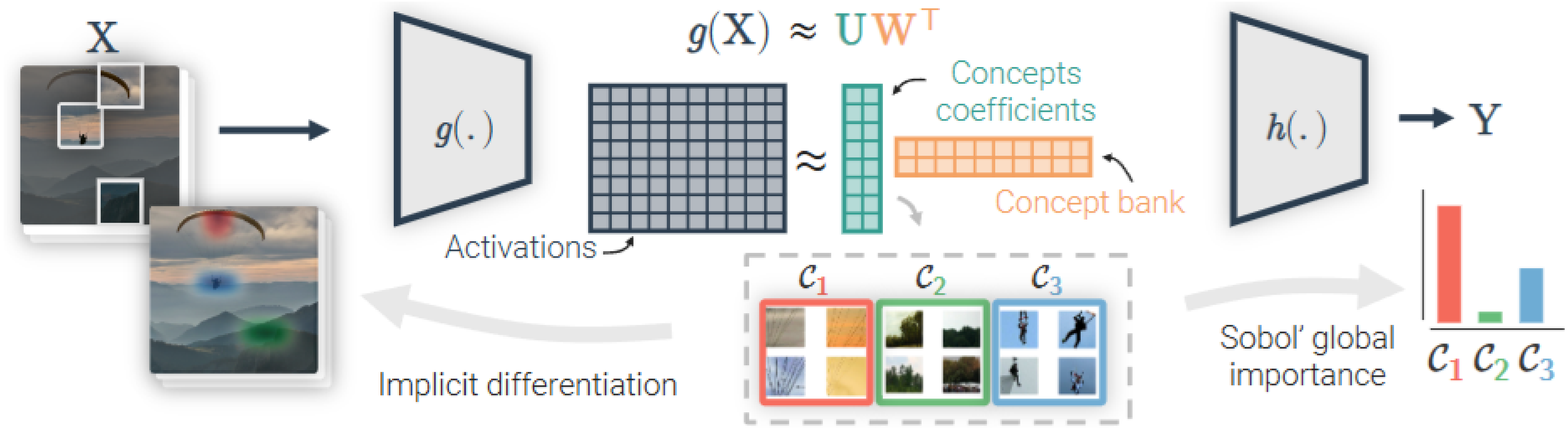
**Input:** Image $x \in \mathbb{R}^{H \times W \times 3}$
**Output:** Binary mask $M \in \{0, 1\}^{H \times W}$

**Stage 1:** $\mathcal{M} = \{M_1, M_2, \ldots, M_n\} \leftarrow \mathrm{SAM2}(x)$ {Generate candidate masks}
**for** $i = 1$ to $n$ **do**
  $A_i \leftarrow \sum_{p \in M_i} 1$ {Calculate area}
  $C_i \leftarrow \frac{1}{A_i} \sum_{p \in M_i} p$ {Calculate centroid}
  $d_i \leftarrow \frac{\|C_i - C_I\|}{\sqrt{H^2 + W^2}}$ {Normalized distance to image center}
  $S_i \leftarrow A_i \cdot (1 - d_i)^\alpha$ {$\alpha$: center bias parameter}
**end for**
$M_k \leftarrow \arg\max_i S_i$ s.t. $A_i > \beta \cdot H \cdot W$ {$\beta$: min area ratio}
**Stage 2:** $M' \leftarrow 1 - M_k$ {Invert mask}
$\{C_1, C_2, \ldots, C_m\} \leftarrow \mathrm{ConnectedComponents}(M')$
$j \leftarrow \arg\max_i |C_i|$ {Find largest component}
$M[p] \leftarrow \begin{cases} 1, & \text{if } p \in C_j \\ 0, & \text{otherwise} \end{cases}$ {Final binary mask}
**return** $M$
**Stage 3:** Experts manual validation of generated masks.


Ravi N, Gabeur V, Hu YT, Hu R, Ryali C, Ma T, Khedr H, Rädle R, Rolland C, Gustafson L, Mintun E. Sam 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714. 2024 Aug 1.

# Contributions

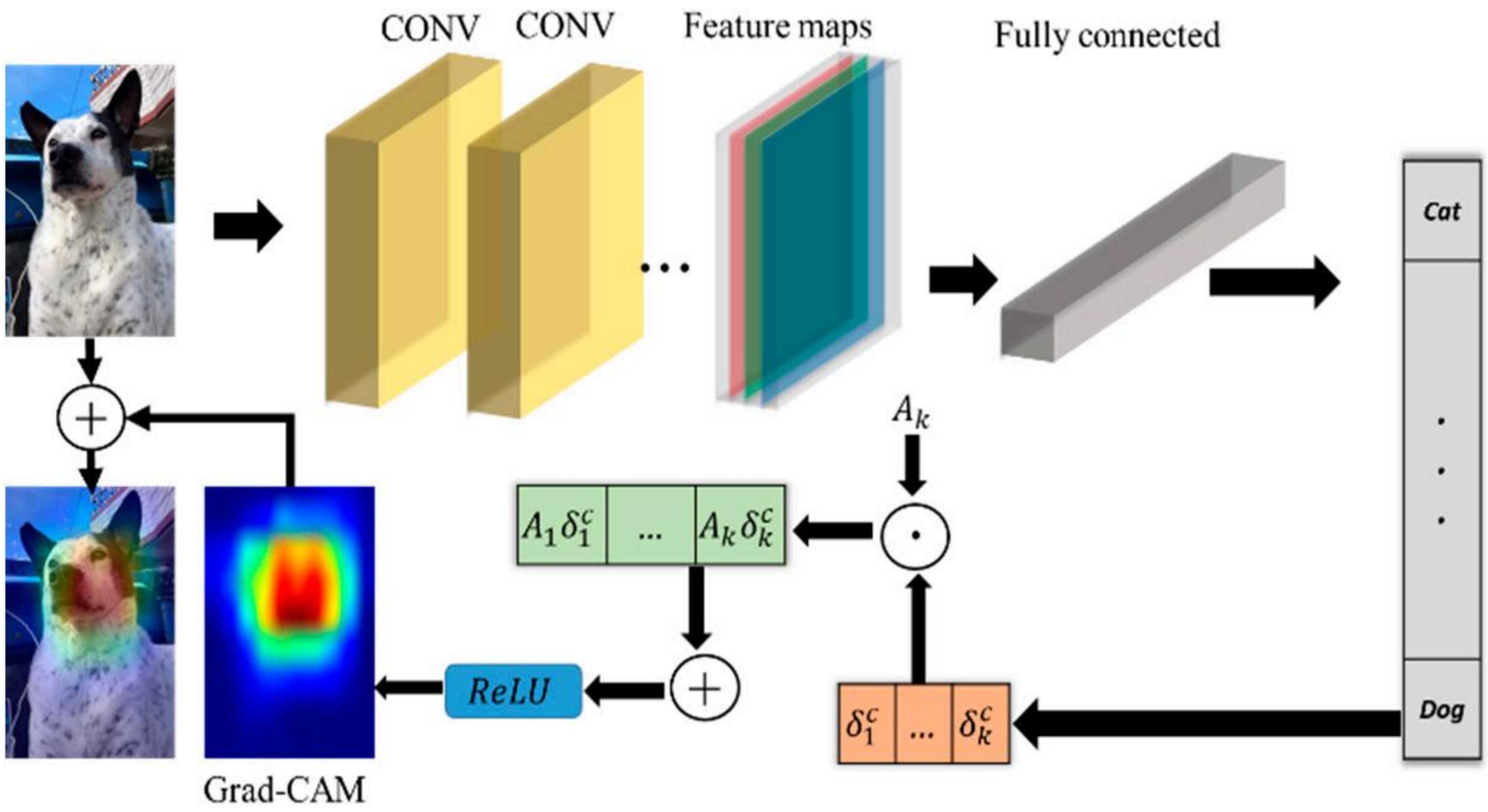
## Explaining image classifications with CRAFT




Fel, Thomas, et al. "Craft: Concept recursive activation factorization for explainability." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.

# Contributions

## Explaining image classifications with GradCAM

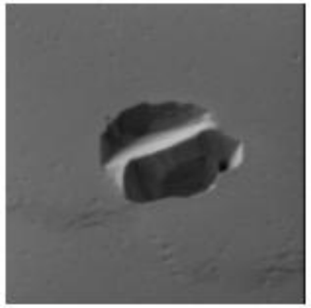


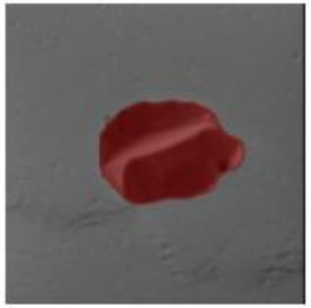

Selvaraju, Ramprasaath R., et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization." Proceedings of the IEEE international conference on computer vision. 2017.

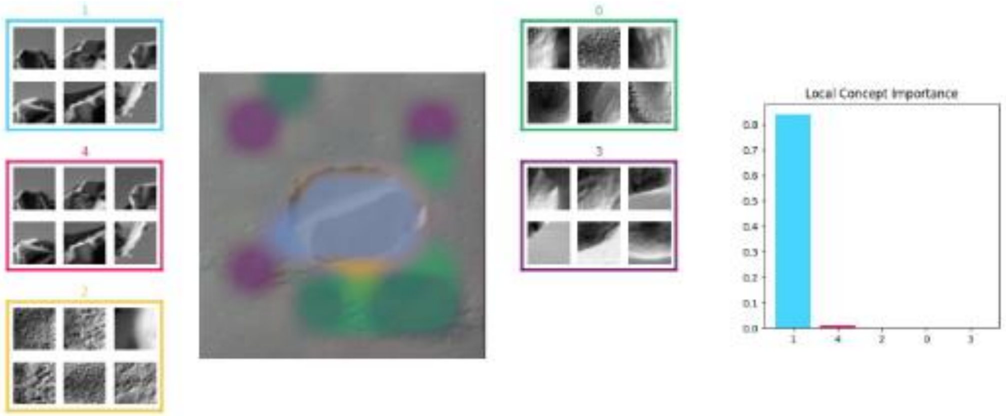
# Results

## Insights from qualitative evaluation of XAI methods

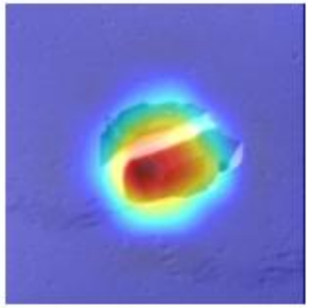
– GradCAM and CRAFT highlight critical regions relevant to the model's decision-making process
– CRAFT is resilient to imperceivable image changes
– CRAFT enhances interpretability by visualizing meaningful concepts that influence predictions


Original image


Ground truth



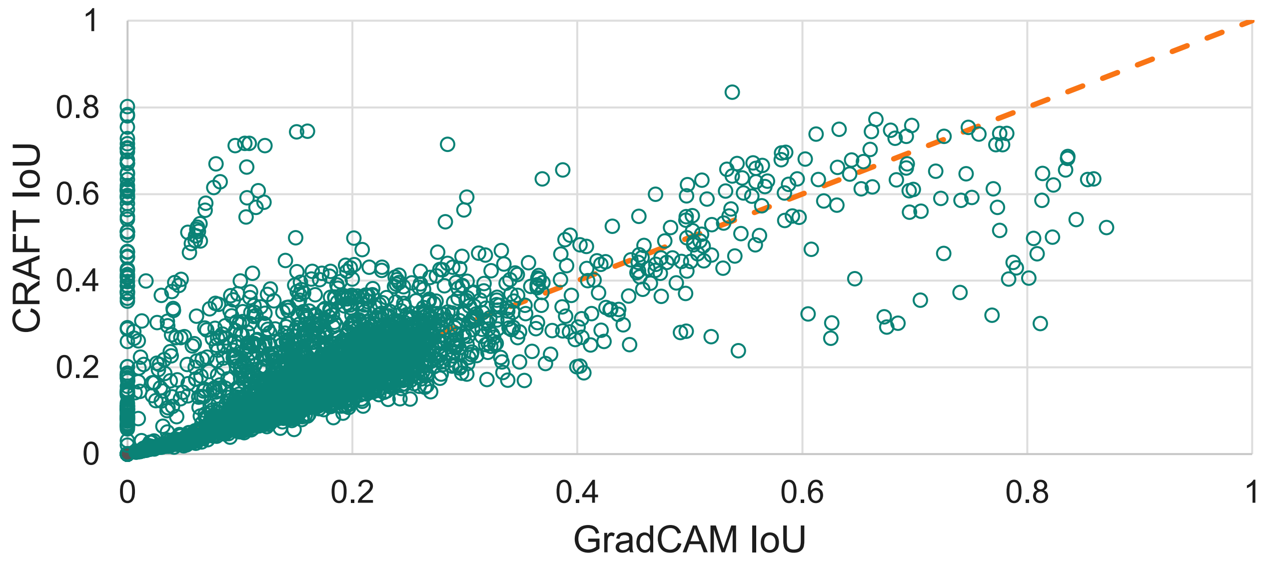
CRAFT


GradCAM

# Results

## Quantitative evaluation of XAI methods

| Method | Accuracy | Precision | Recall | Dice | IoU |
|---|---|---|---|---|---|
| CRAFT | 0.899 ± 0.028 | 0.179 ± 0.165 | 0.925 ± 0.232 | 0.269 ± 0.171 | 0.168 ± 0.132 |
| GradCAM | 0.907 ± 0.075 | 0.164 ± 0.138 | 0.851 ± 0.295 | 0.254 ± 0.154 | 0.156 ± 0.118 |




**Algorithm 2** Binary Mask Extraction from GradCAM

**Input:** GradCAM explanation $e_{gc} \in \mathbb{R}^{H \times W \times 3}$, Threshold $\tau \in [0, 1]$

**Output:** Binary mask $M_{gc} \in \{0, 1\}^{H \times W}$

$e \leftarrow e_{gc}[:, :, 0]$ {Red activation}

$e_{\text{norm}} \leftarrow \frac{e}{\max(e)}$

$M \leftarrow \mathbf{1}_{e_{\text{norm}} > \tau}$

**return** $M$

**Algorithm 3** Binary Masks Extraction from CRAFT

**Input:** CRAFT explanation $e_c \in \mathbb{R}^{H \times W \times 3}$, Concept colors $\mathcal{C} = \{c_1, c_2, \ldots, c_k\}, c_i \in \mathbb{R}^3$, Color tolerance $\delta \in \mathbb{R}^+$

**Output:** Binary concept masks $M = \{M_1, M_2, \ldots, M_k\}, M_i \in \{0, 1\}^{H \times W}$

$e \leftarrow Crop(e_c)$ {Crop to heatmap explanation}

**for** $i \in \{1, 2, \ldots, k\}$ **do**

  $D_i(x, y) \leftarrow ExtractConceptMasks(e_i, c_i)$ {Extract concept masks based on color distance}

  $M_i \leftarrow \mathbf{1}_{D_i \leq \delta}$

**end for**

**return** $M$

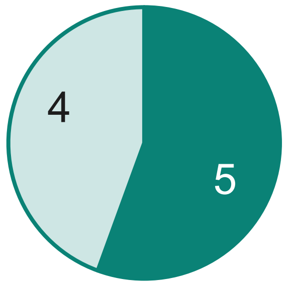
# Results

## Key findings from expert user study

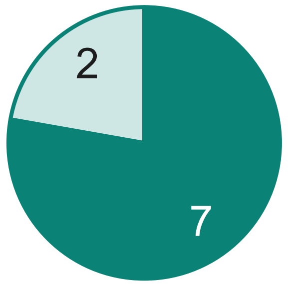
Participation of 9 experts

Evaluation using 5-point Likert explanation satisfaction

| Questions | |
|---|---|
| Q1 | From the explanation, I understand how the model works. |
| Q2 | This explanation of how the model works is satisfying. |
| Q3 | This explanation of how the model works has sufficient detail. |
| Q4 | This explanation of how the model works seems complete. |
| Q5 | This explanation of how the model works tells me how to use it. |
| Q6 | This explanation of how the model works is useful to my goals. |
| Q7 | This explanation of the model shows me how accurate the model is. |
| Q8 | This explanation lets me judge when I should trust and not trust the model |




5/9 experts found the explanations understandable (3.6/5) and satisfying (3.4/5)




7/9 experts found explanation helped judging when to trust or distrust the model's prediction (3.8/5)

R. R. Hoffman, S. T. Mueller, G. Klein, and J. Litman. Metrics for explainable ai: Challenges and prospects. arXiv preprintarXiv:1812.04608, 2018.

# Conclusion & outlook

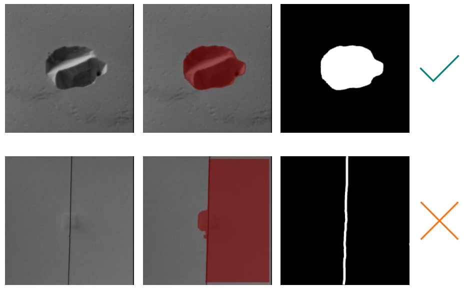
## Conclusion

- Proposal of algorithm for automatic generation of ground truth masks for SEM defect images
- Introduction of Carinthia-S dataset with expert validated ground truth segmentation masks
- Novel application of CRAFT to enhance explainability in SEM defect image classification

## Outlook

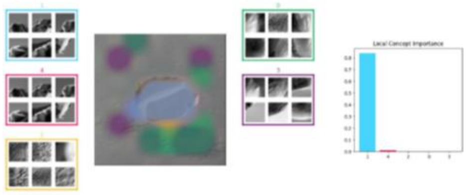
- Application of methods on an internal more complex dataset
- Enhancing robustness of automatic ground truth segmentation algorithm
- Dynamic hyperparameter tuning for CRAFT
- Improving CRAFT heatmap generation

# Your input matters!



## Automatic ground truth mask generation

- Who has experience with automatic ground truth mask generation?
- Are there recommendations for segmentation models with proven effectiveness?



CRAFT

## XAI methods

- Who has experience with XAI methods for image classification?
- Are there recommendations for XAI methods with proven effectiveness?

## Computer vision expertise

- Who among you is currently working on computer vision topics or knows someone who is?
- Collaboration and sharing insights on computer vision topics are always of great interest to us!

# Thank you for your attention!

IPCEI Microelectronics and Communication Technologies

This work is funded by the Austrian Federal Ministry of Climate Action, Environment, Energy, Mobility, Innovation and Technology, the Austrian Federal Ministry of Digital and Economic Affairs, and implemented by Austria Wirtschaftsservice (AWS) and the Austrian Research Promotion Agency (FFG) in the frame of the Important Project of Common European Interest (IPCEI) on Microelectronics and Communication Technologies (ME/CT).

Federal Ministry
Republic of Austria
Education, Science
and Research

The AIMS5.0 project is supported by the Chips Joint Undertaking and its members, including the top-up funding by National Funding Authorities from involved countries (FFG project no. FO999899258) under grant agreement no. 101112089. Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Chips Joint Undertaking. Neither the European Union nor the Chips Joint Undertaking can be held responsible for them.