CHAPTER 1

1. INTRODUCTION

1.1 PROJECT OVERVIEW

Pipelines are critical infrastructure components, facilitating the transportation of various fluids such as oil, gas, water, and chemicals over long distances. Ensuring the integrity of these pipelines is paramount to prevent leaks, ruptures, and environmental disasters. Traditional manual inspection methods, although effective to some extent, are often labour-intensive, time-consuming, and subject to human error. In recent years, there has been a growing interest in leveraging advancements in computer vision and machine learning to automate pipeline integrity assessment processes. Automated pipeline integrity assessment using object detection models represents a promising approach to streamline inspection workflows, enhance accuracy, and reduce reliance on manual labour. By harnessing the power of artificial intelligence, operators can analyse visual data captured from drones, cameras, or specialized robots with greater efficiency and precision. The adoption of object detection models for pipeline inspection offers several advantages. Firstly, it enables the detection and localization of anomalies such as cracks, corrosion, dents, and other structural defects with high accuracy and reliability. Secondly, it reduces the need for manual inspection efforts, thereby saving time, resources, and operational costs. Moreover, automated analysis can facilitate proactive maintenance interventions, preventing potential failures and ensuring the safety and integrity of the pipeline network. Automated pipeline integrity assessment using object detection models holds great promise for revolutionizing the way pipeline inspections are conducted. By harnessing the capabilities of artificial intelligence, operators can improve efficiency, accuracy, and safety while minimizing environmental risks and ensuring the reliability of critical infrastructure networks.

1.2 OBJECTIVES

Automated pipeline integrity assessment using object detection models aims to revolutionize traditional inspection methods by enhancing efficiency, accuracy, and safety. By automating the analysis of visual data collected from drones, cameras, or specialized robots, operators can streamline inspection processes, reducing time and resources required while improving the precision of anomaly detection. The primary objectives include enhancing efficiency by facilitating more frequent and thorough assessments, improving accuracy to minimize the risk of false positives or negatives, and reducing operational costs through optimized resource allocation. Additionally, the deployment of automated systems aims to mitigate risks associated with pipeline failures by identifying potential hazards early and enabling proactive maintenance interventions. By leveraging data-driven insights from automated inspections, operators can make informed decisions regarding maintenance prioritization and resource allocation, fostering continuous improvement in pipeline management practices. Ultimately, the overarching goal is to ensure the long-term reliability, safety, and sustainability of pipeline infrastructure networks while meeting regulatory requirements and industry standards.

1.3 PROBLEM STATEMENT

Despite the critical importance of maintaining the integrity of pipelines for the safe transportation of fluids, traditional manual inspection methods are often labor-intensive, time-consuming, and prone to human error. These methods may not adequately detect subtle anomalies or defects, increasing the risk of leaks, ruptures, and environmental disasters. Moreover, the growing complexity and expansiveness of pipeline networks further exacerbate the challenges associated with manual inspection. To address these limitations, there is a pressing need to develop and implement automated pipeline integrity assessment systems using advanced technologies such as object detection models. However, several challenges must be overcome to effectively deploy and utilize these systems in real-world scenarios.

Data Quality and Availability: Obtaining high-quality visual data from diverse sources such as drones, cameras, or robots while ensuring coverage of the entire pipeline network poses a significant challenge. Variations in lighting conditions, weather, and environmental factors can also impact data quality and consistency.

Anomaly Detection Accuracy: Object detection models must accurately identify and classify various types of anomalies, including cracks, corrosion, dents.

1.4 DEEP LEARNING

Deep Learning is a subset of machine learning that involves training artificial neural networks with multiple layers to recognize patterns in data. Deep learning algorithms can be used for a wide range of tasks such as image and speech recognition, natural language processing, and even playing games like Go and Chess. The main advantage of deep learning over traditional machine learning approaches is its ability to automatically learn features from raw data without the need for manual feature engineering. This is accomplished by stacking multiple layers of neurons, each of which performs a nonlinear transformation of the input data. The output of one layer serves as the input for the next layer, allowing the network to gradually learn increasingly complex representations of the input data. Popular deep learning algorithms include Convolutional Neural Networks (CNNs) for image and video processing, Recurrent Neural Networks (RNNs) for sequential data processing such as natural language processing, and Generative Adversarial Networks (GANs) for generating realistic images and videos. Training deep learning models requires large amounts of labeled data and significant computational resources. However, recent advancements in hardware and software have made it easier to train deep learning models on a wide range of applications.

Deep learning algorithms are based on artificial neural networks, which are inspired by the structure and function of the human brain. The networks consist of layers of interconnected nodes, or neurons, that process information in a hierarchical manner. The input data is fed into the first layer of the network, which extracts basic features. The output of this layer is then passed to the next layer, which extracts more complex features based on the previous layer's output, and so on. The process of training a deep learning model involves adjusting the weights and biases of the network's neurons to minimize the difference between the predicted output and the actual output. This is done by using a loss function that quantifies the difference between the predicted and actual output, and an optimization algorithm that updates the network's weights and biases to minimize this loss function. The most commonly used optimization algorithm is called stochastic gradient descent.

One of the key advantages of deep learning is its ability to handle unstructured data such as images, video, and text. Convolutional Neural Networks (CNNs) are particularly effective at processing images and video, while Recurrent Neural Networks (RNNs) are better suited for sequential data processing such as natural language processing. Deep learning has had a significant impact on a wide range of industries, including healthcare, finance, and transportation. For example, deep learning algorithms are used in medical imaging to help diagnose diseases such as cancer, in finance to detect fraudulent transactions, and in transportation to improve self-driving cars' performance. However, deep learning is not without its challenges.

One of the biggest challenges is the need for large amounts of labeled data to train the models effectively. This can be particularly challenging for applications where the data is scarce or expensive to collect. Additionally, deep learning models are often black boxes, meaning it can be challenging to interpret how the model arrives at its predictions. This can be problematic for applications where interpretability is important, such as in healthcare or finance.

1.5 DEEP LEARNING ALGORITHMS

There are several types of deep learning algorithms, each of which is designed to solve different types of problems. Some of the most popular deep learning algorithms include:

Convolutional Neural Networks (CNNs): These are commonly used for image and video processing. They use a technique called convolution to extract features from the input image or video.

Recurrent Neural Networks (RNNs): These are used for sequential data processing, such as natural language processing. They can capture the context and relationship between different elements in a sequence.

Generative Adversarial Networks (GANs): These are used for generating new data that is similar to the input data. They consist of two networks: a generator network that generates new data and a discriminator network that evaluates whether the generated data is similar to the real data.

Autoencoders: These are used for unsupervised learning and feature extraction. They consist of an encoder network that compresses the input data into a lower-dimensional representation, and a decoder network that reconstructs the original input from the compressed representation.

Deep Belief Networks (DBNs): These are used for unsupervised learning and feature extraction. They consist of multiple layers of restricted Boltzmann machines (RBMs) that can learn hierarchical representations of the input data.

Long Short-Term Memory (LSTM) Networks: These are a type of RNN that is designed to handle long-term dependencies in sequential data. They use memory cells and gates to selectively remember or forget information from previous time steps.

Each of these deep learning algorithms has its own strengths and weaknesses, and the choice of algorithm depends on the specific problem being solved.

1.6 MACHINE LEARNING VERUS DEEP LEARNING

Machine learning and deep learning are both subsets of artificial intelligence, but they differ in the types of problems they are best suited for and the techniques they use.

Machine learning algorithms are typically used for supervised and unsupervised learning tasks. In supervised learning, the algorithm is trained on labeled data, and the goal is to predict the output for new, unseen data. In unsupervised learning, the algorithm is trained on unlabeled data, and the goal is to find patterns or structure in the data.

Deep learning, on the other hand, is a subset of machine learning that uses neural networks with multiple layers to learn hierarchical representations of the input data. Deep learning is particularly effective at handling unstructured data such as images, video, and natural language text. One of the key differences between machine learning and deep learning is the amount of labeled data required to train the models effectively. Machine learning algorithms typically require a smaller amount of labeled data than deep learning algorithms. This makes machine learning more suitable for applications where labeled data is scarce or expensive to obtain.

Another difference is in the interpretability of the models. Machine learning models are often easier to interpret than deep learning models, as the features learned by machine learning algorithms are typically more transparent. This can be an advantage in applications where interpretability is important, such as in healthcare or finance. In summary, machine learning and deep learning are both powerful tools in the field of artificial intelligence, but they differ in the types of problems they are best suited for, the amount of labeled data required, and the interpretability of the models.

Another important difference between machine learning and deep learning is the computational resources required to train the models. Deep learning algorithms typically require more computational resources, including specialized hardware such as graphics processing units (GPUs), to train the models effectively. This can make deep learning more expensive and time-consuming than machine learning. Additionally, while machine learning algorithms can be effective for many types of problems, deep learning algorithms are particularly well-suited for problems that involve high-dimensional data, such as images and video. This is because deep learning models can learn hierarchical representations of the input data that capture complex patterns and relationships in the data. In contrast, machine learning algorithms are often used for problems that involve structured data, such as tabular data in databases. For example, machine learning algorithms can be used to predict customer churn based on customer data such as age, gender, and purchase history. It's worth noting that the line between machine learning and deep learning is not always clear-cut, as many techniques and algorithms can be classified as both. For example, decision trees and support vector machines are often classified as machine learning algorithms, but they can also be used as building blocks for deep learning models. Ultimately, the choice between machine learning and deep learning depends on the specific problem being solved, the amount of labeled data available, and the computational resources available. Both machine learning and deep learning have their own strengths and weaknesses, and the choice of technique depends on the context and requirements of the problem at hand.

1.7 ADVANTAGES OF DEEP LEARNING

Deep learning has several advantages over traditional machine learning techniques. Some of the key advantages of deep learning include:

Ability to handle large amounts of data: Deep learning algorithms are particularly well-suited for handling large amounts of data, such as images, video, and natural language text. By using neural networks with multiple layers, deep learning models can learn hierarchical representations of the input data that capture complex patterns and relationships in the data. Ability to learn from unstructured data: Deep learning is particularly effective at handling unstructured data, such as images, video, and natural language text. By using convolutional neural networks (CNNs), recurrent neural networks (RNNs), and other types of deep learning models, it is possible to learn meaningful features from unstructured data.

Better performance on complex tasks: Deep learning models are capable of learning complex patterns and relationships in the data, making them well-suited for complex tasks such as image and speech recognition, natural language processing, and autonomous driving.

Higher accuracy: Deep learning models can achieve higher accuracy than traditional machine learning models, particularly on complex tasks. This is because deep learning models can learn more complex patterns and relationships in the data.

Automated feature engineering: Deep learning models can automatically learn features from the input data, eliminating the need for manual feature engineering. This can save time and improve the accuracy of the model.

Overall, deep learning is a powerful tool for solving complex problems in a wide range of domains, including computer vision, natural language processing, and autonomous systems. While deep learning requires more computational resources than traditional machine learning techniques, the benefits of deep learning can be substantial in terms of accuracy, performance, and the ability to handle large amounts of unstructured data.

1.8 DISADVANTAGES OF DEEP LEARNING

Despite its many advantages, deep learning also has several disadvantages that should be considered. Some of the key disadvantages of deep learning include:

Requires large amounts of data: Deep learning algorithms typically require large amounts of labeled data to train the models effectively. This can be a disadvantage in applications where labeled data is scarce or expensive to obtain.

Requires substantial computational resources: Deep learning algorithms require substantial computational resources, including specialized hardware such as graphics processing units (GPUs), to train the models effectively. This can make deep learning more expensive and time-consuming than traditional machine learning techniques.

Can be difficult to interpret: Deep learning models can be difficult to interpret, as the features learned by the models are often complex and opaque. This can make it difficult to understand how the model is making its predictions, which can be a disadvantage in applications where interpretability is important, such as in healthcare or finance.

Prone to overfitting: Deep learning models can be prone to overfitting, which occurs when the model performs well on the training data but poorly on new, unseen data. This can be a problem when the model is used for prediction or decision-making in real-world applications.

Limited transferability: Deep learning models are often highly specialized to the specific task they were trained on, which can limit their transferability to other tasks or domains. This means that deep learning models may not generalize well to new, unseen data or tasks.

Limited interpretability: As mentioned earlier, deep learning models can be difficult to interpret, making it difficult to understand how the model is making its predictions. This can be a disadvantage in applications where explainability is important, such as in healthcare or finance.

Lack of transparency: Deep learning models can be opaque, meaning that it can be difficult to understand how they arrive at their decisions. This can be a disadvantage in applications where transparency is important, such as in legal or regulatory contexts.

Requires large amounts of training data: Deep learning models typically require large amounts of training data to perform well. This can be a disadvantage in applications where labeled data is scarce or difficult to obtain.

May require specialized hardware: As mentioned earlier, deep learning models can require specialized hardware, such as GPUs, to train effectively. This can be a disadvantage in applications where resources are limited or where the cost of specialized hardware is prohibitive. Overall, while deep learning is a powerful tool for solving complex problems, it also has several important limitations that should be taken into consideration when choosing a machine learning technique for a given problem.

Anomaly Detection Accuracy: Object detection models must accurately identify and classify various types of anomalies, including cracks, corrosion, dents.

CHAPTER 2

2. LITERATURE SURVEY

2.1 TITLE: LEAK DETECTION FOR NATURAL GAS GATHERING PIPELINES UNDER MULTIPLE OPERATING CONDITIONS USING RP-1DCONVLSTM-AE AND MULTIMODEL DECISION

AUTHOR: ZHONGLIN ZUO, HAO ZHANG, LI MA, TONG LIU AND SHAN LIANG

In this article, a novel leak detection method based on RP1dConvLSTM-AE and the multimodel decision was proposed for natural gas gathering pipelines under MOCs. First, anomaly point processing and long–short sequence construction were introduced, which provide reliable inputs for the detection model. Then, a feature learning network, RP-1dConvLSTM-AE, was developed for MTS of pipeline data. Consequently, the statistics and semantics representations of pipeline data were learned through reconstructing input sequences and predicting future sequences. The time dependencies and local features of MTS were captured using long–short sequences and ConvLSTM. In addition, a multimodel decision scheme with OCSVM was designed by building the health model for each operating condition, which effectively addresses pipeline leak detection under MOCs. Through empirical analyses, the proposed method achieved outstanding performance in leak detection under MOCs. In addition, the proposed method outperforms the compared semi-supervised leak detection methods. The PCA-SVDD and MSKS-GMM partially alleviate the problem of shallow SVDD and GMM lacking feature learning by constructing hand-crafted features. Different from them, the proposed method can automatically learn deep features from SCADA data by using the RP-1dConvLSTM-AE. Besides, the tnGAN-based method addressed data recovery and leak detection simultaneously via adversarial training. However, this method requires a small number of leaked samples to be involved in model training. Moreover, compared with the tnGAN, the proposed method develops the multimodel decision specifically for MOCs. Due to the unique advantages of ConvLSTM in processing spatiotemporal data, ConvLSTM-AE shows passable performance. However, pipeline leaks are reflected in not only statistical changes but semantic shifts. By comparison, the proposed RP-1dConvLSTM-AE pays more attention to the learning of semantic features through predicting future sequences.

2.2 TITLE: MACHINE-LEARNING-ASSISTED LEAK DETECTION USING DISTRIBUTED TEMPERATURE AND ACOUSTIC SENSORS

AUTHOR: HAYDEN GEMEINHARDT AND JYOTSNA SHARMA

There are over two million miles of pipelines in the world, with over two-thirds contained in the U.S. alone. Unfortunately, there have also been about 10 000 oil-and-gas pipeline failures to date in the U.S. since 2002, and over half of the pipelines in the U.S. are over 40 years old. Although pipelines are often seen as the safest and most economical method of transporting fluids pipelines age over time, causing corrosion and degradation that opens small cracks and holes. Small leaks are more difficult to detect using traditional methods, such as pressure gauges and flow meters, due to their limited sensitivity and capability to only measure at discrete locations, which is often insufficient to monitor long pipelines. However, even small leaks can accrue over time resulting in severe environmental contamination and ecological and economic damage. For example, in 2006, the North Slope in Alaska had a spill of over 200 000 gallons of oil which went undetected for five days, originating from a hole only a quarter inch in diameter. It was only found when a worker noticed a petroleum odor when driving along a deserted road. An aging pipeline infrastructure and numerous environmental disasters caused by pipeline leaks have spurred research in pipeline leak detection—both in enhancing existing sensor technology and in developing novel computational techniques for dissecting information from such sensors for creating reliable alerts. Small leaks, many of which often go undetected using conventional gauges, remain an urgent problem for an aging pipeline infrastructure. This article proposes a method that allows for an automated and robust leak detection and localization system by fusing distributed acoustic sensor (DAS) and distributed temperature sensor (DTS) data for machine learning. Distributed fiber-optic sensing creates an advantage over conventional gauges by providing real-time continuous measurements along the entire length of the fiber-instrumented pipeline at a high spatiotemporal resolution and sensitivity that can detect small leaks, as well as the leak location. High sensitivity, however, can create noisy data. Thus, machine learning is applied for a robust method of distinguishing nonleak data from leak signatures for accurate leak detection and localization. The workflow is demonstrated on an experimental pipeline setup exposed to environmental noise.

2.3 TITLE: A NOVEL CUSTOM ONE-DIMENSIONAL TIME-SERIES DENSENET FOR WATER PIPELINE LEAK DETECTION AND LOCALIZATION USING ACOUSTO-OPTIC SENSOR

AUTHOR: UMA RAJASEKARAN AND MOHANAPRASAD KOTHANADARAMAN

The leak detection and localization mechanism for identifying pipeline leaks is crucial in every structural health monitoring system. The state-of-the-art approaches gave high leak detection accuracy using machine learning and deep learning. However , to localize the leak, the only available mechanism is cross-correlation. The cross-correlation also gives good accuracy in localizing leaks but with two sensors and an appropriate noise removal method. So, the existing method's complexity is very high. This paper proposed a novel custom one-dimensional time-series DenseNet, a standalone leak detection and localization architecture with reduced system complexity without any noise removal technique and cross-correlation. This paper also implemented the existing one-dimensional DenseNet-121, three variations of 1DCNN, and cross-correlation to check the credibility of the proposed method. The proposed novel custom onedimensional time-series DenseNet outperforms all the other mechanisms. The number of parameters in the proposed novel custom one-dimensional time-series DenseNet is less compared with the existing one-dimensional DenseNet121. To be precise, the number of parameters in the proposed method is approximately one-quarter of the existing one-dimensional DenseNet-121. The training time of the proposed method is also only one-thirteenth of the existing one-dimensional DenseNet-121. The proposed method and the one-dimensional DenseNet were trained, validated, and predicted with seven classes at two different pressures and gave an average accuracy of 99.08%, 98.63%, and 98.34%, respectively. Moreover, the results demonstrated that the proposed novel custom one-dimensional timeseries DenseNet achieved more accurate leak detection and localization in a shorter time.

2.4 TITLE: VIDEO DETECTION OF SMALL LEAKS IN BURIED GAS PIPELINES

AUTHOR: YUXIN ZHAO, ZHONG SU

This paper proposes a video detection method for tiny leaks in buried gas pipelines in response to the difficult problem of tracking tiny leaks in buried gas pipelines. The method is based on the pipeline leakage dataset collected by the inspection robot in the buried gas pipeline, expanding the dataset, making the label corresponding to the dataset, and establishing the dataset. Introducing the BiFPN structure into the Neck layer in place of the original structure, enhancing the feature fusion ability of the model, and improving the precision of the model's leakage detection. Constructing the small-target detection layer based on the original structure of the Head layer, and adding a new small-target detection Head, to enhance the model's small leak detection ability. Send the data set into the proposed buried gas pipeline small leak video detection model, the model for parameter learning and network training. The collected buried gas pipeline small leak video into the trained model, the model's small leak video detection ability and small leak tracking ability to validate. Set up comparative experiments to prove that this paper's algorithms are better than the mainstream algorithms for target detection mainstream algorithms. Set up ablation experiments to prove that the higher evaluation index of this paper's method is due to the improvement of the YOLOv5s model. Set up a variety of leak detection experiments to prove that this paper's method can detect leaks of different shapes, leaks of different sizes, and leaks of different pipeline conditions, and it has a higher evaluation index and a better generalization ability. For the problem of difficult tracking of small leaks in buried gas pipelines, a video detection of small leaks in buried gas pipelines method is proposed for the detection robot inside buried gas pipelines. Firstly, collecting images and videos of leaks inside buried gas pipelines to establish a dataset.

2.5 TITLE: AN INTELLIGENT IOT AND ML-BASED WATER LEAKAGE DETECTION SYSTEM

AUTHOR: MOHAMMED REZWANUL ISLAM, SALIM ASAM, DEEPIKA MATHUR

This paper introduced a novel approach to leakage detection by developing a compact ML-based battery-powered edge IoT device. The FD-ANN model, segmented ML inference and low power-consuming IoT device promised a more accurate and efficient water leakage detection system. The FD-ANN model was 11 KB in size with 98.96% accuracy. The model was also resistant to external echo, environmental noise, and EMI. In an experimental case study, the model could detect leakage 100% of the time. Running an inference of a model in a desktop PC was easy as it had a lot of computational power and did not suffer from power constraints, but it was challenging in the case of an MCU. The segmented ML inference utilized computational resources efficiently and enhanced battery life, which was critical for remote IoT deployments. The proposed method took less than 4 KB of RAM to run the model on the MCU. This on-device ML inference enabled the possibility of continued monitoring of leakage. Here, we used low-cost SoC and open-source technologies to develop a leakage detection device that was power-efficient and fast compared to other devices. The previous ML-based leakage detection research worked in two ways: recording the signal and taking it to the lab for further processing. This off-site signal processing was not suitable for continuous leakage monitoring. The second way was to transmit raw data from the nodes and process it on a cloud service. Transmitting raw data was power-consuming and required regular battery changes. Our solution did the ML inference locally and only sent data to the central node when it detected a leakage. We developed the device so that it did not require cloud connectivity. The device only sent an alert message to the user if it detected any leakage.

2.6 TITLE: LEAK DETECTION IN WATER SUPPLY NETWORK USING A DATA-DRIVEN IMPROVED GRAPH CONVOLUTIONAL NETWORK

AUTHOR: SUISHENG CHEN, YUN WANG, WEI ZHANG AND HAIRONG ZHAN

Due to pipe aging and external stress changes, water leakage problem often occurs in water supply network, which results in additional energy and sources waste during water treatment and supply, and may cause other severe issues, such as bacterial pollution and poison contamination. Therefore, leakage detection for water supply network is of high significance and raises much attention. This paper proposes a leakage detection algorithm for water supply network based on IGCN model. Compared with traditional GCN methods, both measurement value information and data relationship information are well considered in the proposed algorithm. Moreover, due to its unique self-learning fully connected association graph, the proposed method outperforms several state-of-the-art methods in leakage detection rate. It is noticed that the water demand of each node in the water supply network changes over time, which is not considered in this paper. Therefore, the data autocorrelation will be discussed in our feature work. Due to the complex correlation within data collection, it is a challenging task to detect leakage in the water supply network. The Graph Convolutional Network (GCN) has recently gained significant attention in correlation research. However, most existing GCN-based models assumed that the topology of whole network should be derived from expert knowledge, which is always time-consuming and difficult to acquire. To tackle this problem, a data-driven improved graph convolutional network (IGCN) is proposed based on a self-learning fully connected association graph. Compared with traditional GCN, the proposed model can adaptively learn necessary data relationship information without accurate fixed undirected association graph. On this basis, a leakage-detection IGCN (LD-IGCN) is carried out for leakage detection.

2.7 TITLE: NDAMA: A NOVEL DEEP AUTOENCODER AND MULTIVARIATE ANALYSIS APPROACH FOR IOT-BASED METHANE GAS LEAKAGE DETECTION

AUTHOR: KHONGORZUL DASHDONDOV, MLHYE KIM AND KYURI JO

This study proposed a method consisting of three modules to predict gas leakage. Preparing efficient training data through data preprocessing and data labeling modules has dramatically improved the productive performance of machine learning algorithms. Using this method to create gas leakage data levels for air assessments in Korea is also possible. The environmental feature description of the target dataset. In other words, we used a DAE model to distinguish highly distorted parts from the raw dataset, and the AE model fits the more commonly distributed majority dataset to reconstruct them with a minor error. Therefore, outliers can be easily distinguished by the AE model. The data were normalized using OE transformations and k-means clustering, and the experimental data were ready. The DAE-OE-XGB model had the best results from constructing a predictive model using RF, KNN, XGB, DT, and NB algorithms on the prepared experimental dataset. According to the accuracy scores achieved after k = 10 cross-validation, the NDAMA_XGB model is accurate, which means it predicts the target variable accurately around 99.51% (95% CI, 99.39-99.63) of the time. The proposed NDAMA methods significantly improved the accuracies of the DAE-OE and other baseline methods. Finally, the log loss of the NDAMA_XGB model is 22.79, so it can be concluded that the model has high accuracy, low variability, and low log loss, so it is an effective model. As a consequence, the proposed framework emerged as the best predictive model, capable of significantly outperforming existing state-of-the-art baseline models.

2.8 TITLE: ANALYSIS OF THE PROPAGATION CHARACTERISTICS OF ACOUSTIC WAVES FROM LEAKAGES IN BURIED GAS PIPELINES

AUTHOR: SONG LIU1, ANQI LIU, ZEFANG CAI, CHJUNFENG SUNI

This paper proposed a new method to detect leakages at the buried gas pipelines in a timely and accurate manner. The diffusion of acoustic waves by the soil, the scattering of acoustic waves by particles in the soil and the absorption of acoustic energy by the soil medium are first analyzed to obtain the acoustic wave propagation characteristics of the buried gas pipelines leakages. A calculation model for acoustic wave propagation attenuation of buried pipelines is then established. In addition, the present work optimizes the ESMD denoising method to improve existing shortcomings to further reduce the disturbances of the acoustic signal, resulting in a relatively low noise level in the acoustic signal. Finally, field experiments were carried out to illustrate the performance of the model established in this paper with good accuracy and denoising capability, demonstrating a potential improvement of the leakage detection technology for buried pipelines. In order to accurately detect the leakages in buried gas pipelines and to reduce the leakage amount and false alarms, the propagation characteristics of acoustic waves owing to leakages in buried gas pipelines are analyzed. Firstly, the coupling effect of soil particles and gas, including the diffusion of acoustic waves by the soil, the scattering of acoustic waves by particles in the soil and the absorption of acoustic energy by the soil medium, is considered to establish a propagation attenuation model for acoustic waves resulted from leakages in buried gas pipelines. As acoustic waves are prone to the influence of noise in the process of propagation, an improved extreme-point symmetric mode decomposition (IESMD) acoustic signal denoising algorithm is proposed, which can effectively filter out the noise in the signal.

2.9 TITLE: INTELLIGENT URBAN SENSING FOR GAS LEAKAGE RISK ASSESSMENT

AUTHOR: TAO TAO, ZERONG DENG, ZHUO CHEN, LE CHEN, LIFENG CHEN

Urban underground infrastructure plays an important role in all modern cities. However, its service and maintenance cannot keep abreast of urban development in both developed and developing countries, which causes inconvenience to daily life, even poses a great threat to urban safety. Among them, the leakage of gas pipeline is one of the most destructive threats. In order to mitigate gas leakage damage, it is important to assess systematic risk of gas leakage in utility networks. In practice, leakage detection systems heavily rely on tedious and expensive human efforts, such as manually checking and assessing the risk, and thus only a very limited portion of the gas pipeline network can usually be assessed during a period of time. Therefore, it is critical to develop tools for automatically and systematically evaluating the whole gas pipeline network and revealing high-risk sites for checking in a timely fashion. To this end, in this paper, we develop an intelligent gas leakage risk assessment system based on the analysis of large-scale multi-source data, such as gas pipeline data, Point of Interests (POIs) data, and human mobility data. However, it is a non-trivial endeavor to design such a risk assessment system due to the unclear leakage mechanism, complex environmental conditions, and large size of the pipeline network. To address these challenges, both internal features of gas pipelines and external environmental features should be exploited. Specifically, we design a novel urban sensing technique to extract environmental features by analyzing human mobility data. Then, a joint learning neural network is developed to assess the leakage risk by integrating both internal and external features. Moreover, an intelligent risk assessment system is implemented and deployed for experiments in real-world scenarios.

2.10 TITLE: AN END-TO-END PLATFORM FOR MANAGING THIRD-PARTY RISKS IN OIL PIPELINES

AUTHOR: EDMUNDO CASCUS, LEO RAMOS, CHRISTIAN ROMERO, GONZALO ORELLANA

This paper addressed the critical issue of third-party risk detection near oil pipelines, emphasizing the need for an effective CV system to mitigate potential threats. We proposed a solution that combines the state-of-the-art YOLOv8x neural network model with a comprehensive platform, integrating cameras and alert systems. Our methodology involved the construction of a dataset tailored to aerial perspectives, capturing images in designated oil pipeline zones in Chile. The dataset, comprising 1,003 images across seven classes of objects, including buses, trucks, forklifts, machinery, pickups, tractors, and vehicles, served as the foundation for training and validating our models. The model underwent training and parameter optimization using the Ultralytics Tuner. This optimization resulted in the development of a model that achieves strong performance in metrics such as precision, recall, F1-score, and mAP. Furthermore, visual testing of the optimized model revealed substantial improvements in confidence scores and a notable reduction in false positives compared to the baseline model. These outcomes highlight the effectiveness of hyperparameter tuning in refining our model, ultimately leading to a more robust and reliable detection system. Additionally, a platform has been developed to integrate the model, which contains multiple functionalities. These functionalities allow users to view the alert history, track each alert according to its priority, record actions taken on each alert, visualize alerts on a geographical map, generate and send notifications for identified risks, and generate reports. Our focus is on demonstrating a real-world application of state-of-the-art CV models, like YOLOv8, in operational settings.

CHAPTER 3

3. PERFORMANCE COMPARISON

3.1 TABLE FOR LITERATURE SURVEY

TITLE	AUTHOR	DESCRIPTION	ALGORITHM	PARAMETERS	PERFORMANCE
Leak Detection for Natural Gas Gathering Pipelines Under Multiple Operating Conditions Using RP-1dConvLSTM-AE and Multimodel Decision	Zhonglin Zuo , Hao Zhang , Li Ma , Tong Liu ,and Shan Liang	A serious challenge to accurate leak detection. To address the above-mentioned problems, we propose a hybrid leak detection method	ConvLSTM	Accuracy Precision Recall FRR F1 Score	96.40% 96.64% 96.51% 3.73% 96.58%
Machine-Learning-Assisted Leak Detection Using Distributed Temperature and Acoustic Sensors	Hayden Gemeinhardt and Jyotsna Sharma	Proposes a method that allows for an automated and robust leak detection and localization system	Res-Net framework	Accuracy F1 SCORE TP FP TN FN Average	87% 0.93 82 13 5 0 0.18mm

A Novel Custom One-Dimensional Time-Series DenseNet for Water Pipeline Leak Detection and Localization Using Acousto-Optic Sensor	Umarajasekaran and Mohanaprasad kothanadaraman	To deploy a unified architecture capable of executing both the detection and localization of a leak	One-dimensional convolutional neural networks (1DCNN)	Precision Recall Specificity F1-score R-squared	95.97 99.47 96.87 97.69 98.01
Video Detection of Small Leaks in Buried Gas Pipelines	Yuxin Zhao, Zhong Su	Build a feature fusion network to enhance the model's ability to fuse small leaks	YOLOv5s (You Only Look Once)	Precision Recall Accuracy	94.1 94.8 94.5
An Intelligent IoT and ML-Based Water Leakage Detection System	Mohammed Islam, Salim Asam, Deepika Mathur	Proposed and developed an edge ML-based low-power IoT device to detect water leakage and notify the user	Artificial neural network	Accuracy Precision Recall F1-score	98.96% 98.99% 98% 98.5%
Leak Detection in Water Supply Network Using a Data-Driven Improved Graph Convolutional Network	Suisheng Chen,Yun Wang, Wei Zhang, Hairong Zhan, and Yuchen He	A data-driven improved graph convolutional network (IGCN) is proposed based on a self-learning fully connected association graph	One graph convolutional network (IGCN)	Training accuracy Testing accuracy	99.1% 98.1%

NDAMA: A Novel Deep Autoencoder and Multivariate Analysis Approach for IoT-Based Methane Gas Leakage Detection	Khongorrzul Dashdondov , Mlhye Kim , and Kyuri Jo	Propose a method based on deep learning to predict gas leakage from environmental data	A Novel Deep Autoencoder and Multivariate Analysis	Accuracy MSE F1 SCORE mIOU score	94.70 0.051 93.27 88.52
Analysis of the Propagation Characteristics of Acoustic Waves From Leakages in Buried Gas Pipelines	Song Liu, Anqiliu , Zefang Cai , Chunfeng Suni, and Ruochen Liu	Accurately detect the leakages in buried gas pipelines and to reduce the leakage amount and false alarms	Extreme-point symmetric mode Decomposition (IESMD) acoustic signal denoising algorithm	RMSE SNR	0.0263 25.761
Intelligent Urban Sensing for Gas Leakage Risk Assessment	Tao Tao, Zerong Deng, Zhuo Chen, Le Chen, Lifeng Chen	Develop an intelligent gas leakage risk assessment system based on the analysis of large-scale multi-source data	Deep neural network	Accuracy F1 score RMSE	0.777 0.763 0.401
An End-to-End Platform for Managing Third-Party Risks in Oil Pipelines	Edmundo Casas, Leo Ramos, Christian Romero, Gonzalo Orellana	practical application of cutting-edge computer vision models in real-world operational environments.	YOLO architecture	Precision Recall F1 score Map	0.781 0.718 0.748 0.757

TABLE 1. Existing System Comparison

3.2 COMPARISON TABLE

TITLE	ALGORITHM	COMPLEXITY	SCALABILITY	OVERALL PERFORMANCE
Leak Detection for Natural Gas Gathering Pipelines Under Multiple Operating Conditions Using RP-1dConvLSTM-AE and Multimodel Decision	ConvLSTM	High	Moderate	Excellent for multi-model data
Machine-Learning-Assisted Leak Detection Using Distributed Temperature and Acoustic Sensors	Res-Net Framework	Moderate	Moderate	Good for distributed sensor data
A Novel Custom One-Dimensional Time-Series DenseNet for Water Pipeline Leak Detection and Localization Using Acousto-Optic Sensor	One-dimensional convolutional neural networks (IDCNN)	Moderate	Moderate	Excellent for time-series data
Video Detection of Small Leaks in Buried Gas Pipelines	YOLOV5s (You Only Look Once)	Moderate	High	Excellent for video data
An Intelligent IoT and ML-Based Water Leakage Detection System	Artificial neural network	Low	High	Good for IoT data

Leak Detection in Water Supply Network Using a Data-Driven Improved Graph Convolutional Network	One graph convolutional network (IGCN)	High	High	Excellent for graph-structured data
NDAMA: A Novel Deep Autoencoder and Multivariate Analysis Approach for IoT-Based Methane Gas Leakage Detection	A Novel Deep Autoencoder and Multivariate Analysis	High	Moderate	Good for multivariate time-series data
Analysis of the Propagation Characteristics of Acoustic Waves From Leakages in Buried Gas Pipelines	Extreme-point symmetric mode decomposition (IESMD) acoustic signal denoising algorithm	Moderate	Moderate	Good for acoustic signal processing
Intelligent Urban Sensing for Gas Leakage Risk Assessment	Deep neural network	High	High	Excellent for multi-source data fusion
An End-to-End Platform for Managing Third-Party Risks in Oil Pipelines	YOLO architecture	Moderate	High	Excellent for real-time object detection

TABLE 2. Comparison for Literature Survey

CHAPTER 4

4. PROPOSED SYSTEM

4.1 EXISTING SYSTEM

Ultrasonic testing (UT), magnetic flux leakage (MFL), and inline inspection (ILI) tools are critical components of modern pipeline integrity assessment strategies, complementing manual inspections to provide detailed evaluations of pipeline conditions. UT, also known as ultrasonic inspection, utilizes high-frequency sound waves transmitted into the pipeline material. These sound waves travel through the material until they encounter a boundary between different mediums, such as the metal of the pipeline and any internal defects or anomalies. At these boundaries, the sound waves are reflected back to the UT sensor, allowing the system to detect and analyze internal defects such as cracks, corrosion, and wall thickness variations. UT provides precise measurements and detailed imaging of internal pipeline conditions, offering valuable insights into the integrity of the pipeline. Magnetic flux leakage (MFL) tools are specifically designed for the inspection of ferromagnetic pipelines, such as those made of steel. These tools utilize powerful magnets to create a magnetic field that saturates the pipeline material. When the pipeline is free of defects, the magnetic flux remains contained within the pipe. However, when there are anomalies such as corrosion or metal loss, the magnetic flux leaks out of the pipe and is detected by sensors in the MFL tool. Overall, UT, MFL, and ILI technologies play crucial roles in supplementing manual inspections and enhancing the accuracy and reliability of pipeline integrity assessments. By leveraging these advanced inspection tools, operators can identify potential issues early, prioritize maintenance activities, and mitigate the risk of pipeline failures, ultimately ensuring the safety, reliability, and longevity of pipeline infrastructure.

4.1.1 DISADVANTAGES

- Manual testing can be done
- Cost is high because of sensor implementation
- Time complexity can be high

4.2 PROPOSED SYSTEM

Pipeline leak detection using YOLO (You Only Look Once) entails harnessing this state-of-the-art object detection algorithm to pinpoint and locate leaks within pipeline infrastructure. The process begins with the collection of a dataset comprising images or videos featuring instances of pipeline leaks. These data are meticulously annotated with bounding boxes encompassing the leaks, serving as training material for the YOLO algorithm. Subsequently, the dataset undergoes preprocessing to ensure consistency and optimize compatibility with the YOLO model. Following this, the YOLO algorithm is trained on the annotated dataset, enabling it to recognize and accurately delineate leaks within pipeline imagery or video frames.

The training phase is iterative and involves adjusting model parameters to refine leak detection performance. Upon completion of training, the trained YOLO model undergoes evaluation to assess its accuracy and performance metrics. Validation ensures that the model can reliably detect leaks across a variety of conditions and scenarios, thereby instilling confidence in its effectiveness. Once validated, the trained YOLO model is ready for deployment in real-world settings for continuous leak detection.

This deployment phase may involve integration with cameras or sensors positioned along the pipeline to capture imagery or video feeds. In operation, the YOLO-based system continuously analyzes these feeds, promptly detecting and flagging instances of leaks as they occur. By leveraging YOLO for pipeline leak detection, operators can enhance the efficiency and effectiveness of their leak detection efforts, enabling timely intervention and mitigation to prevent environmental damage and operational disruptions.

4.2.1 ADVANTAGES

- Time complexity can be reduced
- Automated computer vision techniques
- Accurate leak detection from images

4.3 BLOCK DIAGRAM

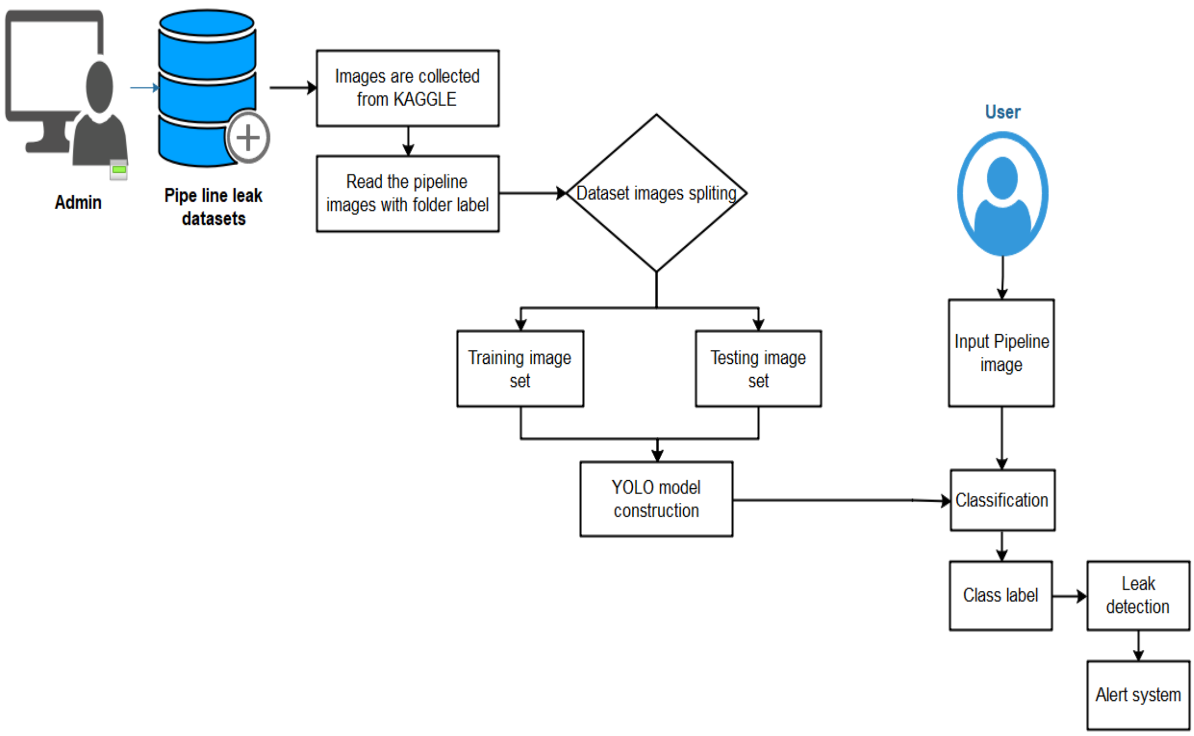

Figure 1. Block Diagram

4.3.1 Data Images Splitting

Data image splitting is a crucial step in machine learning, especially for computer vision tasks. It involves dividing a dataset into three subsets: training, validation, and testing. The training set is used to train the model, the validation set helps fine-tune hyperparameters, and the testing set evaluates the model's performance on unseen data. This process prevents overfitting, ensures reliable model evaluation, and enables optimal hyperparameter tuning. Common splitting techniques include random splitting, stratified splitting, and time-based splitting. Libraries like Scikit-learn, TensorFlow/Keras, and PyTorch provide tools for efficient image splitting and preprocessing

4.3.2 Training Image Set

A training image set is a collection of images used to train a machine learning model, typically for computer vision tasks. This dataset provides the model with examples of the patterns and features it needs to learn to make accurate predictions on new, unseen data. A good training image set is large, diverse, high-quality, accurately labeled, and often augmented to improve model performance. Popular datasets like ImageNet, CIFAR-10, MNIST, Pascal VOC, and COCO are commonly used for training various computer vision models.

4.3.3 Testing Image Set

A testing image set is a collection of images that a machine learning model has not seen during training. It is used to evaluate the model's performance on unseen data, providing an unbiased assessment of its generalization ability. This helps identify potential overfitting or underfitting issues and ensures the model's reliability in real-world scenarios. A good testing set should be representative of the real-world data distribution and should not be used during the training or validation phases to maintain its objectivity.

CHAPTER 5

5. CONCLUSION AND FUTURE ENHANCEMENT

5.1 CONCLUSION

In conclusion, the integration of YOLO (You Only Look Once) object detection technology into pipeline leak detection systems presents a significant advancement in ensuring the safety, reliability, and efficiency of pipeline infrastructure. By leveraging YOLO's real-time detection capabilities, operators can effectively identify and localize leaks within pipeline imagery or video feeds, enabling prompt intervention and mitigation efforts. The development of a YOLO-based alert system enhances response times by promptly notifying relevant personnel or systems upon detecting a leak, allowing for swift action to minimize potential environmental damage, operational disruptions, and safety hazards. Furthermore, the seamless integration of YOLO with control systems and automated shut-off valves streamlines emergency response measures, further enhancing the resilience of pipeline networks. Overall, the adoption of YOLO technology in pipeline leak detection not only improves detection accuracy but also contributes to proactive risk management, regulatory compliance, and sustainable infrastructure management practices. As advancements in object detection and artificial intelligence continue, YOLO-based systems are poised to play a crucial role in safeguarding pipeline assets and ensuring the safety and security of communities and environments they traverse.

5.2 FUTURE ENHANCEMENT

Continuously improving the accuracy and robustness of the YOLO model for leak detection by refining training strategies, incorporating additional data sources, and optimizing model architecture to handle diverse pipeline environments and conditions. Extending the YOLO model to detect and classify different types of pipeline anomalies beyond leaks, such as corrosion, dents, cracks, or intrusion events, enabling comprehensive pipeline integrity assessment in real-time.

REFERENCES

[1] Zuo, Zhonglin, et al. "Leak detection for natural gas gathering pipelines under multiple operating conditions using RP-1dConvLSTM-AE and multimodel decision." IEEE Transactions on Industrial Electronics (2023).

[2] Gemeinhardt, Hayden, and Jyotsna Sharma. "Machine Learning-Assisted Leak Detection using Distributed Temperature and Acoustic Sensors." IEEE Sensors Journal (2023).

[3] Rajasekaran, Uma, and Mohanaprasad Kothandaraman. "A Novel Custom One-Dimensional Time-Series DenseNet for Water Pipeline Leak Detection and Localization Using Acousto-Optic Sensor." IEEE Access (2024).

[4] Zhao, Yuxin, et al. "Video detection of small leaks in buried gas pipelines." IEEE Access (2023).

[5] Islam, Mohammed Rezwanul, et al. "An Intelligent IoT and ML-Based Water Leakage Detection System." IEEE Access (2023).

[6] Chen, Suisheng, et al. "Leak detection in water supply network Using a data-driven improved graph convolutional network." IEEE Access (2023).

[7] Dashdondov, Khongorzul, Mi-Hye Kim, and Kyuri Jo. "NDAMA: A novel deep autoencoder and multivariate analysis approach for IOT-based methane gas leakage detection." IEEE Access (2023).

[8] Liu, Song, et al. "Analysis of the Propagation Characteristics of Acoustic Waves from Leakages in Buried Gas Pipelines." IEEE Access (2023).

[9] Tao, Tao, et al. "Intelligent Urban Sensing for Gas Leakage Risk Assessment." IEEE Access 11 (2023): 37900-37910.

[10] Casas, Edmundo, et al. "An End-to-End Platform for Managing Third-party Risks in Oil Pipelines." IEEE Access (2024).