The project has received funding from the European Union's Horizon Europe programme under Grant Agreement N°101070115.

Scalable and Modular robotic tools for pipeline inspection and repair

TUBERS

DELIVERABLE

D7.1 VISUAL AND ULTRASONIC DATASETS FOR LEAK AND CORROSION DETECTION

Document Info

Deliverable Number	D7.1
Work Package Number and Title	WP7 Explainable machine learning algorithms development
Lead Beneficiary	UESS
Due date of deliverable	30/11/2023
Deliverable type[1]	DATA
Dissemination level[2]	PU
Author(s)	Antonios Porichis (UESS)
Internal reviewer(s)	Nikos Panagopoulos
Version - Status	v1.1 - Final

Document History

Version	Date	Modifications	Editor(s)
0.1	23/11/2023	First draft	Antonios Porichis
0.2	29/11/2023	Ultrasonic Dataset descriptions	Vahid Hassani
0.3	10/12/2023	Visual dataset descriptions	Antonios Porichis
0.4	18/12/2023	Partner comments, suggestions	Consortium members
1.0	21/12/2023	Partner suggestion integration	Antonios Porichis
1.1	27/12/2023	Format checking	Eirini Angeli

Disclaimer:

[1] **Please indicate the type of the deliverable using one of the following codes**:
R = Document, report
DEM = Demonstrator, pilot, prototype, plan designs
DEC = Websites, patents filing, press & media actions, videos
DATA = data sets, microdata
DMP = Data Management Plan

[2] **Please indicate the dissemination level using one of the following codes**:
PU = Public
SEN = Sensitive

Table of Contents

List of Figures

1. Introduction

The TUBERS project aims to introduce an ecosystem of robotic tools to holistically address inspection and repair of pipelines within water networks and beyond by combining cutting-edge robotic and AI-based tools. A crucial element of the TUBERS toolchain is the automatic analysis of the visual and ultrasonic sensor streams obtained by the robotic platforms that underpin the Decision Support System facilitating optimal inspection and maintenance scheduling.

The lack of sufficient data poses a significant challenge in the development of AI models for the inspection of water pipelines. Effective AI models require extensive and diverse datasets to learn and generalize patterns, anomalies, and potential issues within the pipelines. In the context of water pipelines, the scarcity of comprehensive data limits the ability of AI algorithms to accurately identify and predict various conditions, such as leaks, corrosion, or structural vulnerabilities. Without an ample dataset that encompasses the diverse conditions and challenges that water pipelines may face, AI models may struggle to achieve the necessary level of precision and reliability. Adequate data is crucial for training AI models to recognize subtle patterns and variations, enabling them to offer valuable insights and contribute to the proactive maintenance and monitoring of water infrastructure.

To tackle this challenge within TUBERS, we have pursued a combination of synthetic image generation based on real images from pipeline inspection as well as ultrasonic data collection from appropriately designed physical mock-ups.

The purpose of this report is to describe the datasets collected and offer some guidelines on how these can be used to stimulate further research and increase the adoption of AI-driven techniques in the crucial sector of water pipeline inspection.

2. Visual Datasets

The visual datasets comprise RGB images of pipeline interiors both real and synthetic in nature. The next paragraphs explain the collected images and the methodology of producing the synthetic instances.

2.1 Collected data

The real images have been collected with help from the end-user partners of the project, namely Evides, Brabant Water and Vitens. These contain highly diverse images collected by different camera sensors mounted on different platforms and within pipelines of different diameters and materials. Example screenshots are provided in Figure 1 below.

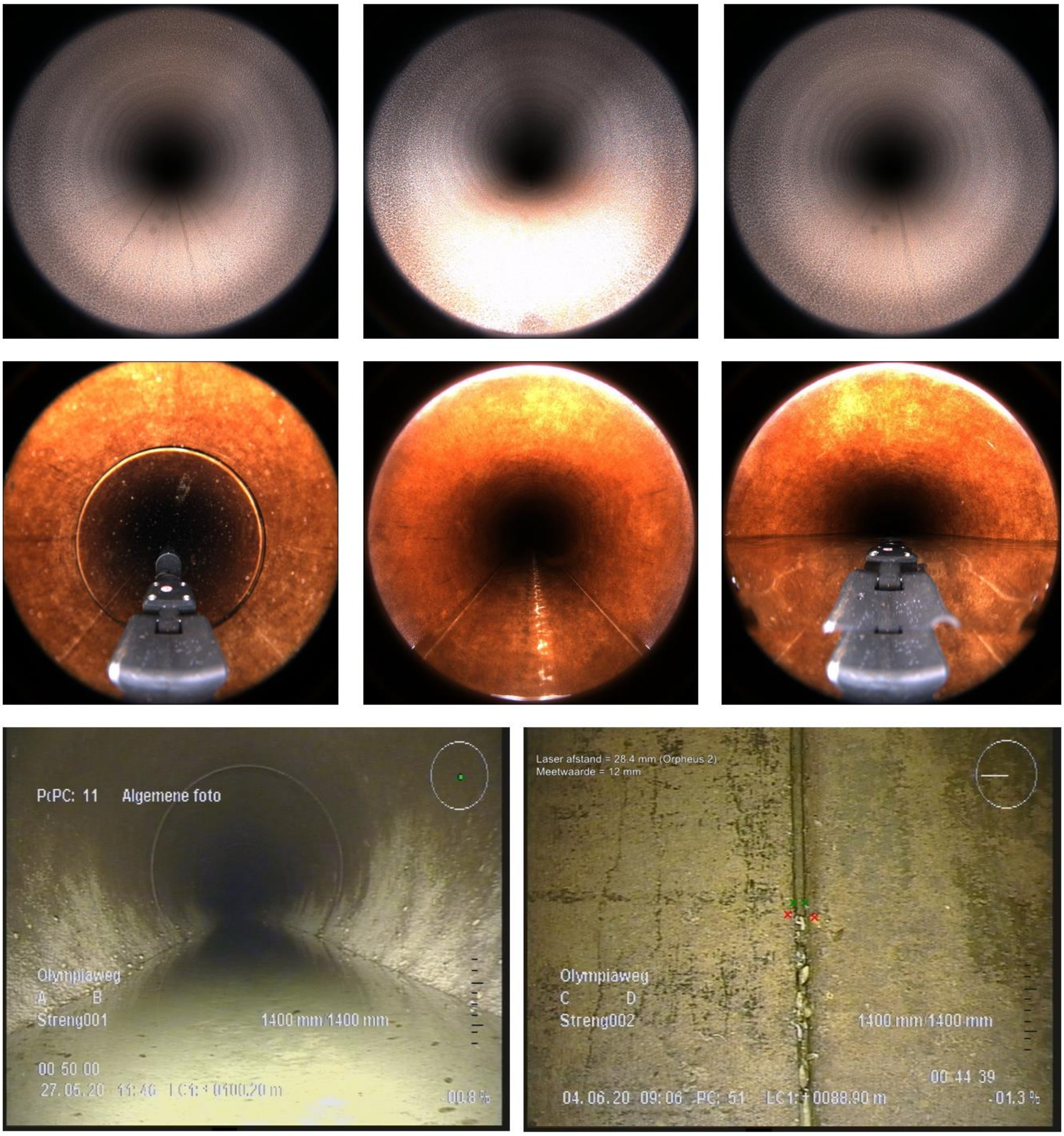

Figure 1 Images of pipeline interiors collected within the scope of TUBERS.

2.2 Synthetic crack and texture generation

To generate synthetic images containing cracks, we first create synthetic crack maps. To achieve this, we make use of gradient noise functions which we finely tuned to produce seamless, organic flow of values across two-dimensions, implicitly generating patterns that appear both random and structured simultaneously.

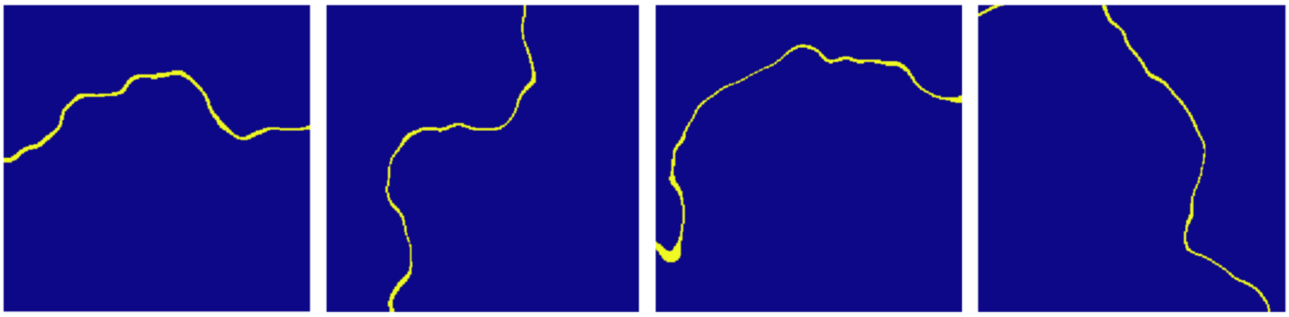

Figure 2 Examples of synthetic crack maps.

Leveraging the power of Generative AI models, we can combine arbitrary background textures with synthetic crack masks, extracted from maps such as these shown in Figure 2 above. To achieve this, we leverage generative models that are appropriately conditioned and trained on a small number of background texture images so that the local texture characteristics at multiple scales are captured by the models. With these models in place, we then proceed to iteratively blend the synthetic crack onto the background image texture. Figure 3 illustrates this effect, showing the progression from a crude mask overlay towards a realistic blend of the crack on the background texture.

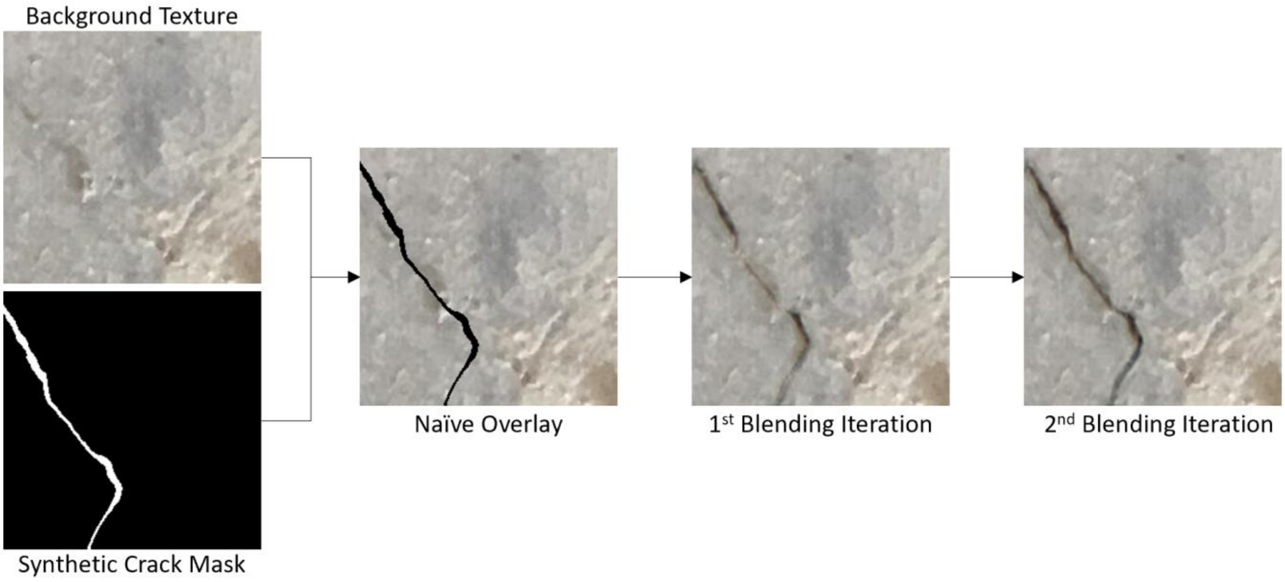

Figure 3 Synthetic crack blending onto a real texture.

Besides smooth and realistic blending of cracks onto real textures, the generative models employed also can generate synthetic textures based on the original. This is an important advantage of our approach as it enables us to create effectively infinite combinations of new textures and cracks with high realism. Figure 4 shows three examples of synthetic textures created from a real background texture segment. The real background texture was manually extracted from the pipeline interior image.

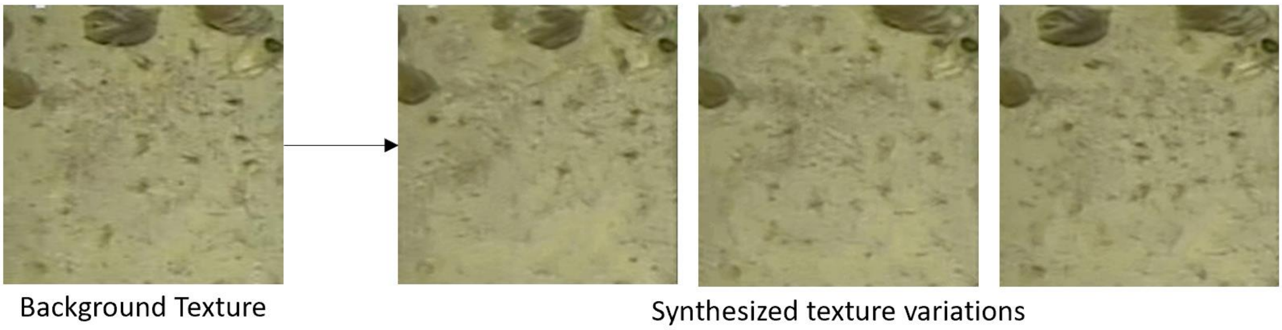

Figure 4 Synthesized textures based on original real background texture.

2.3 Dataset structure and Indicative Examples

Based on the above we have created and publicly shared over 1000 synthetic images with dense crack segmentation provided as image-mask pairs. We offer these pairs both with and without modelling the 3D structure of the pipe as shown in the figures below.

These image sets are organized within folders. Each folder contains two separate subfolders, *Images* and *Masks*. The correspondence of each image with its respective mask is made with the filename; each image and the associated mask share the same filename of the following format <No.>.png where <No.> is a natural number. The image sets are stored within .zip files named with the prefix *Visual Data* at the TUBERS' Zenodo Community[3].

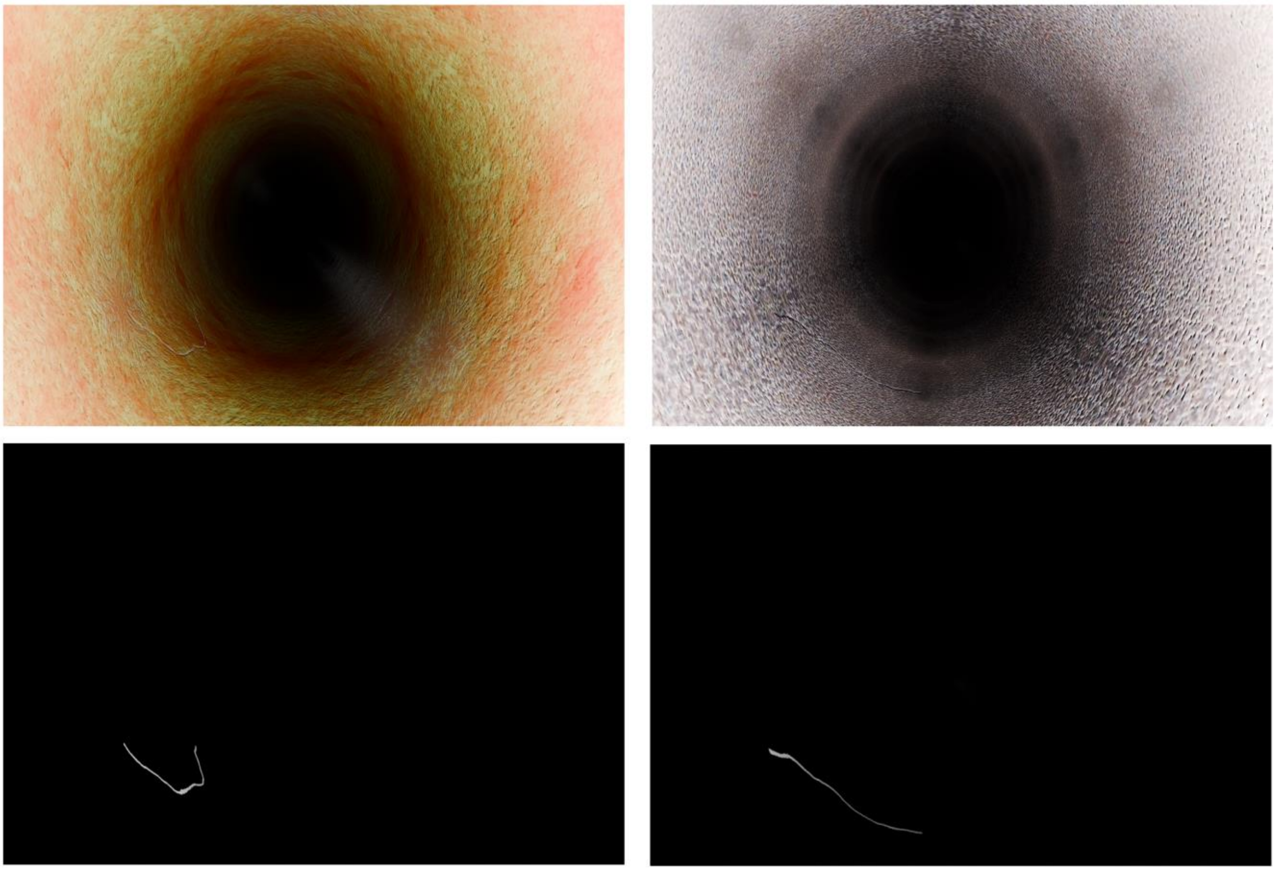

Figure 5 Examples of synthetic image-mask pairs with 3D pipeline modelling.

[3] Available at: https://zenodo.org/communities/tubers_project.

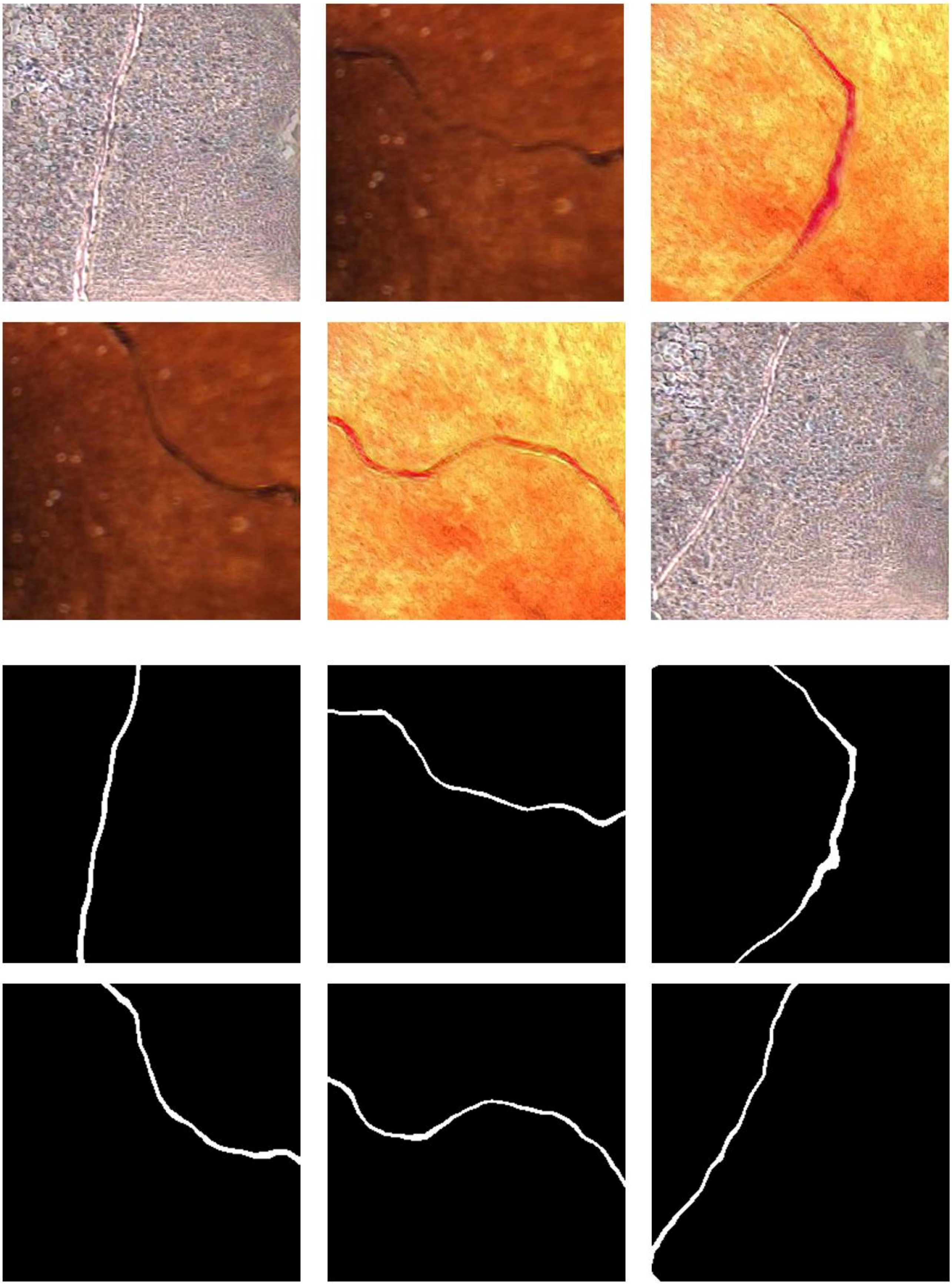

Figure 6 Examples of synthetic image mask pairs without 3D pipeline structure modelling.

3. Ultrasonic Datasets

The Ultrasonic Datasets have been compiled through extensive testing on various mock-ups both in open-air and underwater conditions. These datasets provide a solid baseline for data-driven and rule-based AI systems regarding wall thickness loss measurements.

3.1 Mock-up description

Various sensors have been explored within the ultrasonic testing context including two circular sensors with diameters measuring 2mm (10 MHz bandwidth) and 5mm (5MHz bandwidth) as well as a rectangular sensor with a cross section of 13mm x 13mm (5MHz bandwidth). Figure 7 shows the latter sensor during tests with a specimen with predetermined structural characteristics.

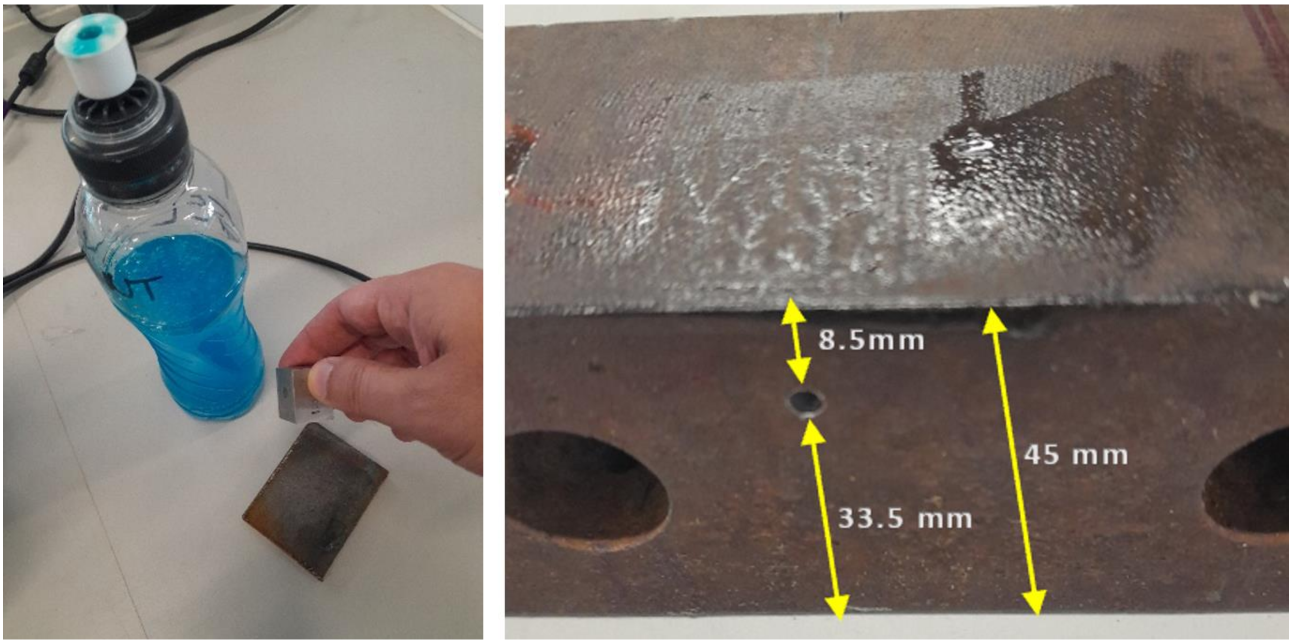

Figure 7 Left: Ultrasonic Sensor and coupling liquid, Right: Test specimen.

3.1.1 Plate Tests

Extensive tests have been carried out in two different metallic plates, made of carbon and stainless steel respectively. The first plate contained three while the second contained seven artificial defects.

The specifications of each plate are listed in Table 1 below. Figure 8 shows the two plates used for testing. Plate No.1 was scanned with both the 5MHz and the 10MHz sensors, while Plate No. 2 was scanned with the 10MHz sensor. The sampling rates were set at 200MHz and 100 MHz respectively.

Plate No.1 was manually scanned with 2mm increments starting from the larger defect to the smaller on the opposite side of the one the defects where located. For Plate No.2, a linear guide was adopted to automatically scan along the vertical direction passing through each defect. Each increment was approximately 1~1.3 mm.

Table 1 Specifications of plates used for testing

Specifications	**Plate No.1**	**Plate No.2**
Material	Carbon steel	Stainless steel
UT speed	Approx. 5890 m/s	Approx. 5740 m/s
Thickness	4 mm	11.5 mm

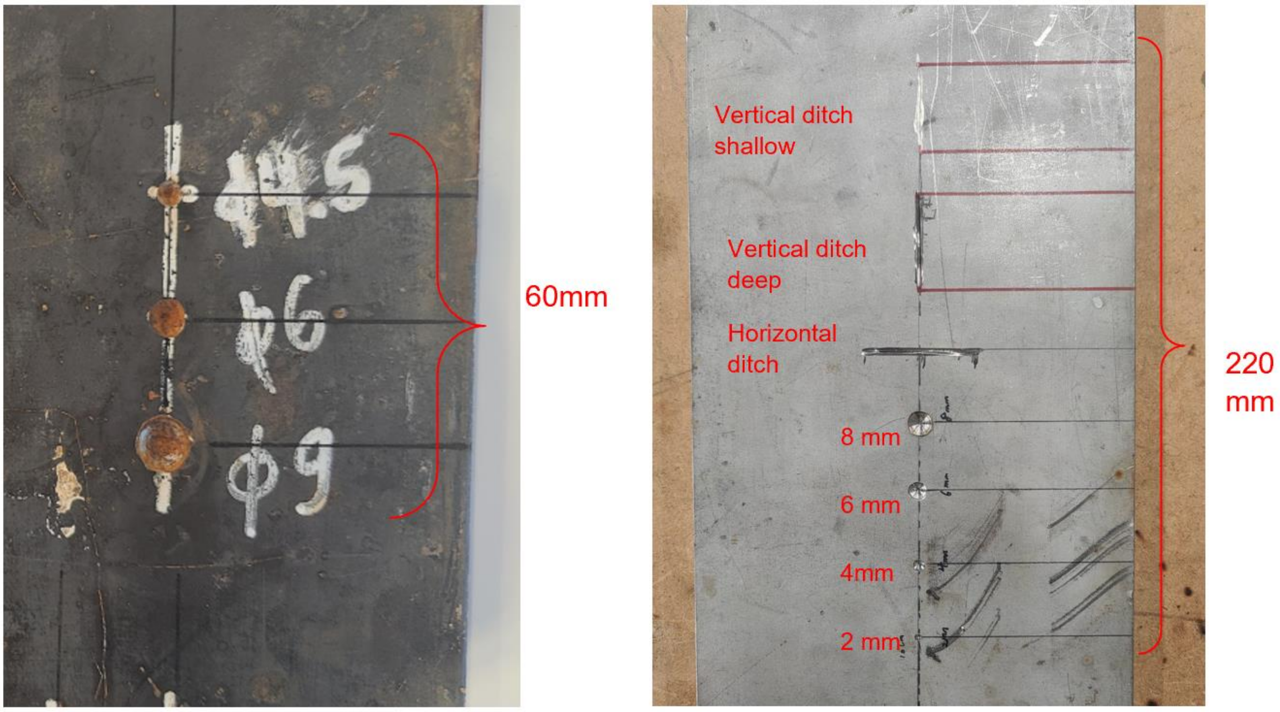

Figure 8 Plate tests; Left: Plate No.1 with three defects, Right: Plate No.2 with seven defects

3.1.2 Wedge Tests

In addition to the above, underwater experiments were carried out on a wedge of varying thicknesses, ranging from 1mm to 8mm, as depicted in Figure 9. The outcomes regarding the measured thickness in comparison to the expected thickness are detailed in Table 2. As indicated in the latter, the disparities between the sensor's measured values and the standard thicknesses of the wedge are very low.

Table 2 Comparison between Measured Values by Sensor and the Nominal Values of Wedge Thicknesses

Nominal Values (mm)	Measured Values by Sensor (mm)
1	0.98
2	2.01
3	3.04
4	3.99
5	5
6	6
7	6.98
8	8.1

Figure 9 Wedge ultrasonic measurement setup in underwater conditions; Left: Front view, Right: Side View

3.2 Dataset Structure and Examples

The ultrasonic measurements are stored within .zip files named with the prefix *Ultrasonic Data* at the TUBERS' Zenodo Community[3]. The directory contains folders with prescriptive names such as Plate No.1, Plate No.2 and the wedge tests respectively. Each folder contains comma-separated value (CSV) files with two columns, one representing the timestamp of each measurement and one representing the value of the measurement in Volts. An example is shown in the following figure.

x-axis	1
second	Volt
1.24E-05	
1.24E-05	-8.62E-02
1.24E-05	-7.72E-02
1.24E-05	-8.86E-02
1.24E-05	-1.25E-01
1.25E-05	-1.78E-01
1.25E-05	-2.41E-01
1.25E-05	-3.05E-01
1.25E-05	-3.52E-01
1.25E-05	-3.63E-01
1.25E-05	-3.33E-01
1.25E-05	-2.42E-01
1.25E-05	-7.93E-02
1.25E-05	1.49E-01
1.25E-05	4.56E-01
1.26E-05	7.30E-01
1.26E-05	8.37E-01
1.26E-05	8.97E-01
1.26E-05	9.11E-01
1.26E-05	8.81E-01
1.26E-05	8.01E-01
1.26E-05	6.53E-01
1.26E-05	3.17E-01
1.26E-05	-9.28E-02
1.26E-05	-4.18E-01
1.27E-05	-6.16E-01

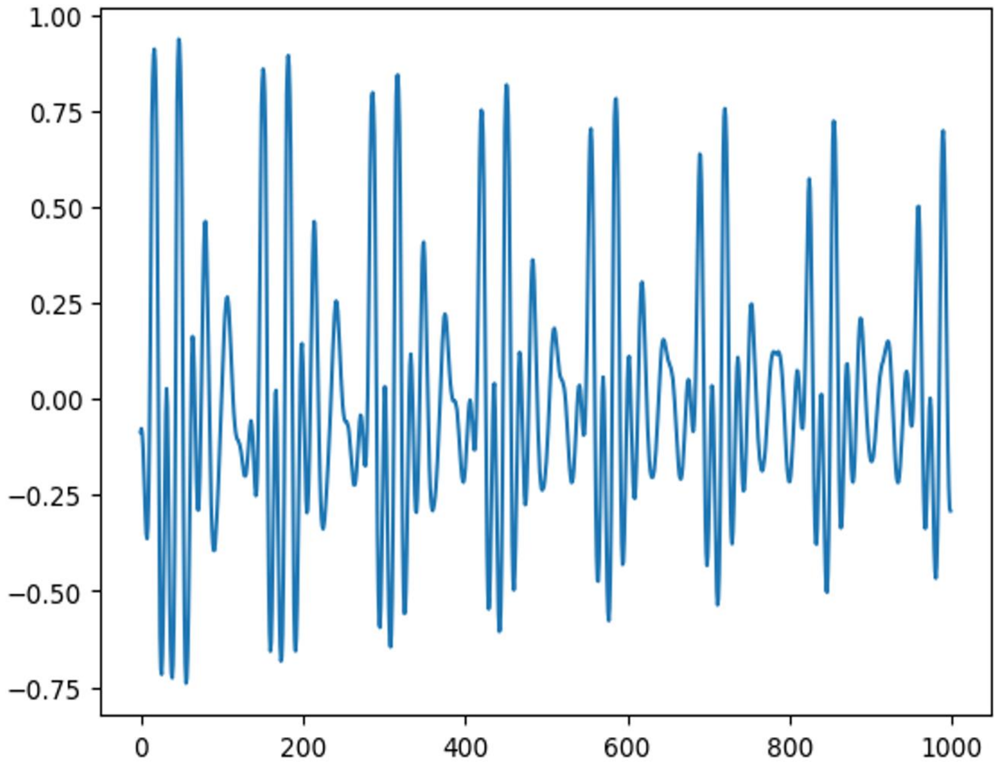

Figure 10 Example of ultrasonic measurements; Left: Screenshot of part of the CSV file, Right: Visualisation of the measurement as a timeseries.

4. Conclusions

We have provided a detailed overview of the TUBERS open datasets, including a description of the folder structure as well as the methodologies used to produce realistic synthetic images of cracks and to capture ultrasonic measurements. Synthesising realistic crack images and appropriate segmentation tasks was particularly challenging, requiring careful manual segmentation of real images as well as meticulous fine tuning of AI models. Ultrasonic datasets collection necessitated appropriate configuration of different sensors, sourcing of suitable samples and artificial defect introduction. We believe that the provided datasets will help make significant progress and explore new directions in leveraging AI techniques for water pipeline inspection and maintenance.

Future research directions could involve the exploration of cross-modal fusion techniques to enhance leak detection accuracy by combining information from both image and ultrasonic datasets. Investigating weakly supervised learning methods could reduce dependence on precise segmentation masks, making models more adaptable to diverse real-world scenarios. Transfer learning approaches might be employed to create models that can generalize across different types of water pipelines, thereby reducing the need for extensive labelled data for each specific case.

Additionally, there is potential for research in real-time leak detection algorithms, optimizing existing approaches for speed and efficiency or exploring novel techniques prioritizing real-time performance.

In practical terms, these datasets will facilitate the development of new digital tools enabling autonomous pipeline inspection, reducing the need for manual intervention, and saving valuable resources. Predictive maintenance strategies based on dataset insights could help plan preventive measures, minimizing downtime and avoiding catastrophic failures. The impact on reducing water loss and environmental damage is significant, contributing to water conservation efforts.

Looking towards the long-term roadmap, close engagement with end users and regular updates and expansions of the dataset are necessary to keep models relevant and effective. Incorporating new technologies, such as advancements in robotics, sensors, and AI, ensures the pipeline inspection process stays at the forefront of innovation. The integration of reinforcement learning allows for adaptive systems that can improve their leak detection performance over time based on feedback from operators and changing environmental conditions. Collaborative learning with operators, collecting feedback on false positives/negatives, and implementing adaptive learning plans that evolve based on real-world feedback contribute to the continuous improvement of the system.

In summary, TUBERS' datasets offer a foundation for groundbreaking research and practical applications in water pipe leakage detection. The suggested directions encompass a wide spectrum of possibilities, from refining detection algorithms to ensuring seamless integration with existing infrastructure and emphasizing the long-term evolution of the system through collaboration and adaptation.