

VERONA, 15-19 JUNE

 **ECNDT** ^{14th}
2026
EUROPEAN CONFERENCE ON NON-DESTRUCTIVE TESTING

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Machine learning-driven flaw detection for ultrasonic pipe inspections with acoustic lens

Thiago E. Kalid

Universidade Tecnológica Federal do Paraná, Curitiba, Brazil

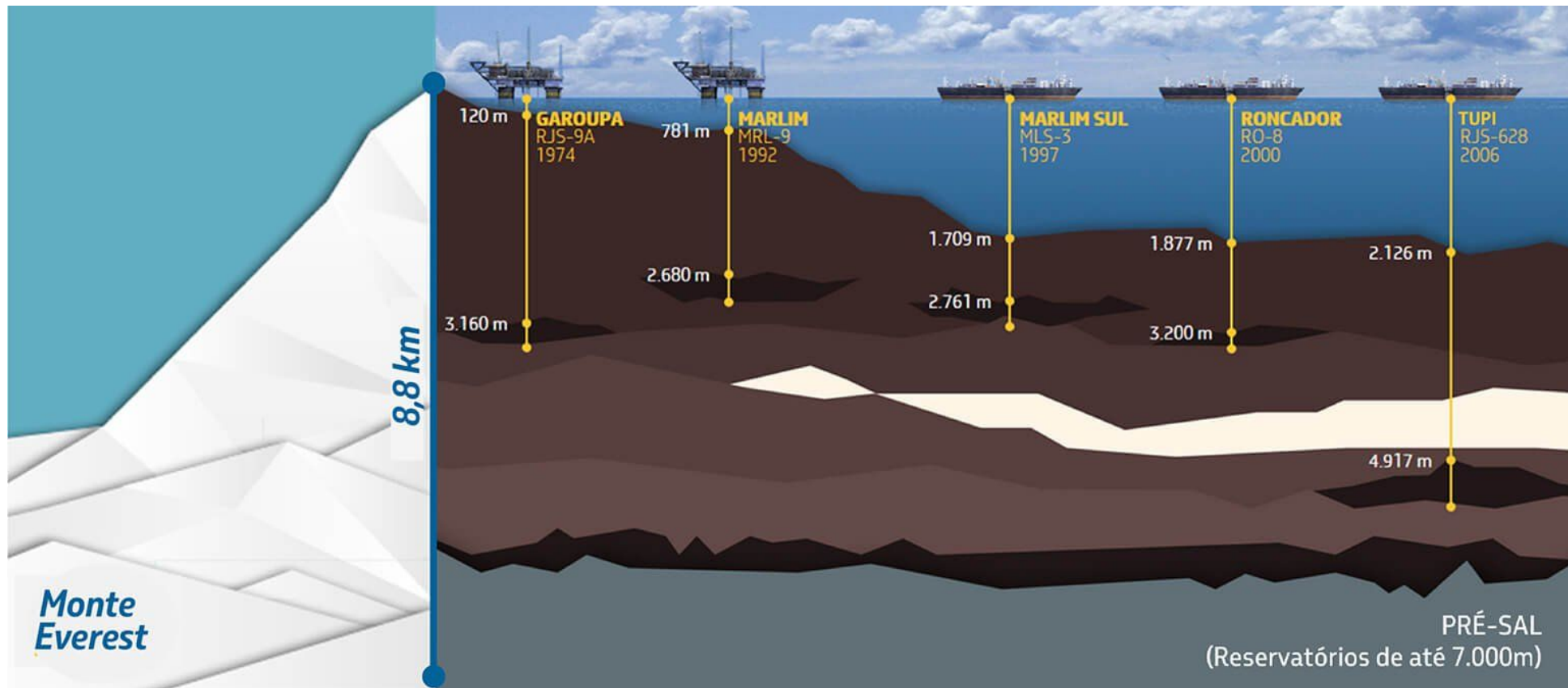
The oil and gas sector has many critical components in its infrastructure that are likely to be subject to significant wear and tear.



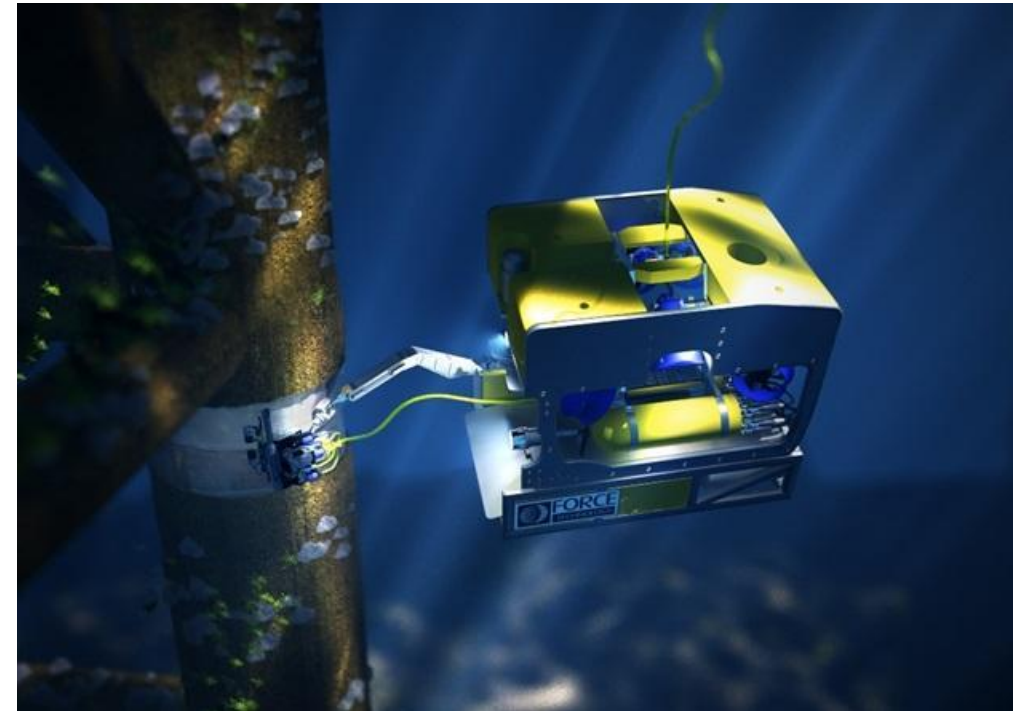
Non-Destructive Testing (NDT) is a solution for monitoring the structural health of these components.



The extraction of oil in the Brazilian pre-salt imposes significant technological and logistic challenges for NDT

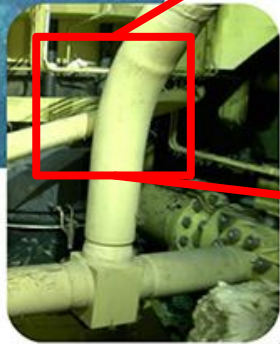


Inspection of deepsea components requires a huge amount of resources.

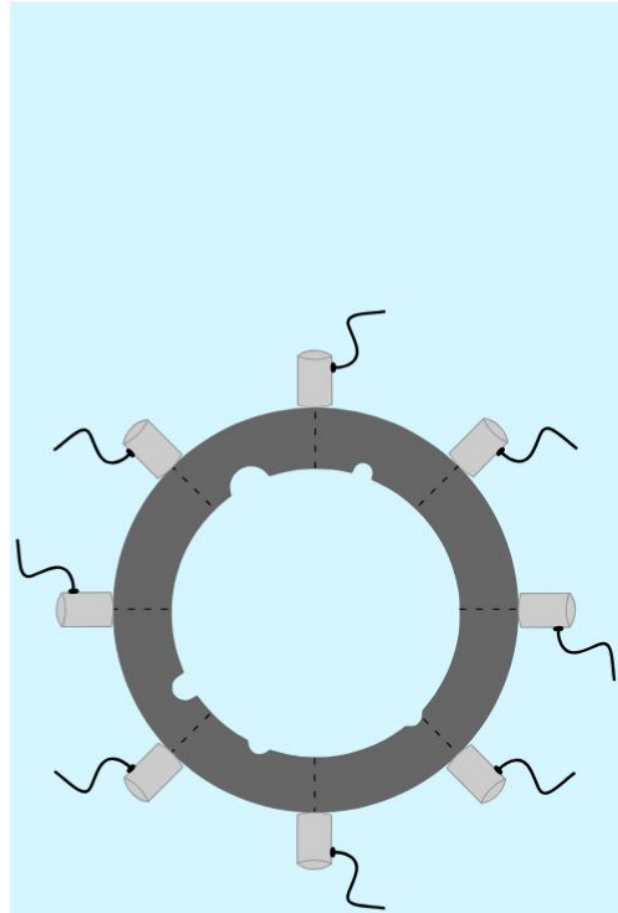


Reinjection water has a high CO₂ content, causing a non-uniform corrosion pattern.

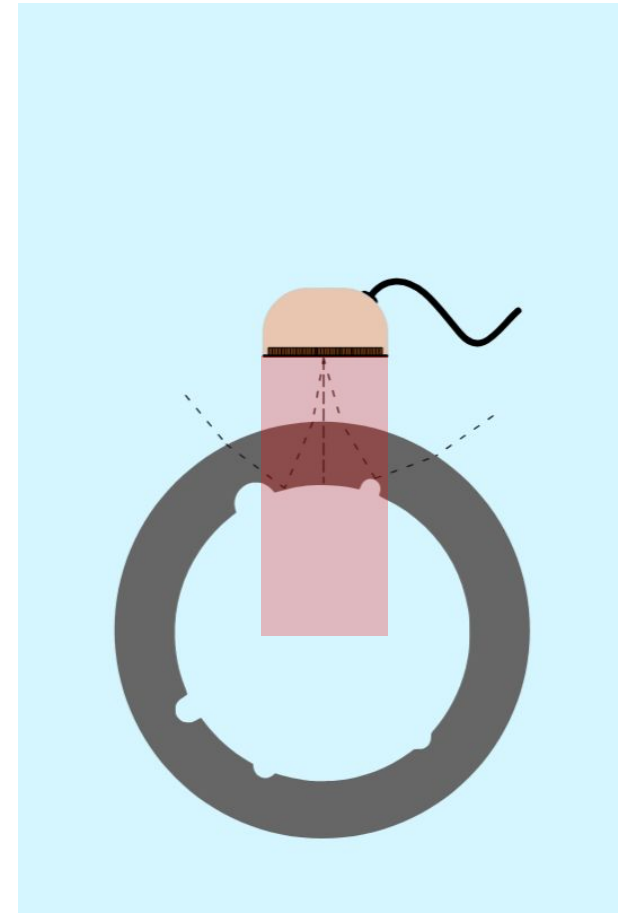
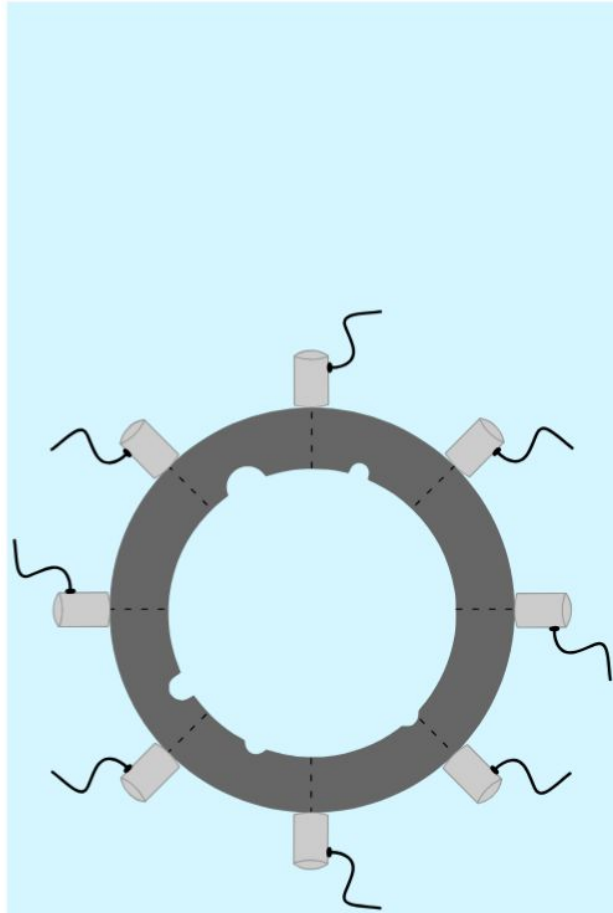
Piping of subsea equipment subject to pitting corrosion



One way to increase the probability detection is to assess wall thickness at multiple positions.

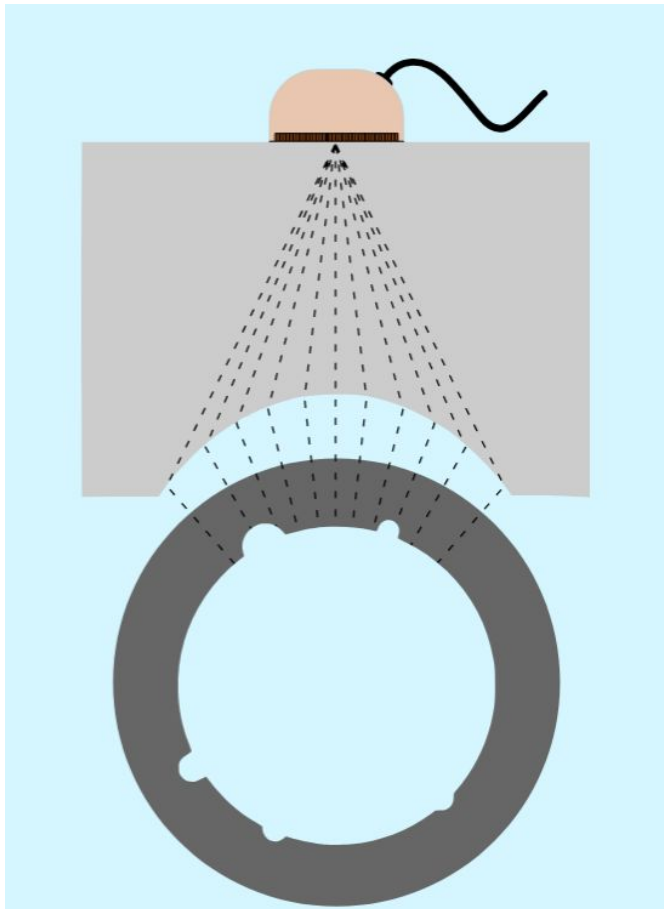


Phased array probes increases the inspected area, but still is roughly restricted to the probe active aperture.



$\sim 26^\circ$ sector

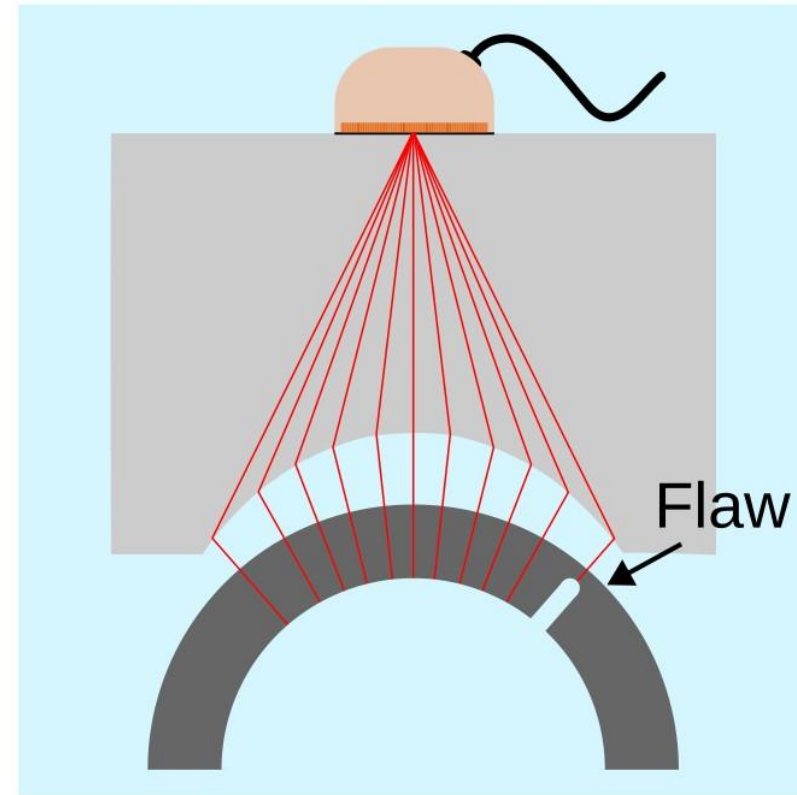
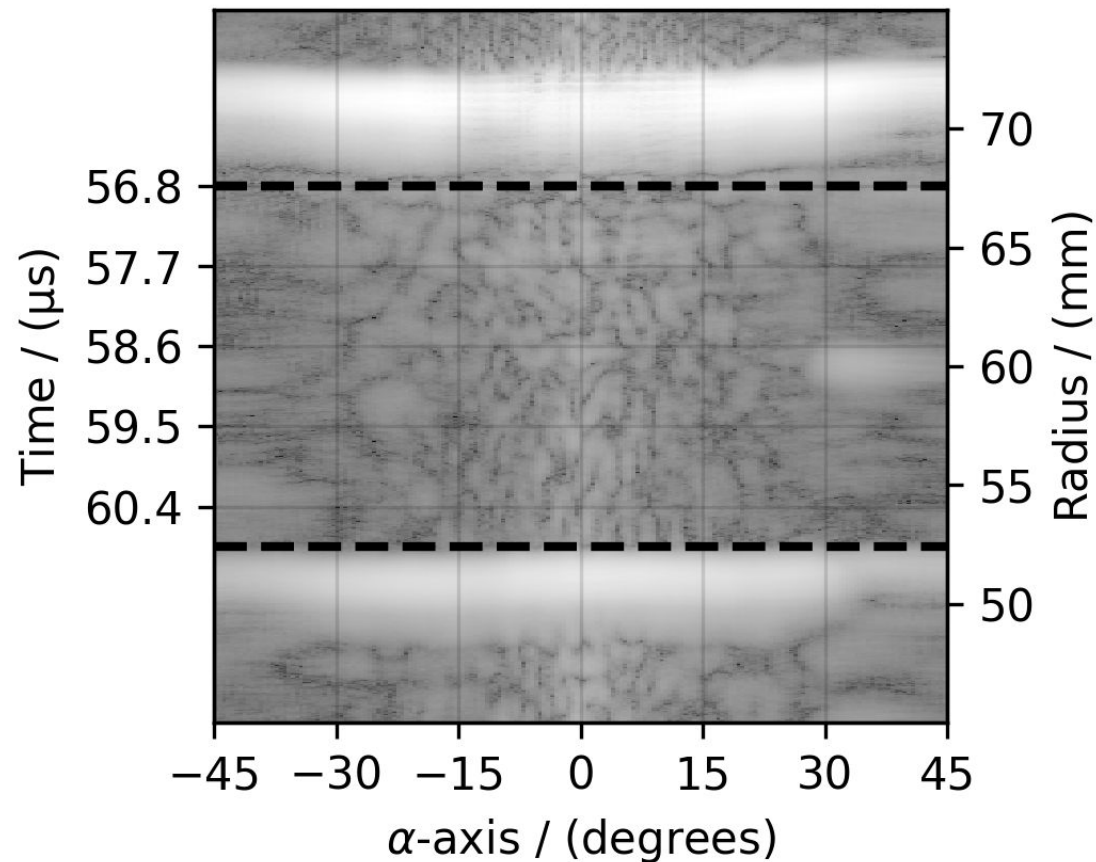
To reduce the time required to inspect a pipe, we previously proposed the use of an **acoustic lens**.



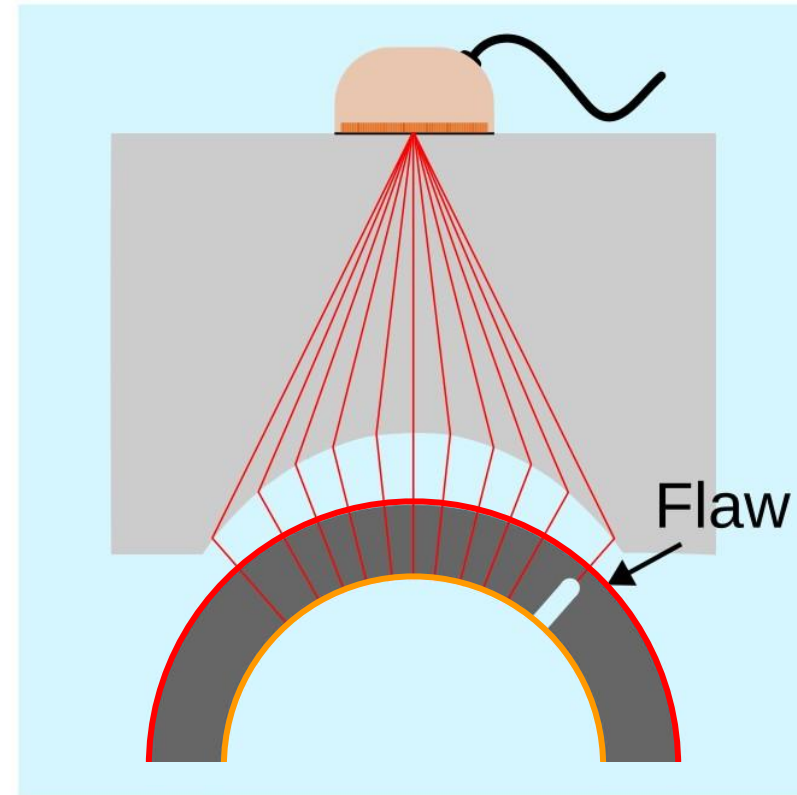
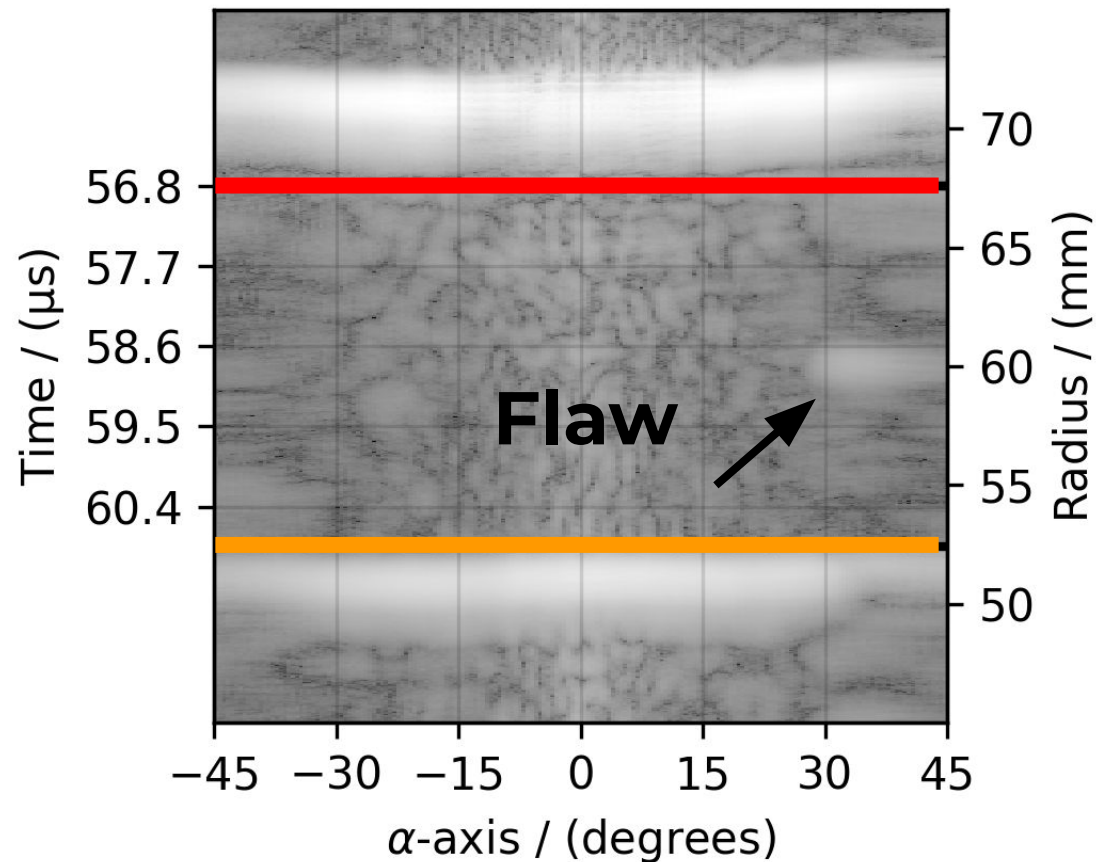
~90° sector



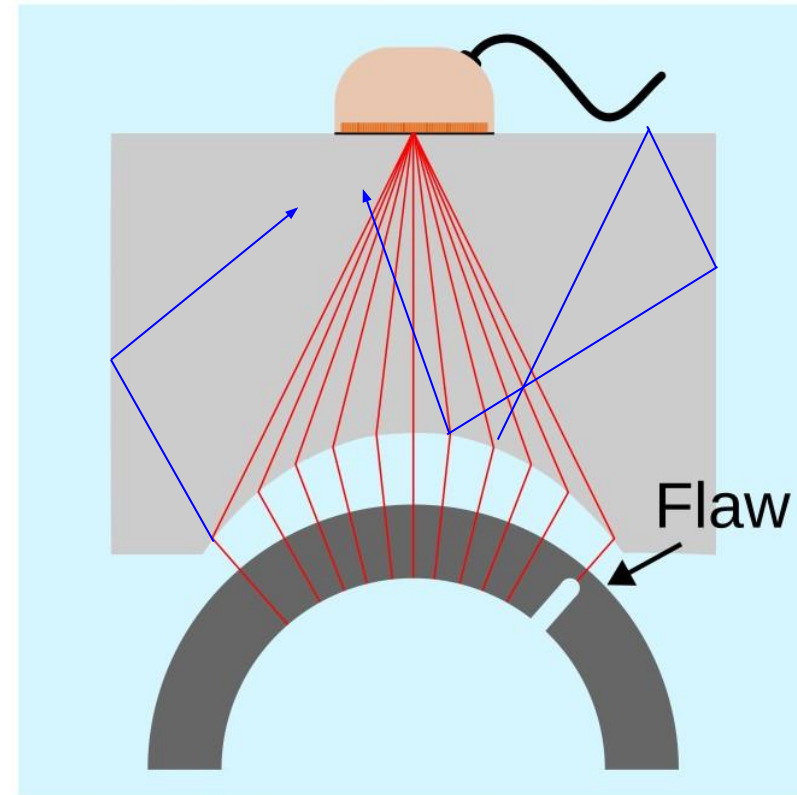
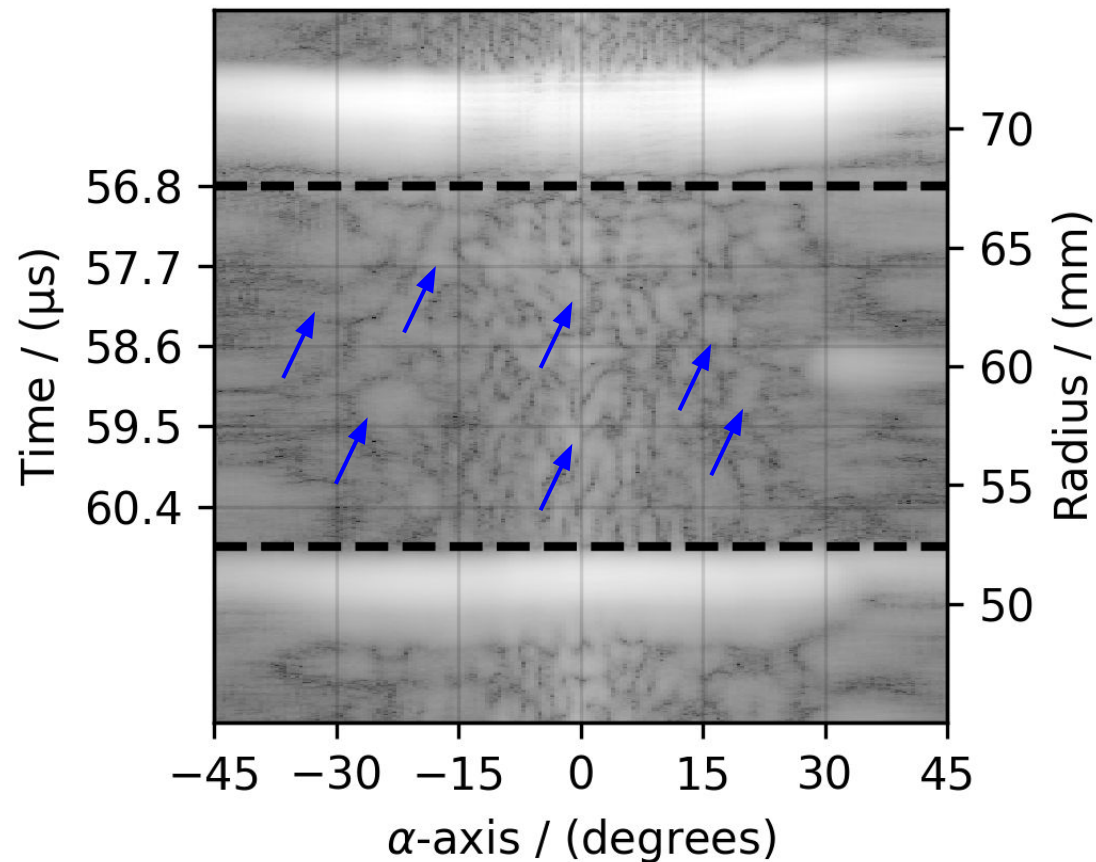
The inner and outer surface of a cylindrical specimen (pipe) appears as parallel horizontal lines.



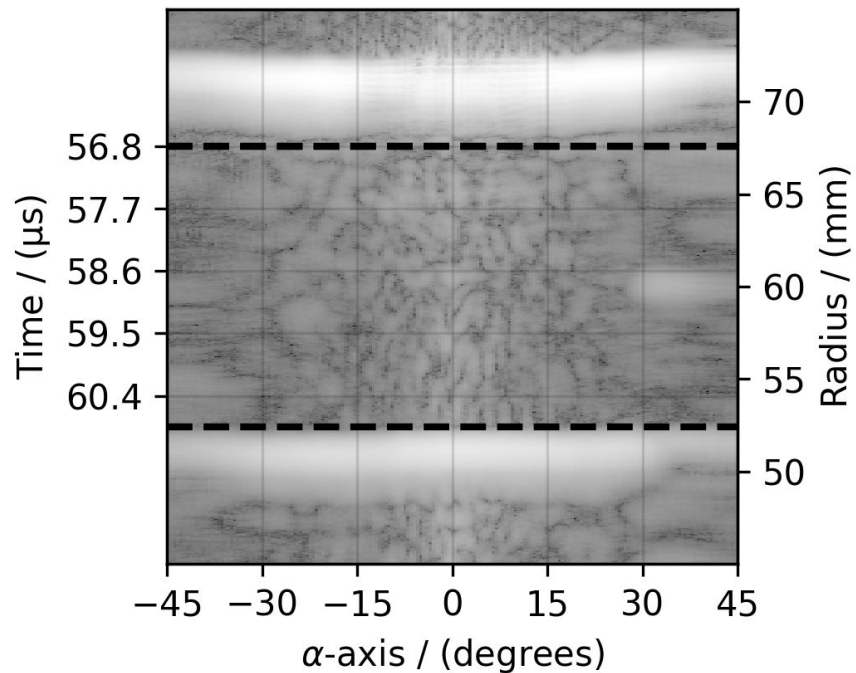
The inner and outer surface of a cylindrical specimen (pipe) appears as parallel horizontal lines.



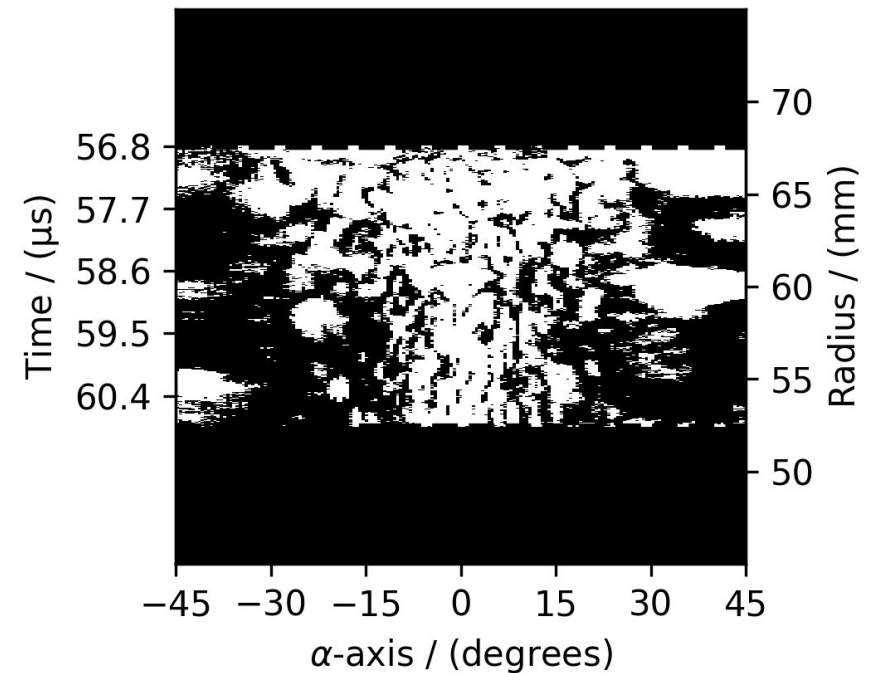
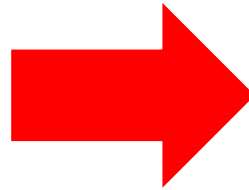
Although the acoustic lens allows us to see a larger section of the pipe, it introduces geometrical artifacts.



Classic methods, based on thresholding, perform poorly on bad SNR.



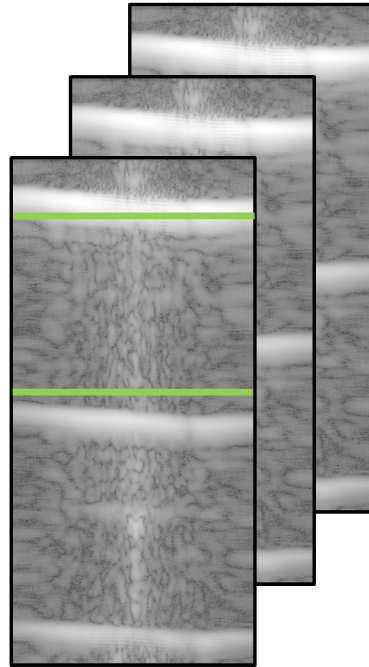
-6 dB
threshold



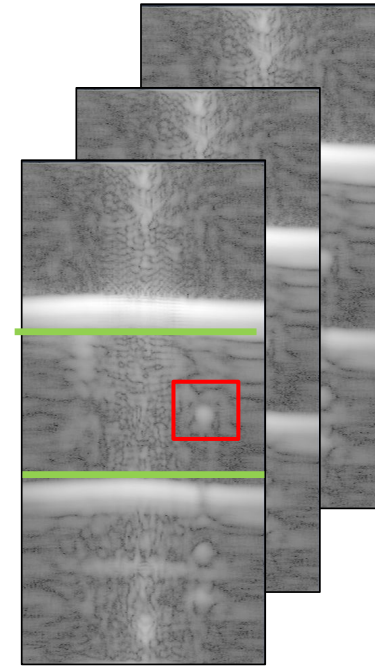
Proposed solution

Our ideal flaw detection model:

Trained on few flawless pipes
(in lab)



Tested on pipes with
different types of flaws.



We had two major constraints to our original model:

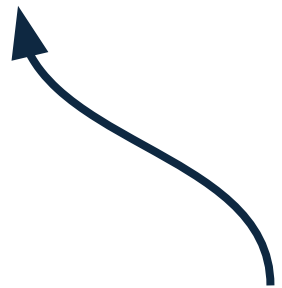
- **Our dataset was relatively small**

We had two major constraints to our original model:

- Our dataset was relatively small.
- **We wanted to create a model that didn't have to be trained on specific flaw geometries since pitting corrosion is unique by itself.**

We had two major constraints to our original model:

- Our dataset was relatively small (data augmentation and simulation were not an option).
- We wanted to create a model that didn't had to be trained on specific flaw geometries since pitting corrosion is unique by itself.



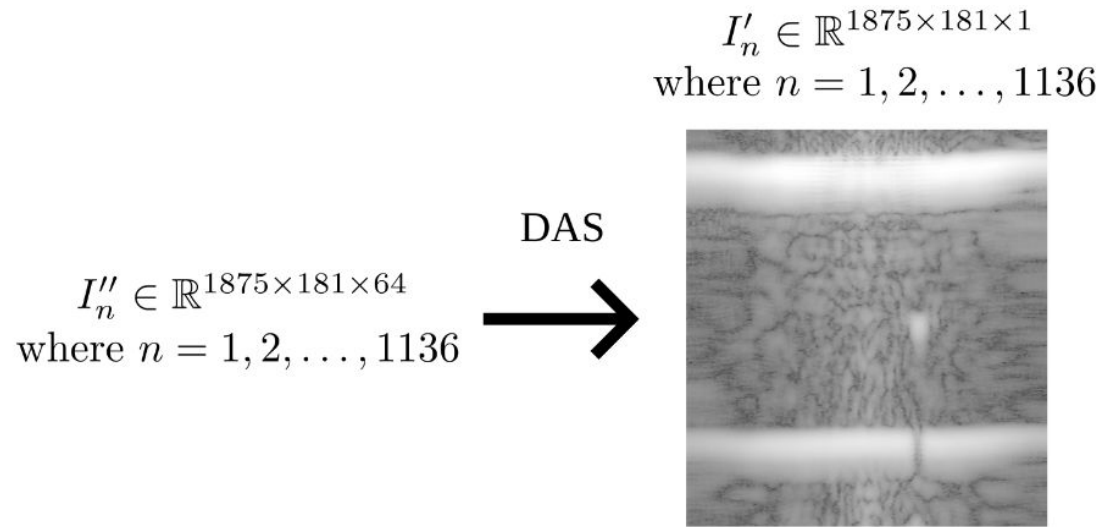
To solve these two constraints, we adopted a model based on machine learning **anomaly detection algorithms** to detect flaws.

Our original problem, based on ultrasound RF data, has a huge dimension, we must reduce it.

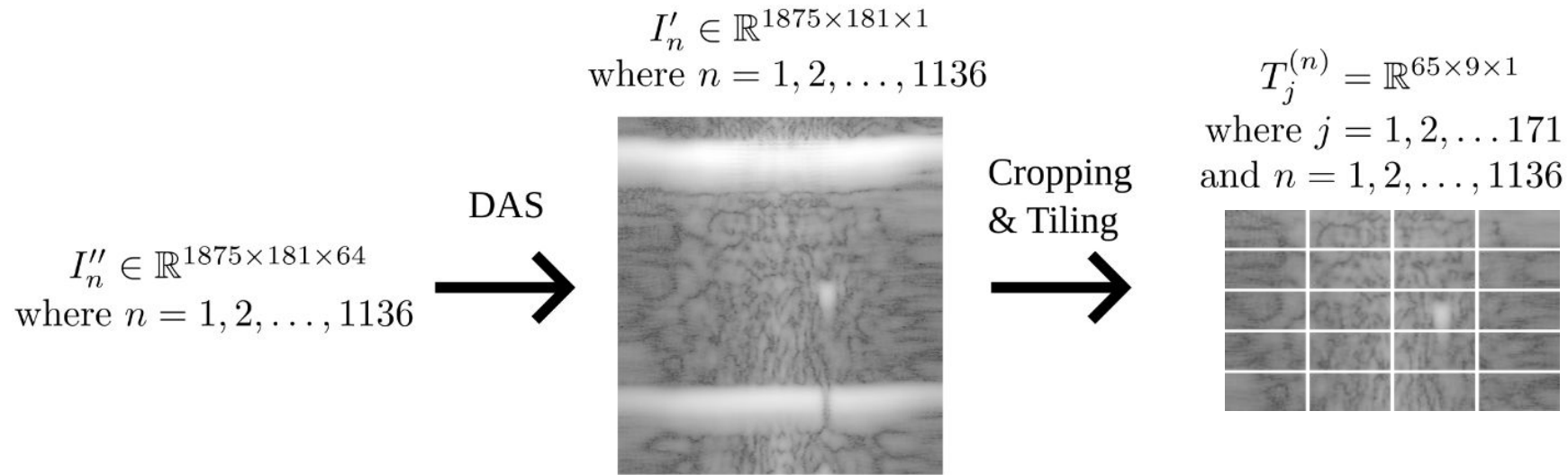
$$I''_n \in \mathbb{R}^{1875 \times 181 \times 64}$$

where $n = 1, 2, \dots, 1136$

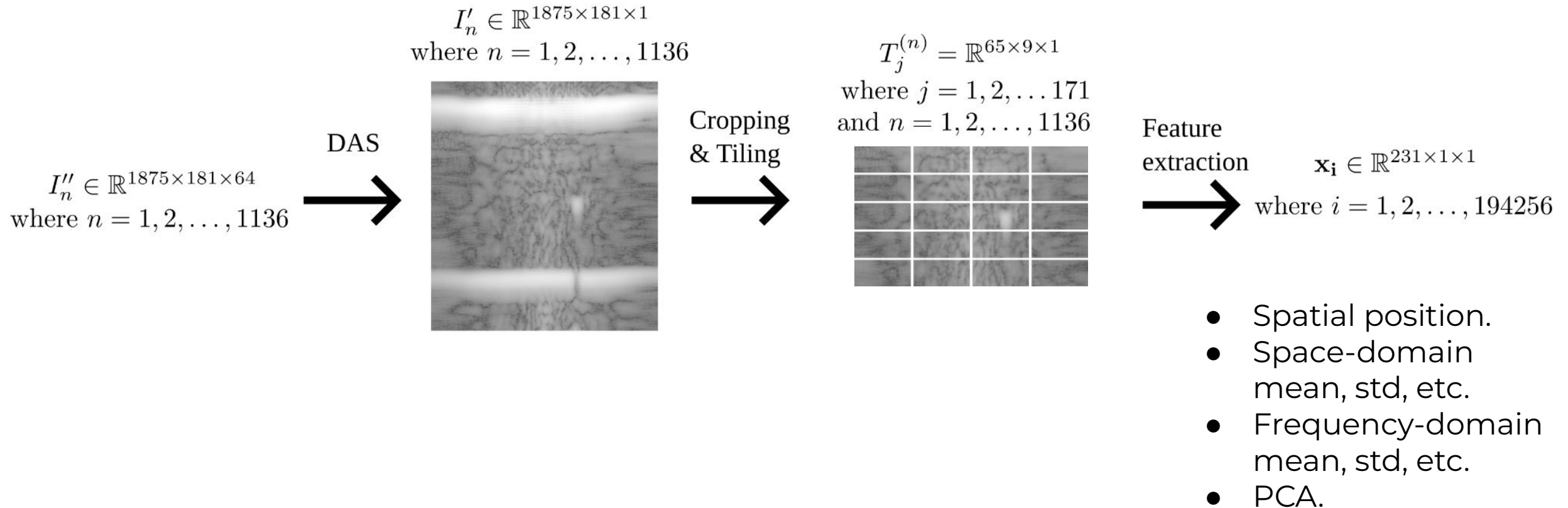
Our original problem, based on ultrasound RF data, has a huge data dimension, we must reduce it.



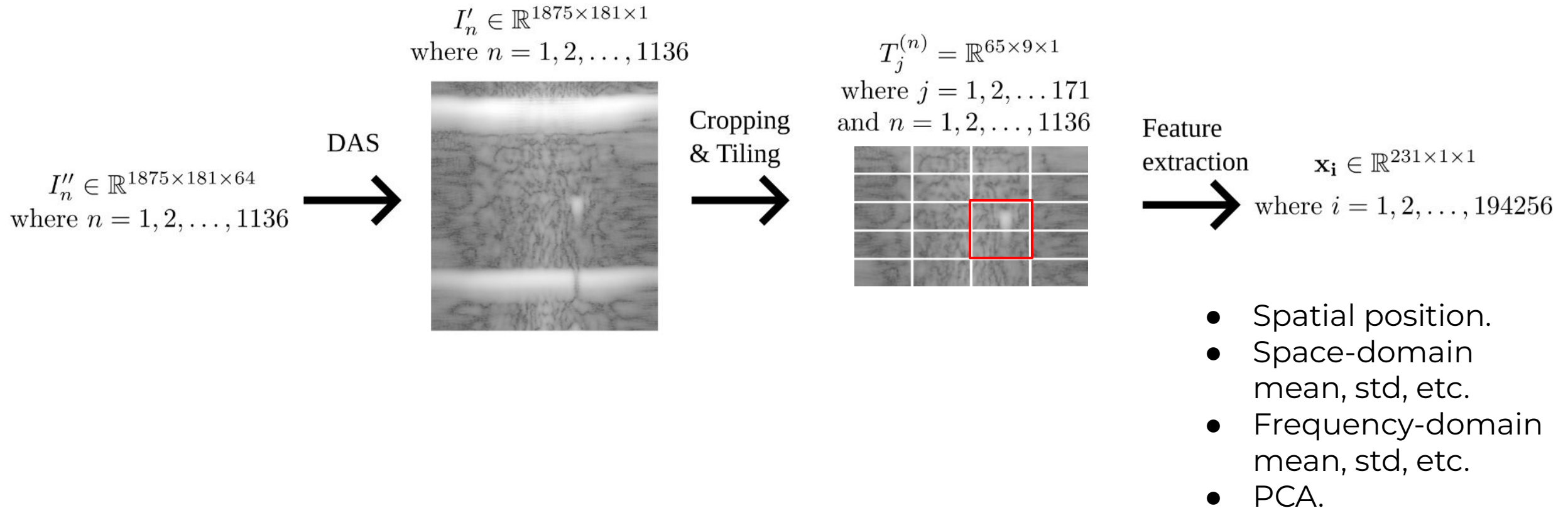
After that, the sectorial image was cropped into a smaller region of interest, and tiled.



After that, the sectorial image was cropped into a smaller region of interest, and tiled.

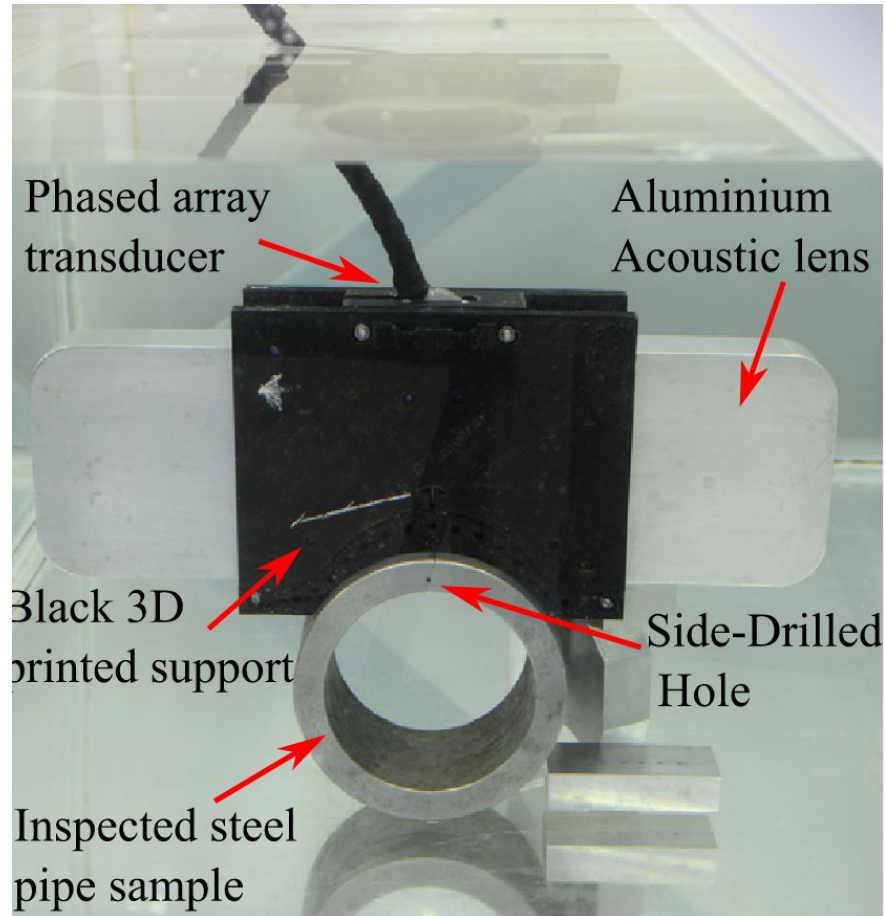


Based on the extract features from each tile we could spot which one would have anomalies.



Dataset

We inspected 4 cylindrical specimens with different flaws geometries.



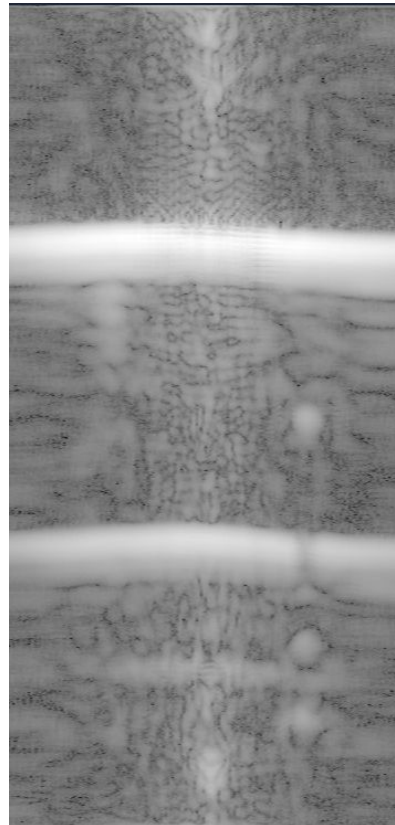
Each inspection yielded more than one acquisition, especially when there was a mechanical sweep along the pipe.

~40 inspections
from 4 pipes

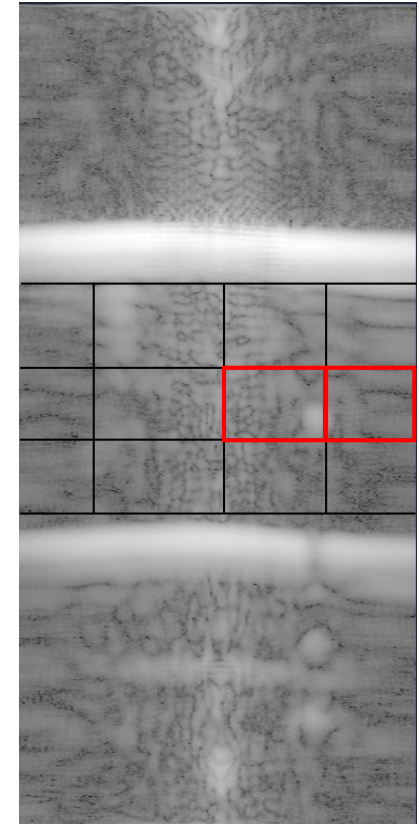
Raw
ultrasound RF
data



1100 S-scans

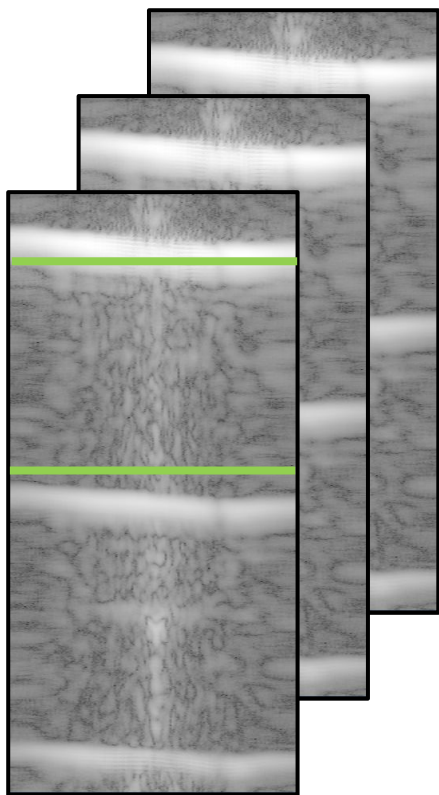


Binary tile label

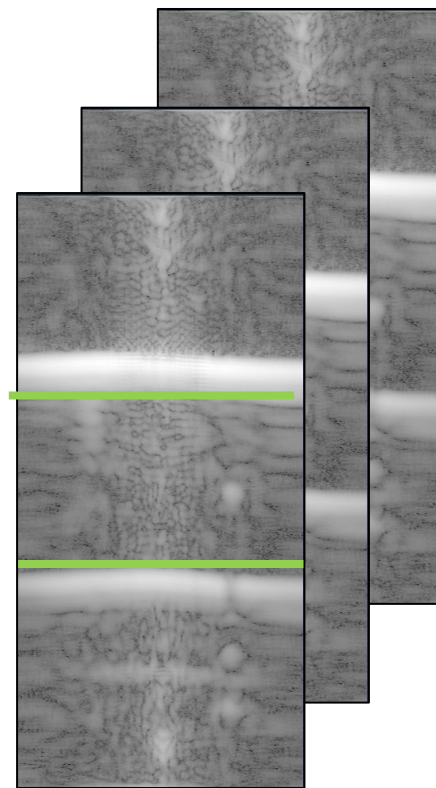


It will be a 70/30 **semi-supervised** novelty detection problem:

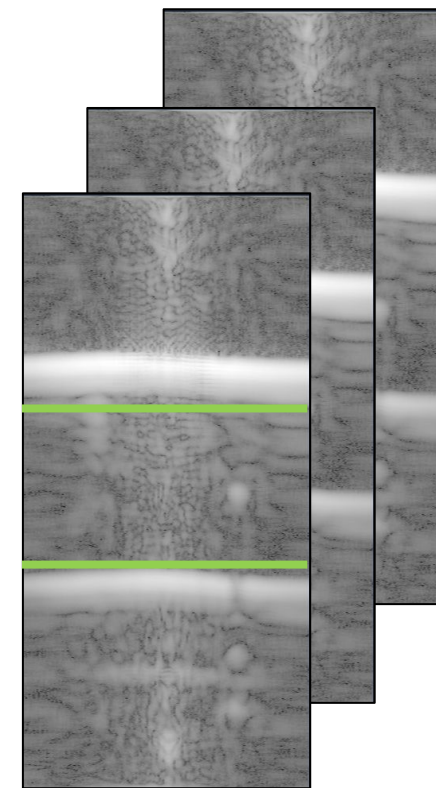
70 % train:
Tiles **without** flaws



15 % validation:
Tiles **with** and **without** flaws**



15 % test:
Tiles **with** and **without** flaws



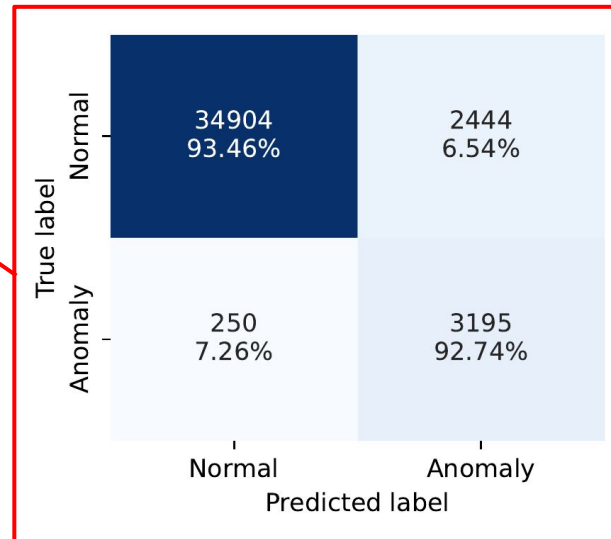
Results

I tested a set of classical machine learning anomaly detection algorithms by using the PyOD python library:

- Local Outlier Factor (lof)
- K-nearest neighbours (knn)
- Isolation forest (iforest)
- Local-threshold (classical approach)

Quantitative results for different models:

Model	Accuracy	Precision	Recall	F_2 -Score	ROC AUC
Local Outlier Factor	0.934	0.567	0.927	0.823	0.960
K-Nearest Neighbors	0.918	0.510	0.745	0.682	0.935
Isolation Forest	0.887	0.343	0.369	0.363	0.768
Threshold	0.913	0.481	0.364	0.382	0.664



Local Outlier Factor (lof) confusion matrix:

Conservative approach
(maximizing **F2**-score):

True label	Normal	Anomaly
	Predicted label	Predicted label
Normal	34904 93.46%	2444 6.54%
Anomaly	250 7.26%	3195 92.74%

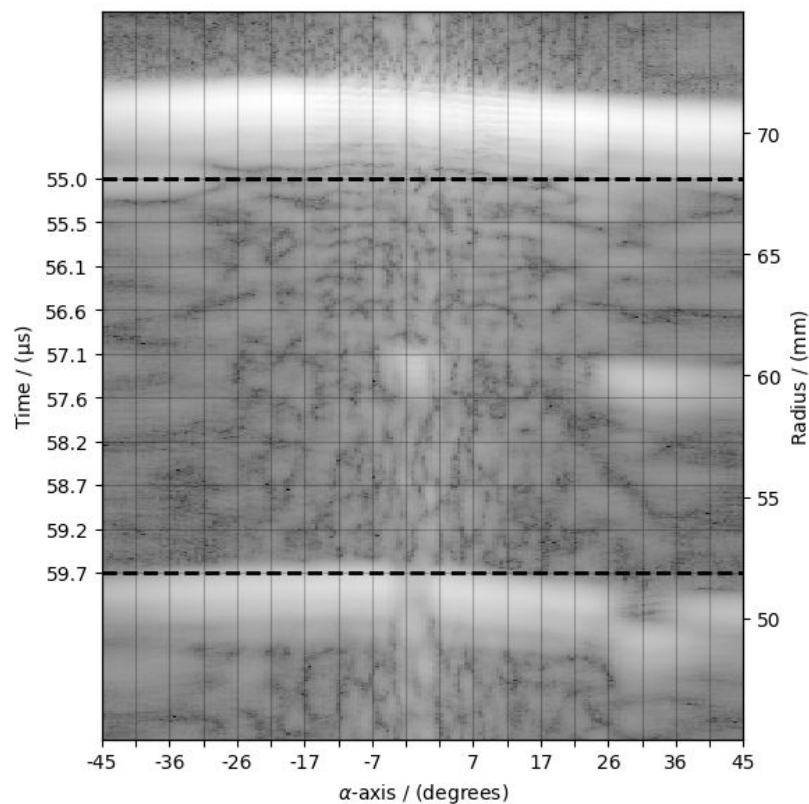
Balanced approach
(maximizing **F1**-score):

True label	Normal	Anomaly
	Predicted label	Predicted label
Normal	36122 96.72%	1226 3.28%
Anomaly	443 12.86%	3002 87.14%

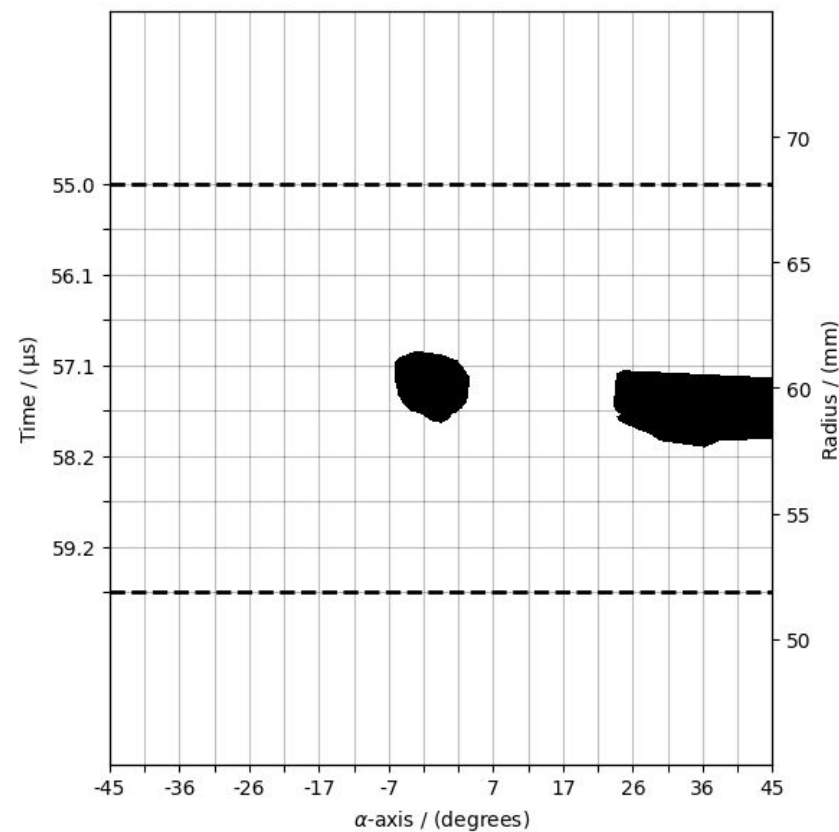
Qualitative result: testing the proposed method (**easy** scenario)



Acquired S-scan



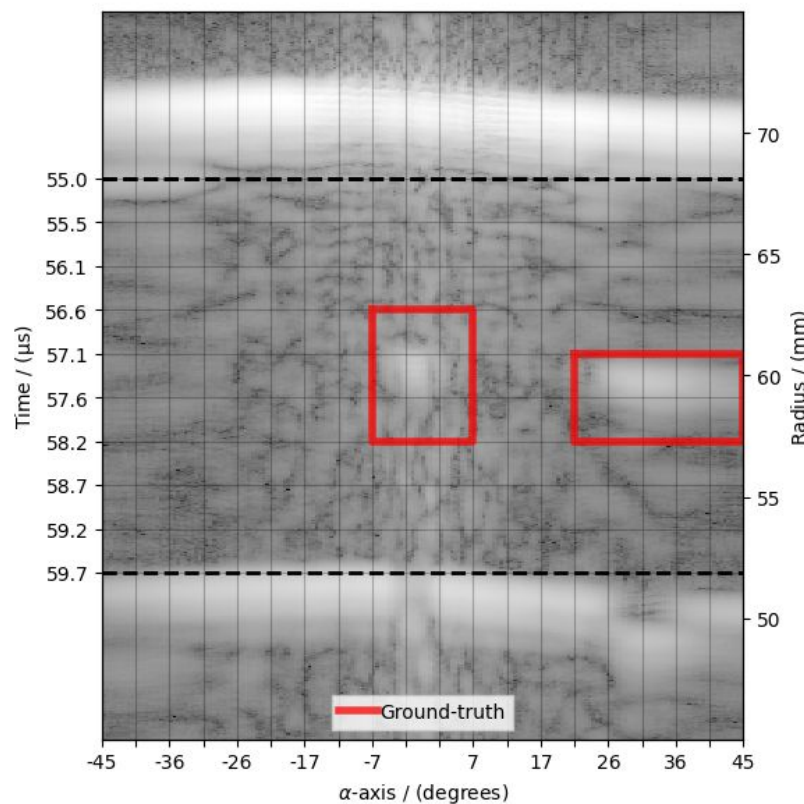
Ground-truth



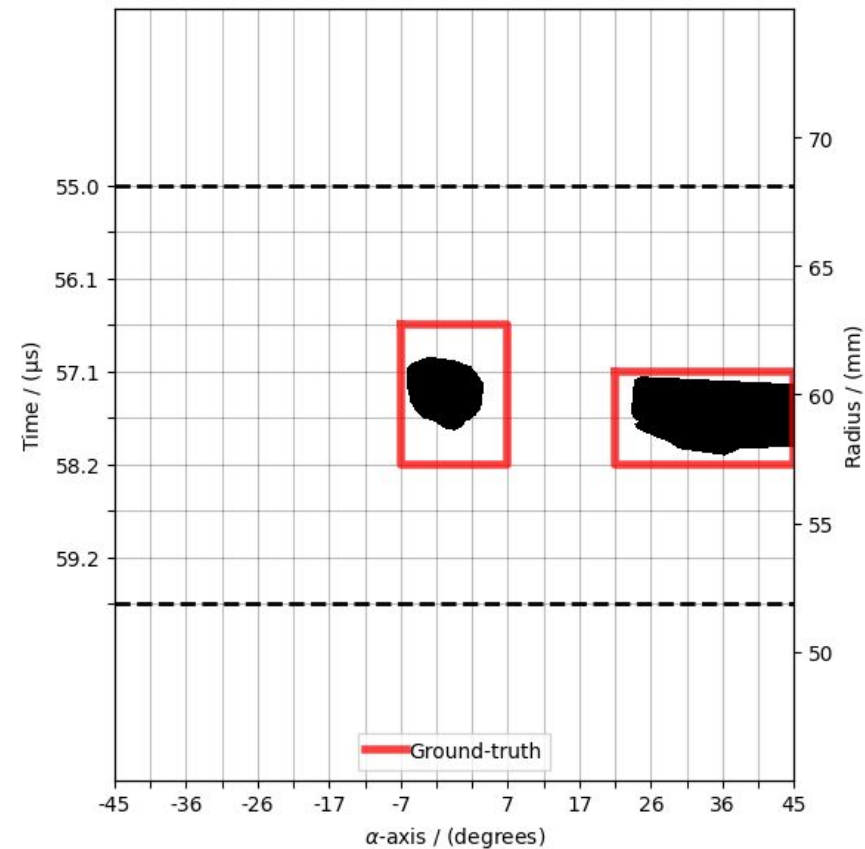
Qualitative result: testing the proposed method (**easy** scenario)



Acquired S-scan



Ground-truth

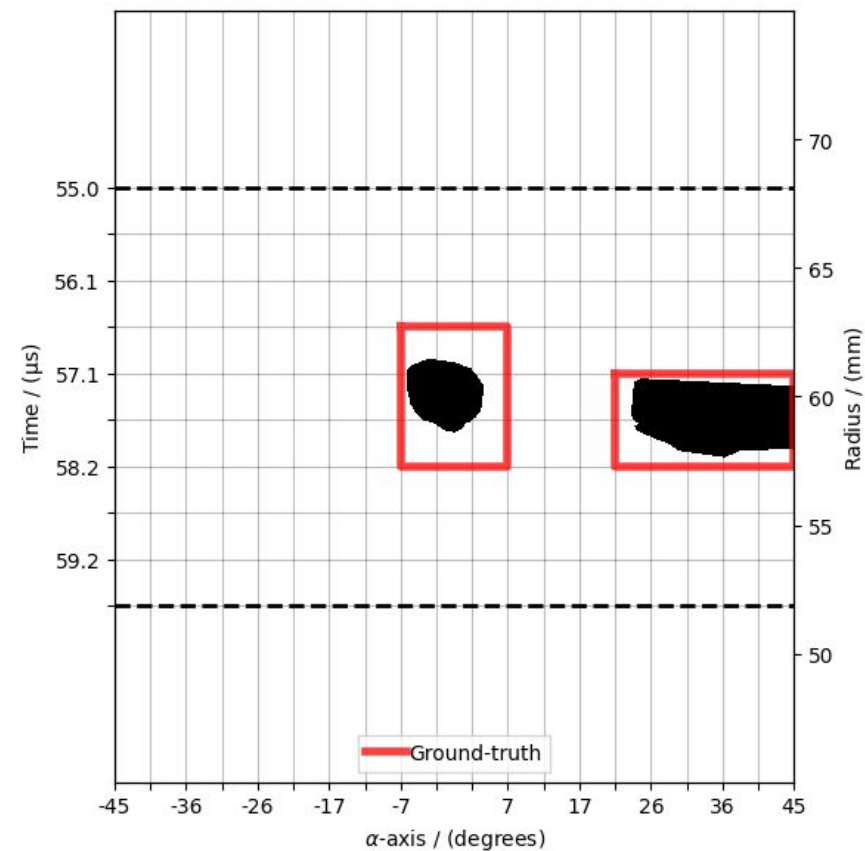


Qualitative result: testing the proposed method (**easy** scenario)

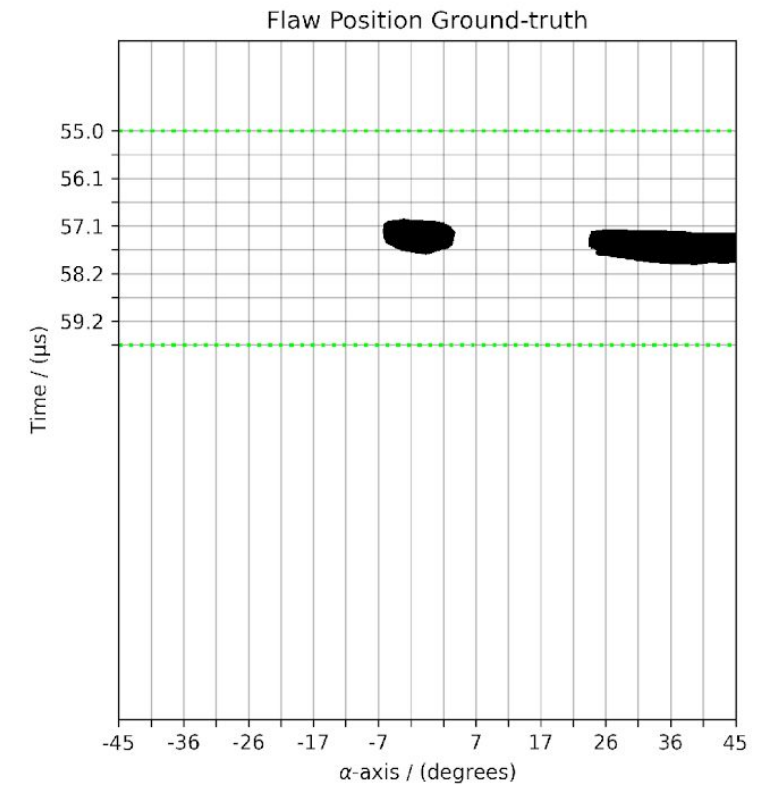
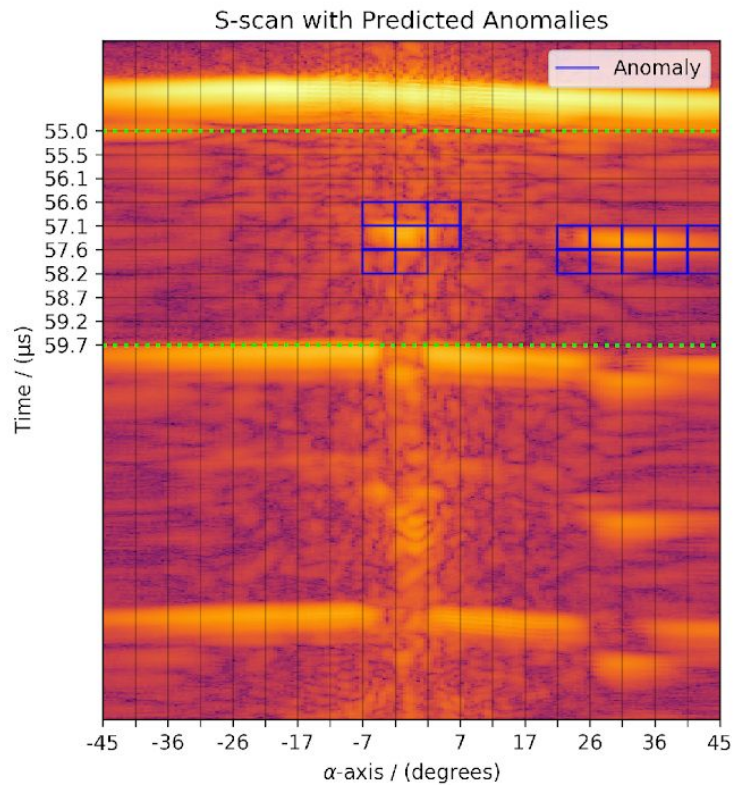
Acquired S-scan



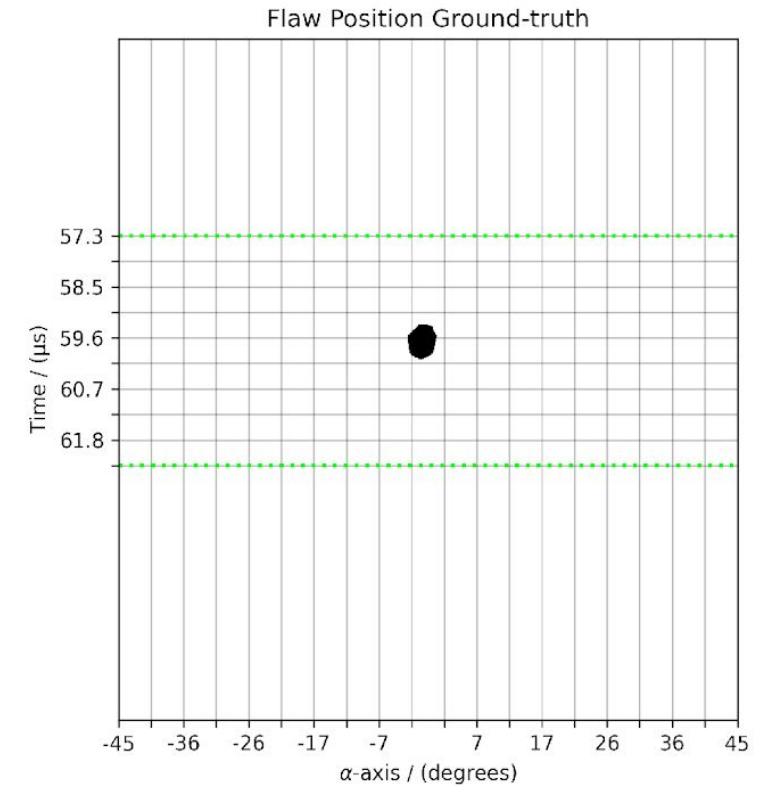
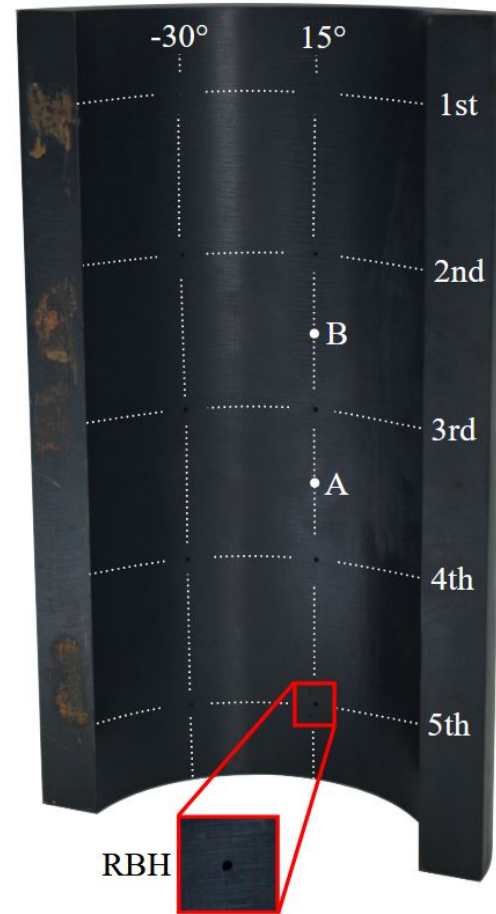
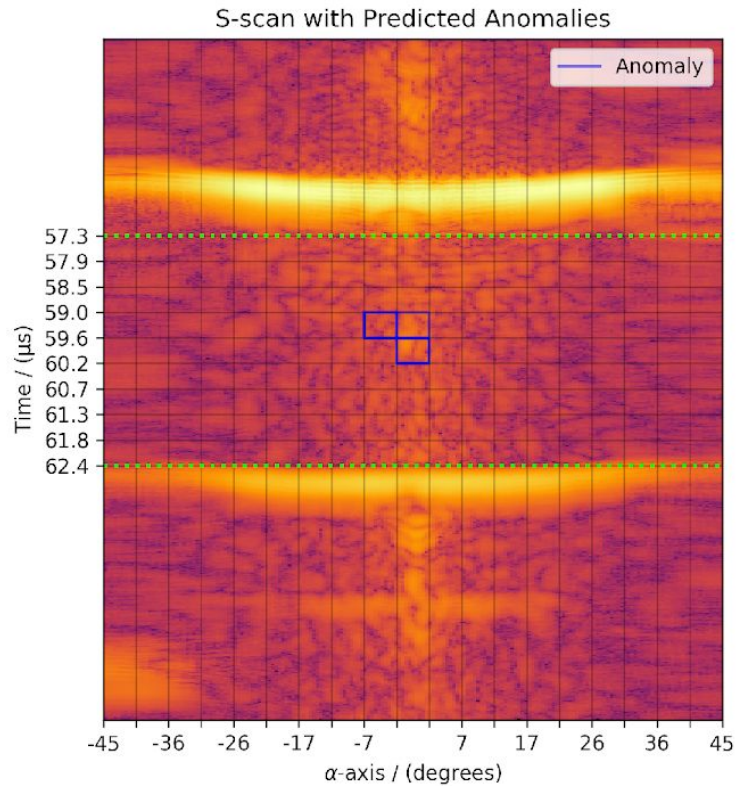
Ground-truth



Qualitative result: testing the proposed method (**easy** scenario)

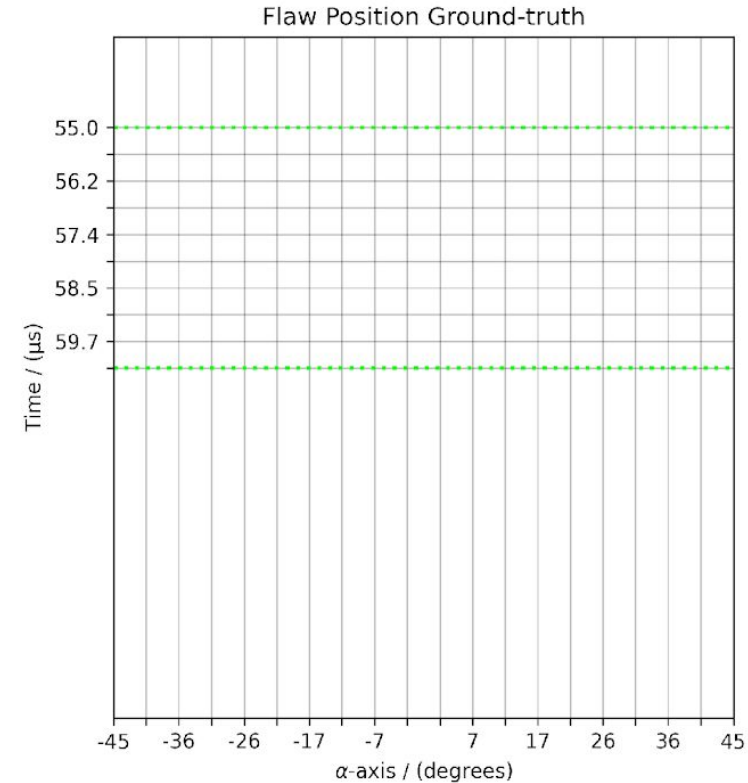
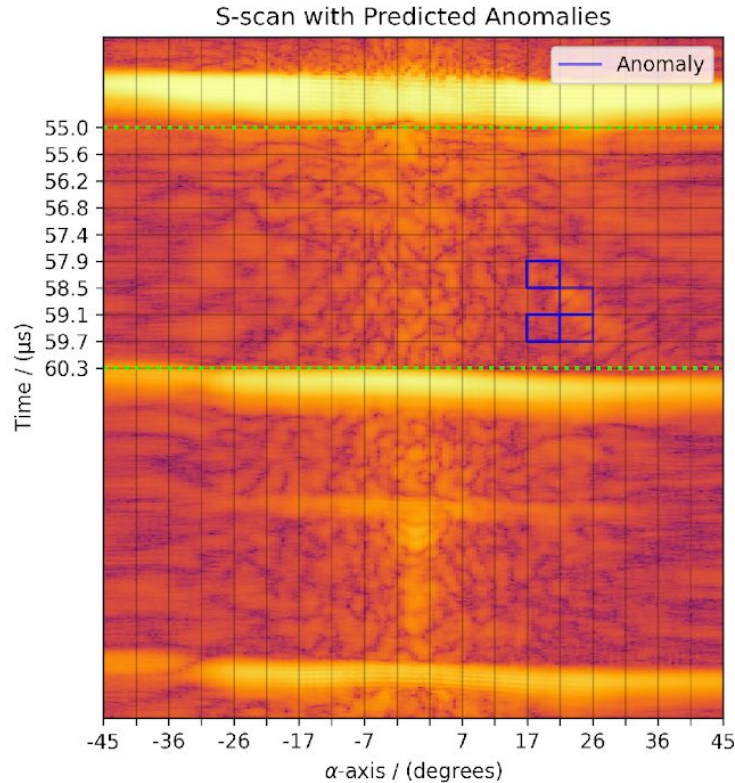


Qualitative result: testing the proposed method (**hard** scenario)

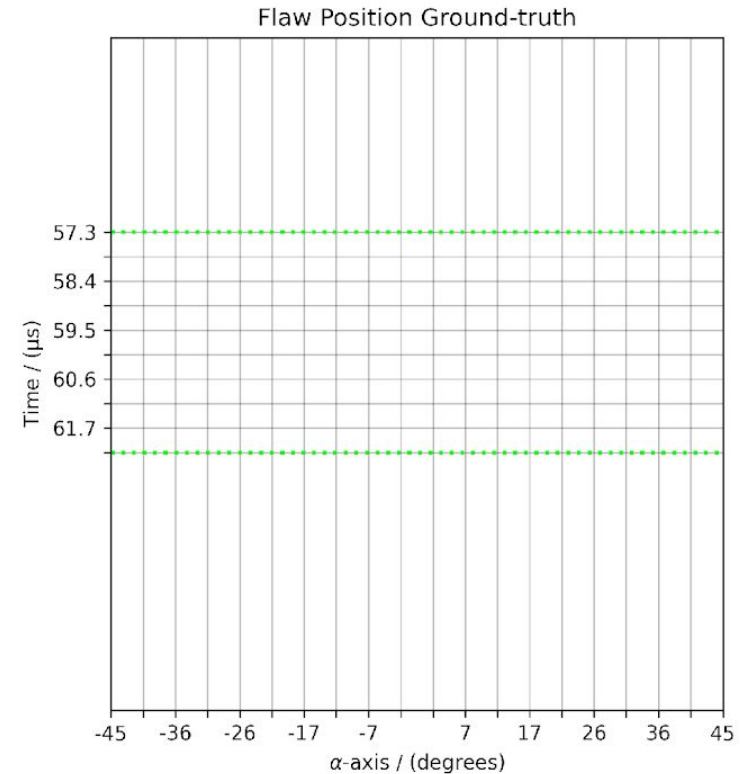
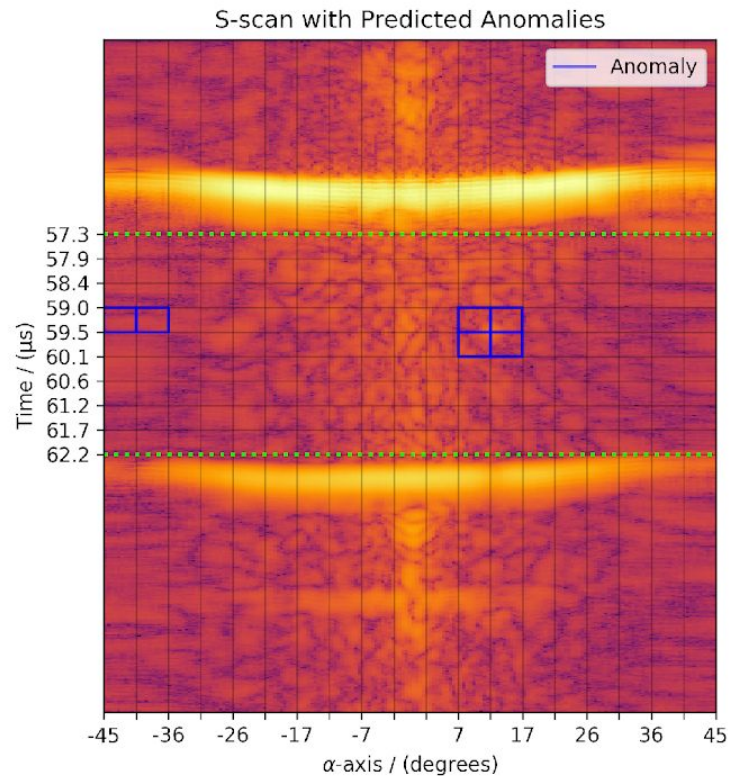


Discussion

Classifying **any** anomaly as a flaw is the best approach?



Sometimes the model caught real flaws that were mislabeled as normal response



Conclusions

Conclusion

- **The best model yielded 0.96 ROC AUC, which indicates a good performance.**

Conclusion

- The best model yielded 0.96 ROC AUC, which indicates a good performance.
- **Since the model does not rely on training on specific flaw shapes, it can identify different shapes and geometries.**

Conclusion

- The best model yielded 0.96 ROC AUC, which indicates a good performance.
- Since the model does not rely on training on specific flaw shapes, it can identify different shapes and geometries.
- **It has a 118 % improvement on F2-score compared to classic threshold-based method.**

Conclusion

- The best model yielded 0.96 ROC AUC, which indicates a good performance.
- Since the model does not rely on training on specific flaw shapes, it can identify different shapes and geometries.
- It has a 118 % improvement on F2-score compared to classic threshold-based method.
- **The proposed methodology of identifying flaws through anomaly detection could be valid alternative to threshold-based methods, specially when high-amplitude geometrical artifact is present.**

Future works

- **Extract features through more advanced techniques (Convolution Neural Networks).**

Future works

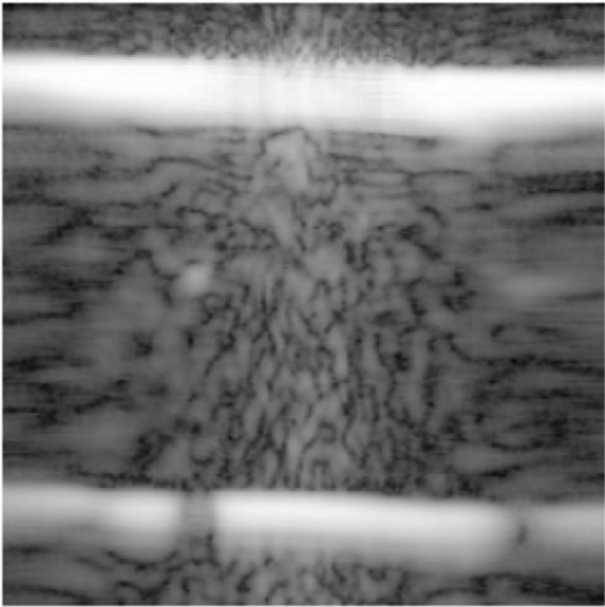
- Extract features through more advanced techniques (Convolution Neural Networks).
- **Explore more complex architectures for anomaly detection.**

Future works

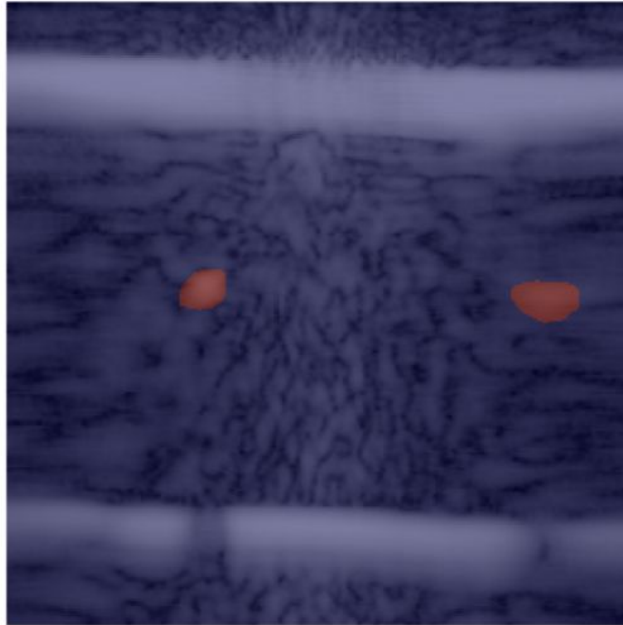
- Extract features through more advanced techniques (Convolution Neural Networks).
- Explore more complex architectures for anomaly detection.
- **Not only detect but also characterize flaws (e.g., based on their shape and size).**

Sneak peek: preliminaries results based on deep architectures.

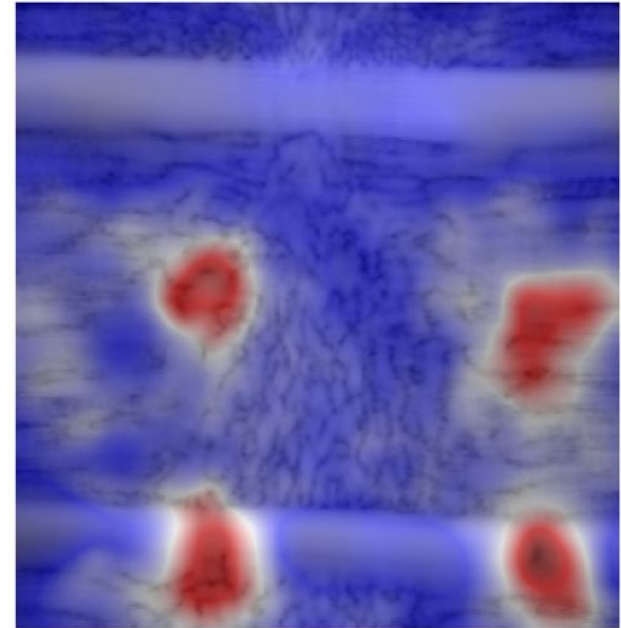
Input image



Ground-truth



Anomaly map



Acknowledgements

Thanks to

- UTFPR
- Petróleo Brasileiro S.A. (Petrobras)
- ANP / PRH-21.1
- Fundação Araucária
- CAPES



utfpr.curitiba.br/lassip/



PETROBRAS





EF
European Federation for
Non-Destructive Testing
NDT



VERONA, 15-19 JUNE

MECNDT 14th
2026

EUROPEAN CONFERENCE ON NON-DESTRUCTIVE TESTING

Thank you!

Thiago E. Kalid

thiagokkalid@gmail.com

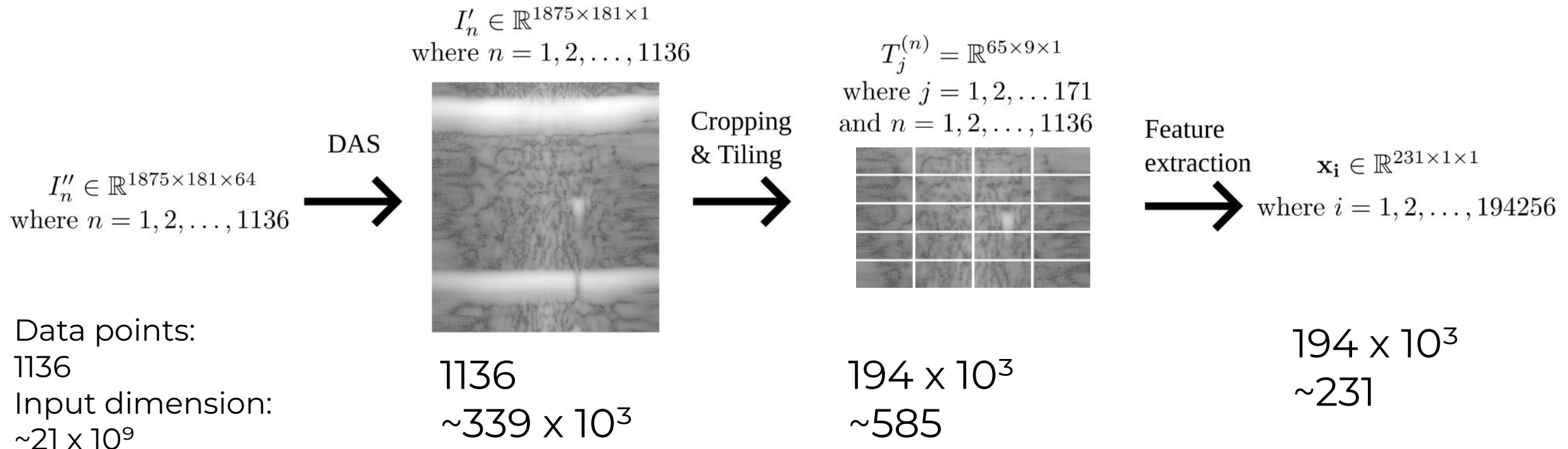
*Federal University of Technology – Parana (UTFPR)
Laboratory for Statistical Signal Processing and Inverse
Problems (LASSIP)*

*All the source code and data is openly provided as additional
material at*

<https://doi.org/10.5281/zenodo.19410832>



At each step of the data-processing step, the data complexity is reduced through non-compromising steps.



Receiver Operating Characteristic (ROC) curve:

