The Indonesian Journal of Computer Science
www.ijcs.net
Volume 13, Issue 5, October 2024
https://doi.org/10.33022/ijcs.v13i5.4380

Enhancing Inspection Tasks: A Dataset for Corrosion Defects in Pipelines

Faten Aljalaud [1,2], Yousef Alohali[2]
faten.aljaloud@gmail.com[1], yousef@ksu.edu.sa[2]
[1]Computer Science Department, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11564, Saudi Arabia
[2]Computer Science Department, College of Computer and Information Sciences, King Saud University, Riyadh 11451, Saudi Arabia

Article Information

Received : 24 Aug 2024
Revised : 29 Aug 2024
Accepted : 3 Sep 2024

Keywords

corrosion; automation; dataset; inspection; pipes;

Abstract

Inspection plays a crucial role in ensuring the longevity, security, and dependability of critical public infrastructure for both governments and businesses. However, traditional inspection processes are often labor-intensive and pose various risks. Consequently, there is a growing need for automation in such tasks. This research paper presents a comprehensive dataset that can be utilized to develop algorithms and systems for automating the inspection process, a critical area in the field of computer vision. The dataset encompasses a diverse range of inspection scenarios and serves as a valuable resource for advancing automation technology specifically for the inspection of steel pipes to detect corrosion defects. Real-life pipe maps have been used to derive scenarios that represent varying levels of corrosion. By leveraging this dataset, researchers and practitioners can contribute to the development of more efficient and accurate automated inspection systems, thus greatly improving the overall efficiency and long-term safety of infrastructure inspection.

A. Introduction

Inspection tasks play a crucial role in the civil world, ensuring the safety, reliability, and longevity of infrastructure. They serve as a proactive approach to identifying and addressing potential issues before they escalate into more significant problems. By conducting regular inspections, engineers and professionals can detect structural defects, corrosion, deterioration, and other forms of damage that may compromise the integrity of buildings, bridges, dams, and other structures.

The typical maintenance and inspection processes of civil infrastructure and mechanical systems face challenges due to the bureaucratic and labor-intensive nature of the methodologies and protocols used [1]. Several studies emphasize the significance of employing robots, such as UAVs, to mitigate the hazards of inspection duties and other challenges related to inspection [1], [2], [3]. In this study, we will focus on steel pipes in this research since they are the most common type of pipe [4], [5]. According to one assessment of damage, the corrosion factor is responsible for more than half of the failures in oil and gas pipelines [6], [7], [8].

Because pipeline corrosion management is becoming more expensive in the oil and gas industry, operators are becoming increasingly concerned about corrosion management planning at all stages of production [9]. There are many types of corrosion in oil and gas pipelines, but mainly stress corrosion cracking, pitting corrosion, and erosion–corrosion [9], [10]. Corrosion in pipelines manifests in different distribution patterns. For instance, when a pit begins to form, the corrosive attack concentrates in that area [10].

Pipes are used in many fields, such as oil and gas fields [11], water infrastructure [12], sewer pipes [13], and firefighting systems [14]. Automatic or fixed sprinkler systems are safety measures in buildings required by most civil defense groups worldwide. Since corrosion is one of the main defects that affects steel pipes [15], it is our primary focus.

To the best of our knowledge, there are currently no publicly accessible datasets documenting the inspection of indoor pipes made of steel in sprinkler systems. The closest dataset we have discovered that is accessible to the public is Sun's dataset [16], [17]. They offer a dataset that comprises a grand total of 1,819 photos, consisting of 990 images showing corrosion and 829 images depicting no corrosion. This dataset can be used to build automated corrosion detection models.

Another dataset in this context is the InsPLAD dataset for inspecting power lines [18]. This dataset contains 10,607 images. Its task involves utilizing object detection to identify powerline components in UAV photos, categorizing defects in the extracted powerline component images, and implementing unsupervised anomaly detection on the extracted asset images. This dataset provides annotated powerline components that are divided into 17 distinct classes for object detection.

The general automated framework of infrastructure is explained in the steps listed in Figure 1 [19]. Previous datasets have effectively addressed the deficiencies in the data collection and decision-making capabilities of this system. Nevertheless, they have not contributed to the navigation component of this framework, which is a crucial stage of any automated inspection system.

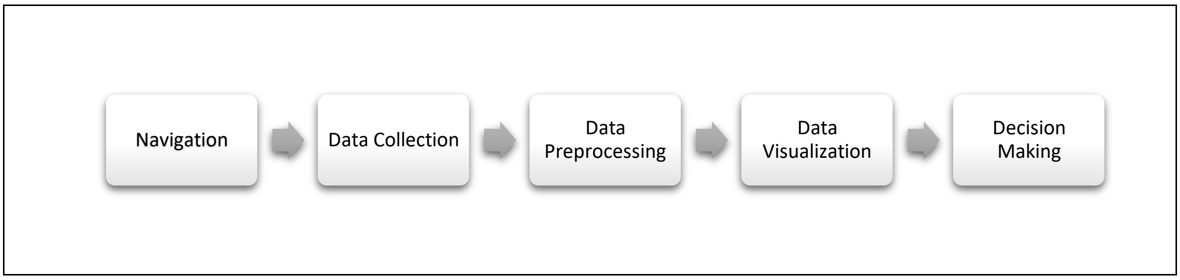


Figure 1. An infrastructure inspection framework that includes automation.

This paper introduces a dataset designed for pipe corrosion detection that addresses a critical gap in the field of automated infrastructure inspection. With a focus on the navigation and decision-making components of automated systems, this dataset accurately simulates the real-life distribution of corrosion in steel pipes. Its compatibility with various systems handling pipe defects makes this dataset remarkably versatile, offering significant advantages for infrastructure maintenance.

This paper is organized as follows: Section 2 presents the significant and application of the dataset. Section 3 outlines the methods employed. Section 4 presents a discussion on where the dataset has been applied before and how it can be used. Finally, Section 5 provides the conclusion of the paper.

B. Significance and Applications

Early detection and intervention facilitated by the proposed dataset can lead to substantial cost reductions and heightened safety in key sectors like construction and public utilities. By providing early warning signs of material degradation or structural issues, infrastructure owners and operators can take proactive measures to address problems before they escalate. This can minimize the need for costly repairs, reduce the risk of service disruptions, and prevent catastrophic failures that could endanger public safety. Furthermore, the dataset can enable more targeted and efficient maintenance scheduling, optimizing resource allocation and minimizing operational downtime.

In the realm of robotics, this dataset can refine the accuracy of autonomous inspection systems, enhancing their safety and efficiency. The detailed information on material properties, environmental conditions, and degradation patterns can help train computer vision algorithms and predictive models to better identify and classify infrastructure defects. This, in turn, can improve the reliability and decision-making capabilities of robotic inspection platforms, allowing them to navigate complex environments, detect subtle issues, and recommend appropriate interventions with a higher degree of precision.

By exploiting this dataset, there is a significant opportunity for substantial improvements in infrastructure inspection practices, yielding better efficiency, accuracy, and safety. The adoption and use of this dataset have the potential to drive significant advancements in automated inspection techniques, ultimately enhancing the quality and reliability of infrastructure inspection practices. This can lead to more proactive and data-driven maintenance strategies, reducing the risk of unexpected failures, prolonging the useful life of assets, and ensuring the continued safe and reliable operation of critical infrastructure systems. Machine learning applications are also substantial, with potential for the dataset to train algorithms in recognizing and predicting patterns of material wear and decay.

However, while poised to revolutionize maintenance and safety protocols, the utility of the dataset is limited by its scope and the genesis of its creation. It may not fully reflect the complexities of corrosion processes encountered in the field. Consequently, algorithms and models that rely on this data may need further refinement and rigorous testing against real-world conditions to ensure their efficacy. Nonetheless, by exploiting this dataset, there is an opportunity for substantial improvements in infrastructure inspection practices, yielding better efficiency, accuracy, and safety. The adoption and use of this dataset have the potential to drive significant advancements in automated inspection techniques, ultimately enhancing the quality and reliability of infrastructure inspection practices.

C. Data Description

The initial dataset comprised 18 files that were utilized in the author's earlier research [20]; however, they were not published at that time. Those files correspond to two real-life sprinkler system layouts, which will be further discussed in the next section. The files are in the csv format and named after each scenario number. The dataset was expanded to include 1,300 instances.

The folder structure of the dataset is shown in Figure 2.

- map#/
    - map#_ nwloc.csv
    - simple/
        - instanceX/
            - instanceX_defects.csv
            - instanceX_seeds.csv
    - average/
        - instanceX/
            - instanceX_defects.csv
            - instanceX_seeds.csv
    - advanced/
        - instanceX/
            - instanceX_defects.csv
            - instanceX_seeds.csv

Figure 2. Folder structure of the dataset

- **map#:** Instances generated based on a certain map, where # is the map number.
- **map#_ nwloc.csv:** The coordination of the locations corresponds to the pipes on the map with a certain number, #. The data in each row are separated by commas (,) and include the following information: the x coordinates of the network location of the pipes and the y coordinates of the network location of the pipes.

- **simple:** Instances generated with a simple severity level in the defects (represented by number of defects).
- **average:** Instances generated with an average severity level in the defects (represented by the number of defects).
- **Advanced:** Instances generated with an advanced severity level in the defects (represented by the number of defects).
- **instanceX:** A file that is labeled with a number ranging from 1 to 200.

- **instanceX_defects.csv:** Every row consists of the following data, which is separated by commas (,): the x coordinates of a defect location in the pipes and the y coordinates of a defect location in the pipes.
- **instanceX_seeds.csv:** Each row contains the following data, delimited by commas (,): the x and y coordinates of the seed position that corresponds to the defect in the pipes.

D. Methods

The data were generated based on a template of a fire sprinkler system (RCP) using the Edrawmax tool. This template was used as an input map to mimic a realistic scenario as much as possible [21] [22]. The first input map is shown in Figure 3, and the second input map is shown in Figure 4. The maps were handled as an occupancy matrix after preprocessing (as per Figure 5 and Figure 6) to form a 500 × 500 grid.

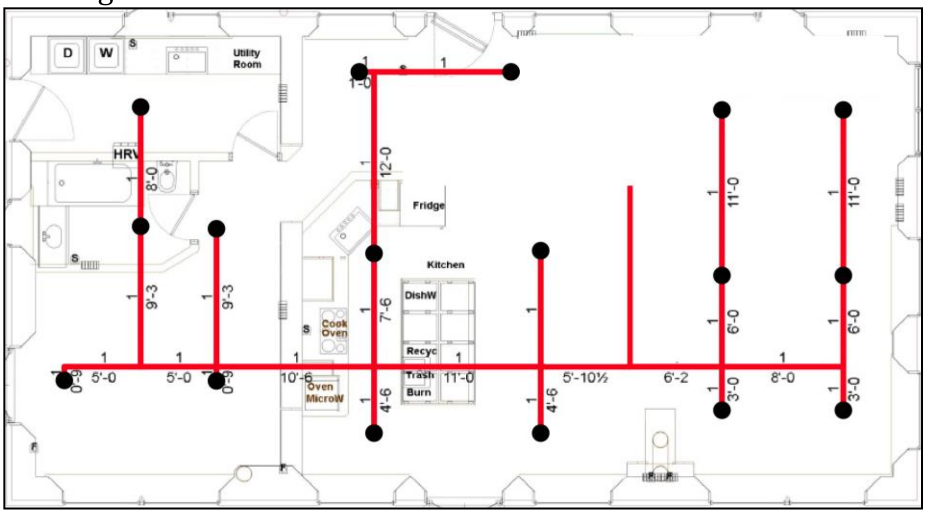


Figure 3. Input map#1 of the pipe network.

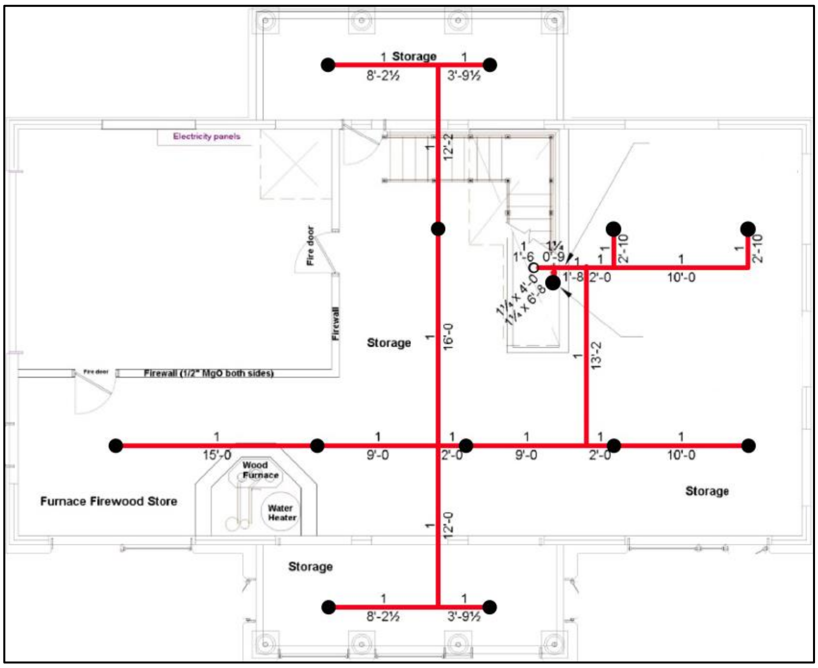


Figure 4. Input map#2 of the pipe network.

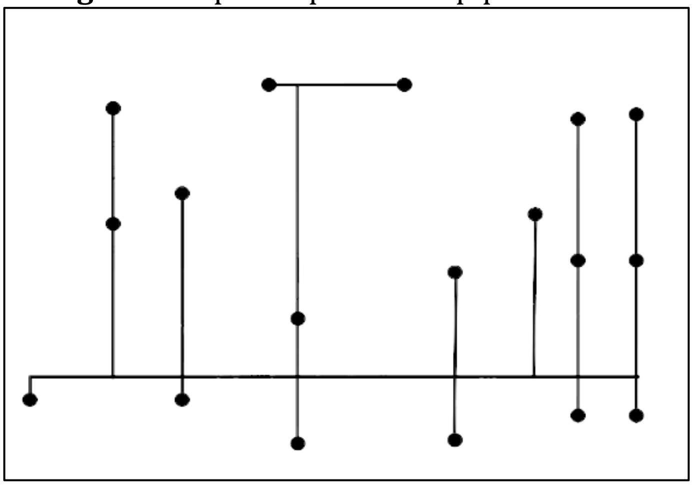


Figure 5. Input map#1 after preprocessing.

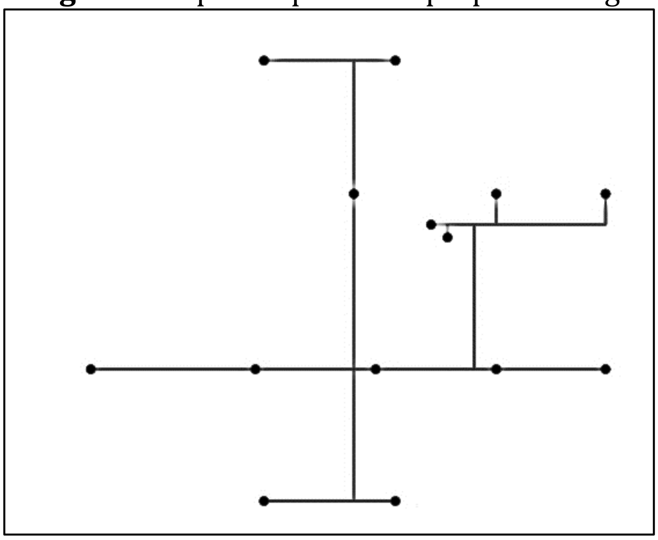


Figure 6. Input map#2 after preprocessing.

The network locations served as inspection locations, and the defect locations were randomly distributed within these points to guarantee representative samples in all tests. Moreover, the defect locations were randomly created by employing uniformly distributed within a disc using the Polar Coordinates Method [23] to simulate a homogeneous spatial distribution within the given radius, where defects were concentrated and had a higher probability of being detected. The method involves two steps: First, a random angle θ $\theta$is generated, uniformly distributed between 0 and 2π 2π, using Equation(1) where U_1 is a random number uniformly distributed between 0 and 1.

$$\theta = 2\pi * U_1 \tag{1}$$

Second, a random radius r is generated using Equation (2), where $raduis$ is the given radius of the disc and U_2 is another random number uniformly distributed between 0 and 1. These polar coordinates (θ, r) are then converted to Cartesian coordinates (x, y) using the equations (3) and (4). By employing Equations (1) through (4), we ensure that the points are uniformly distributed over the area of the disc.

$$r = radius * \sqrt{U_2} \tag{2}$$
$$x = r * \cos\theta \tag{3}$$
$$y = r * \sin\theta \tag{4}$$

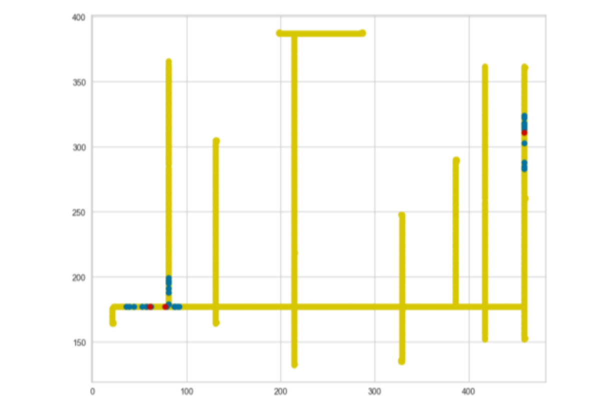
The scenarios were generated by varying the total defect count, incorporating both hotspots and the concentration of defects within a single hotspot. The mean corrosion rate of the steel samples was 1.53% [24], although another study reported a maximum corrosion rate of 17.5% [25].

We chose a range of numbers, specifically, from 1% to 30%, from all the network locations to ensure that all potential severity levels were represented in the steel systems. The values corresponding to the various severity levels of the faults (indicated by the number of defects) can be found in Table 1.

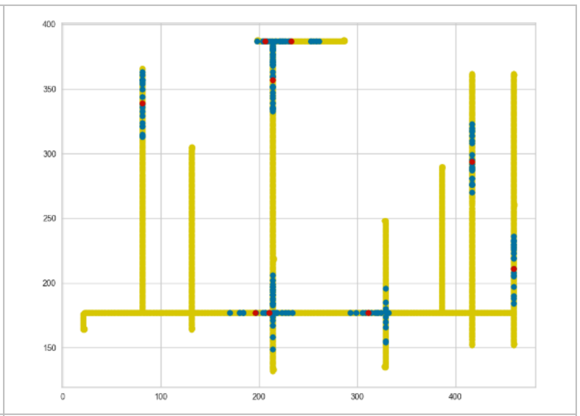
The visual representation of random instances of the dataset can be seen in Figure 7 and Figure 8. The coordinations depicted in this picture correlate to the seeds, defects, and network locations, represented by the colors red, blue, and yellow, respectively.

Table 1. Values of the different severity levels of the defects.

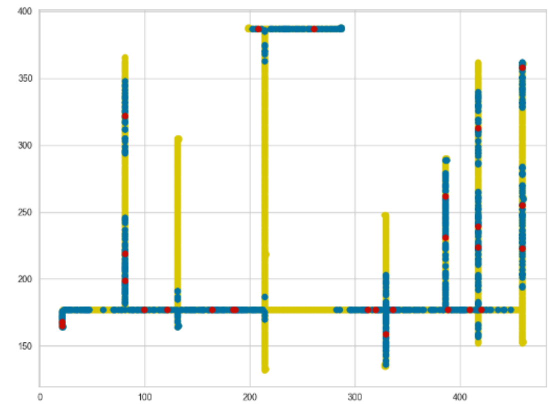
Severity Level	Number of Hotspots	Radius of Hotspot	Number of Defects
Simple	3	30	10
Average	9	30	20
Advanced	27	30	30



(a)

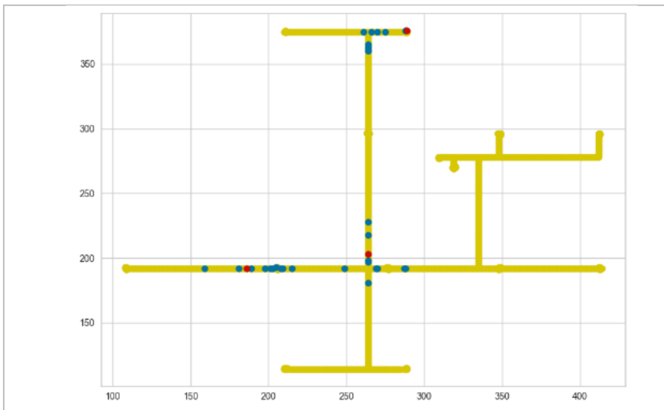


(b)

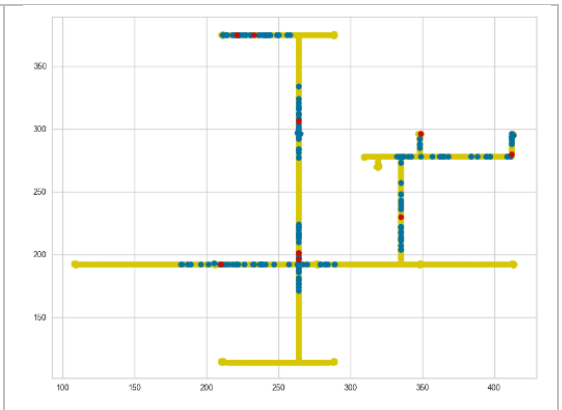


(c)

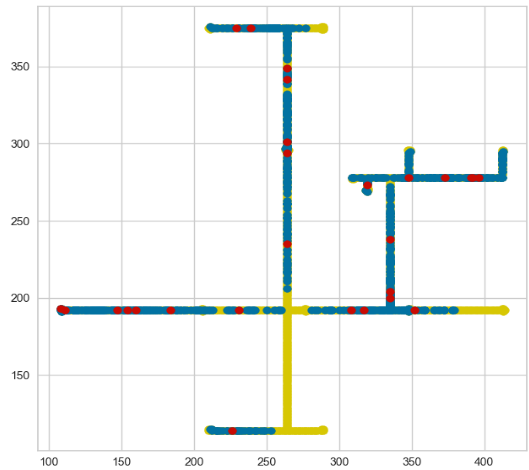
Figure 7. Maps of different severity levels corresponding to the original map#1: **(a)** simple settings; **(b)** average settings; **(c)** advanced settings.



(a)



(b)



(c)

Figure 8. Maps of different severity levels corresponding to the original map#2: **(a)** simple settings; **(b)** average settings; **(c)** advanced settings.

E. Discussion

In our previous research, we developed a novel multi-UAV path planning method that leveraged a dataset unique for its comprehensive coverage of various operational scenarios and UAV behaviors [20]. This dataset was instrumental in enabling the UAVs to autonomously assume diverse roles and efficiently navigate through different zones, adapting their strategies to minimize the total distance traveled. By integrating established algorithms such as Ant Colony Optimization, Particles Swarm Optimization, and OTA strategy, and a baseline random method, adapted for multi-UAV contexts, we were able to rigorously compare their performances using our dataset. The outcomes were compelling; our method significantly outperformed all benchmarks, particularly in energy efficiency and defect detection speed.

For instance, in basic severity settings, our approach enhanced mean detection times by 59% and increased operational speed threefold compared to the random algorithm. These improvements were even more pronounced with a scaling number of UAVs, demonstrating the dataset's critical role in optimizing and scaling UAV operations. The dataset not only supported the development of this innovative path planning technique but also proved essential in establishing the method's superiority across different severity levels and operational metrics. The clear and consistent performance advantages observed in our experiments underscore the dataset's value, marking it as a pivotal asset in pushing the boundaries of UAV path planning research.

To facilitate the reproducibility and further adaptation of our innovative approach, we have carefully designed the dataset using Python, a popular programming language known for its versatility and robust data handling capabilities. We provide a clear, step-by-step guide on how to access and utilize this dataset within Python scripts, aiming to integrate seamlessly into a variety of research workflows. To ensure ease of use and enhance compatibility across different systems, we recommend organizing the dataset within a **data_instances** folder located in the working directory of your Python project, utilizing relative paths for all file references.

The initial step in the data handling process involves loading the **model_nwloc.csv** file. This file is crucial as it contains the geographical coordinates of various locations within the network or pipeline system, each marked as a point. Following the data loading, our script employs a structured approach to iterate through the dataset. In each iteration, the script processes a segment of the dataset, carefully extracting and categorically storing data regarding the positions of seeds and defects. These positions are stored in designated arrays, allowing for systematic access and analysis.

This structured data processing method not only facilitates an organized exploration and manipulation of the dataset but also ensures that researchers can easily adapt the dataset to fit their specific experimental setups. By providing these details, we aim to support other researchers in navigating the dataset effectively, enabling them to leverage it for enhancing the accuracy and efficiency of their own UAV path planning or similar studies.

F. Conclusion

The focus of this study is on the inspection of steel pipes, which are widely used in different industries. Corrosion has been identified as a major issue affecting the integrity of pipelines, particularly in the oil and gas sector. The presented dataset provides a valuable resource for developing algorithms and systems to automate the inspection process, specifically for detecting corrosion defects in steel pipes.

By accurately representing various levels of corrosion derived from real-life pipe maps, this dataset offers researchers and practitioners the opportunity to advance automation technology in the field of infrastructure inspection. The dataset's inclusion of a range of inspection scenarios contributes to the development of efficient and accurate automated inspection systems.

Data Availability Statement: The data presented in this study are openly available **https://doi.org/10.5281/zenodo.10809418**

G. References

[1] M. O. Macaulay and M. Shafiee, "Machine learning techniques for robotic and autonomous inspection of mechanical systems and civil infrastructure," *Auton. Intell. Syst.*, vol. 2, no. 1, p. 8, Dec. 2022, doi: 10.1007/s43684-022-00025-3.

[2] M. Fisher *et al.*, "An Overview of Verification and Validation Challenges for Inspection Robots," *Robotics*, vol. 10, no. 2, p. 67, Apr. 2021, doi: 10.3390/robotics10020067.

[3] A. J. Lee, W. Song, B. Yu, D. Choi, C. Tirtawardhana, and H. Myung, "Survey of robotics technologies for civil infrastructure inspection," *Journal of Infrastructure Intelligence and Resilience*, vol. 2, no. 1, p. 100018, Mar. 2023, doi: 10.1016/j.iintel.2022.100018.

[4] Z. Panossian, N. L. de Almeida, R. M. F. de Sousa, G. de S. Pimenta, and L. B. S. Marques, "Corrosion of carbon steel pipes and tanks by concentrated sulfuric acid: A review," *Corrosion Science*, vol. 58, pp. 1–11, May 2012, doi: 10.1016/j.corsci.2012.01.025.

[5] N. S. Hassan, “The Effect of Different Operating Parameters on the Corrosion Rate of Carbon Steel in Petroleum Fractions,” *Eng. Technol. J.*, p. 12, 2013.
[6] P. Maruschak, O. Prentkovskis, and R. Bishchak, “Defectiveness of external and internal surfaces of the main oil and gas pipelines after long-term operation,” *J. Civil Eng. Manag.*, vol. 22, no. 2, pp. 279–286, Mar. 2016, doi: 10.3846/13923730.2015.1100672.
[7] A. H. Al-Moubaraki and I. B. Obot, “Top of the line corrosion: causes, mechanisms, and mitigation using corrosion inhibitors,” *Arabian Journal of Chemistry*, vol. 14, no. 5, p. 103116, May 2021, doi: 10.1016/j.arabjc.2021.103116.
[8] C. Popescu and M. R. Gabor, “Quantitative Analysis Regarding the Incidents to the Pipelines of Petroleum Products for an Efficient Use of the Specific Transportation Infrastructure,” *Processes*, vol. 9, no. 9, p. 1535, Aug. 2021, doi: 10.3390/pr9091535.
[9] C. I. Ossai, “Advances in Asset Management Techniques: An Overview of Corrosion Mechanisms and Mitigation Strategies for Oil and Gas Pipelines,” *ISRN Corrosion*, vol. 2012, pp. 1–10, Nov. 2012, doi: 10.5402/2012/570143.
[10] Michael Baker and Raymond R. Fessler, “Pipeline Corrosion,” U.S. Department of Transportation, Washington, DC, USA, 2008.
[11] A. Shukla and H. Karki, “Application of robotics in onshore oil and gas industry—A review Part I,” *Robotics and Autonomous Systems*, vol. 75, pp. 490–507, Jan. 2016, doi: 10.1016/j.robot.2015.09.012.
[12] R. Worley *et al.*, “Robot Localization in Water Pipes Using Acoustic Signals and Pose Graph Optimization,” *Sensors*, vol. 20, no. 19, p. 5584, Sep. 2020, doi: 10.3390/s20195584.
[13] H. Plihal *et al.*, “A novel method for rapid inspection of sewer networks: combining acoustic and optical means,” *Urban Water Journal*, vol. 13, no. 1, pp. 3–14, Jan. 2016, doi: 10.1080/1573062X.2015.1076857.
[14] British Standards Institution, “Fixed firefighting systems: components for sprinkler and water spray systems. Part 1. Part 1.,” 1999.
[15] “The Cause of Corrosion in a Fire Sprinkler System.” Accessed: Nov. 12, 2022. [Online]. Available: https://www.ryanfp.com/cause-of-corrosion/
[16] P. Sun, “Deep Learning for Automated Corrosion Detection.” Accessed: Jan. 08, 2023. [Online]. Available: https://github.com/pjsun2012/Phase5_Capstone-Project
[17] J. Yosinski, J. Clune, A. Nguyen, T. Fuchs, and H. Lipson, “Understanding Neural Networks Through Deep Visualization,” Jun. 22, 2015, *arXiv*: arXiv:1506.06579. Accessed: Jan. 09, 2024. [Online]. Available: http://arxiv.org/abs/1506.06579
[18] A. L. B. V. e Silva *et al.*, “InsPLAD: A Dataset and Benchmark for Power Line Asset Inspection in UAV Images,” Dec. 03, 2023. doi: 10.1080/01431161.2023.2283900.
[19] S. Agnisarman, S. Lopes, K. C. Madathil, K. Piratla, and A. Gramopadhye, “A survey of automation-enabled human-in-the-loop systems for infrastructure visual inspection,” *Autom. Constr.*, vol. 97, pp. 52–76, Jan. 2019, doi: 10.1016/j.autcon.2018.10.019.

[20] F. Aljalaud, H. Kurdi, and K. Youcef-Toumi, “Autonomous Multi-UAV Path Planning in Pipe Inspection Missions Based on Booby Behavior,” *Mathematics*, vol. 11, no. 9, p. 2092, Apr. 2023, doi: 10.3390/math11092092.
[21] L. Anderson, “Fire Sprinkler System Rcp.” Accessed: Dec. 15, 2022. [Online]. Available: https://www.edrawmax.com/templates/1021321/
[22] “Interior Infrastructure - Fire Sprinklers,” carnation construction. Accessed: Jun. 27, 2023. [Online]. Available: https://www.carnationconstruction.com/Techniques/07-03-Techniques-InteriorInfrastructure-FireSprinklers.html
[23] G. James, *Random Number Generation and Monte Carlo Methods*. Springer, 2003.
[24] N. Modak, D. Das, and R. V. Kumar, “Behavior of AFRP Composite and its Practical Aspects in the Invigoration of Structural and Materialistic Properties of Corroded SHTS,” *IJEAT*, vol. 9, no. 2, pp. 2549–2557, Dec. 2019, doi: 10.35940/ijeat.B3851.129219.
[25] H. Reis, B. L. Ervin, D. A. Kuchma, and J. T. Bernhard, “Estimation of Corrosion Damage in Steel Reinforced Mortar Using Guided Waves,” *Journal of Pressure Vessel Technology*, vol. 127, no. 3, pp. 255–261, Aug. 2005, doi: 10.1115/1.1989352.