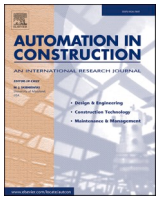
Automation in Construction 179 (2025) 106479

Contents lists available at ScienceDirect

Automation in Construction

journal homepage: www.elsevier.com/locate/autcon



Review

Sewer pipeline condition assessment and defect detection using computer vision

C. Long Nguyen [a], Andy Nguyen [a,*], Jason Brown [a], L. Minh Dang [b,c]

[a] School of Engineering, University of Southern Queensland, Springfield, QLD 4300, Australia
[b] Institute of Research and Development, Duy Tan University, Da Nang 550000, Viet Nam
[c] Faculty of Information Technology, Duy Tan University, Da Nang 550000, Viet Nam

ARTICLE INFO

Keywords:
Sewer pipeline
Computer vision
Defect inspections
Condition assessment
Severity assessment

ABSTRACT:

The structural integrity and operability of sewer pipeline systems are crucial for society's health, urban environment, and economic stability. Advancements in computer vision (CV) have revolutionized sewer defect inspection, offering unprecedented accuracy and efficiency in identifying and assessing pipeline failures. While prior reviews exist, they often lack systematic comparisons of models, detailed dataset analyses, or comprehensive severity assessment frameworks. This paper presents a comprehensive review of CV implementations for sewer defect detection, location, and characterization. It thoroughly evaluates main inspection techniques, diverse datasets, and key performance metrics. State-of-the-art CV models and their applications are critically reviewed, alongside defect severity assessments and their link to maintenance strategies. Key challenges and limitations are identified, leading to recommendations for enhancing inspection efficiency and accuracy. The paper consolidates findings on methodological trends, data analysis advancements, algorithm performance variations, and improved severity assessment approaches.

1. Introduction

Sewer pipelines are critical elements of urban infrastructure, tasked with transporting wastewater from residences, businesses, and industrial sites to treatment facilities. These systems are vital for public health, preventing drinking water contamination and limiting the spread of waterborne diseases. They also play a crucial role in environmental protection by ensuring wastewater is adequately treated before being released into natural water bodies. Economically, efficient sewer systems support sustainable urban growth by enabling safe residential, commercial, and industrial activities.

However, many sewer systems are aging and deteriorating, leading to frequent blockages, collapses, and overflows. These problems can result in significant public health risks, environmental pollution, and substantial economic costs due to emergency repairs and service disruptions. It was reported that, more than $100 million was needed for a four-year program since 2018 to upgrade ageing sewer pipes and maintenance holes across Melbourne, Victoria [1]. Therefore, regular maintenance helps pipeline asset owners identify and repair defects early, thereby avoiding costly and disruptive failures and highlighting the importance of their maintenance and management.

Defects and failures in sewer pipeline systems can be broadly categorised into two main types: operational and structural. Operational issues encompass defects directly impacting the pipeline's functionality, such as root intrusion, blockages, infiltration, sediment accumulation, and deposits [2]. Structural problems, on the other hand, involve failures that compromise the system's structural integrity, including fractures, cracks, deformation, collapses, corrosion, and joint displacement [3].

Historically, sewer inspections have predominantly relied on manual methods conducted by skilled experts. These methods were labour-intensive, time-consuming, and often hazardous, requiring inspectors to enter the sewer system with specialised tools to assess pipeline conditions [4]. Additionally, inaccessible areas posed challenges, limiting the thoroughness of inspections and increasing the potential for errors. As a result, manual inspection of sewer pipelines was challenging, inefficient, and prone to errors. To address these issues, it is crucial to enhance the methods of sewer inspection and monitoring with less intervention from humans. Developing and testing technologies that automate sewer condition assessment are key to realising these improvements.

With the development of sensor technologies, closed-circuit

* Corresponding author.
E-mail address: andy.nguyen@unisq.edu.au (A. Nguyen).

https://doi.org/10.1016/j.autcon.2025.106479
Received 28 January 2025; Received in revised form 12 August 2025; Accepted 15 August 2025
Available online 20 August 2025

television (CCTV) systems have been widely applied for sewer inspection since the last century [5,6] with high-resolution optical sensors, such as cameras. This inspection system obtains images or video footage from a robot with a mounted camera and lighting system. The robot is controlled and navigated by an operator at the control unit on the ground via cable, which is also used for transmitting power supply and image information of the sewer (Fig. 1). The recorded image or video recordings of the sewer are then visually inspected in-field or transferred to the office for offline condition assessment by inspectors. This approach has a few disadvantages, including being slow, expensive and prone to human errors. To address these problems, improved inspection techniques with less interventions from humans and more automation capacities must be developed.

These days, machine Learning (ML) and its specialised subfield, Deep Learning (DL), are transformative technologies that empower computers to learn from data and make data-driven conclusions or predictions. With the advantages of processing and analysing large quantities of complex data, these technologies have been applied in various fields to support humans without being explicitly programmed. When applied to visual data, these advanced learning algorithms manifest as Computer Vision (CV) techniques, which have been widely recognised by researchers and engineers as a key component of an improved inspection procedure for various structures including sewer pipelines [7].

Researchers have utilised CV techniques to improve methods for inspecting sewer defects in pipeline systems [8,9]. For instance, a high-efficiency object detection method for locating sewer defects was proposed to classify and locate two sewer categories faster and safer than traditional sewer inspections [9]. Furthermore, it enables real-time CCTV inspection, eliminating the delays and errors common in manual inspection while safely accessing hazardous or confined spaces through robotics. Despite significant advancements in the application of CV, several critical challenges remain unresolved such as the lack of standardised and publicly available datasets, which are essential for both practising and benchmarking purposes. Furthermore, there is a noticeable gap in connecting defect detection outputs with actionable maintenance strategies, such as assessing defect severity or integrating predictive maintenance systems. Addressing these gaps is essential to advancing the field and ensuring more practical and reliable sewer pipeline inspection practices.

Several review papers have endeavoured to capture these progresses. Published in 2019, the analysis by Moradi et al. [3] is an early and structured review of sewer inspection, offering a foundational understanding of inspection modalities and automation efforts. However, the review lacks detailed analysis of deep learning models, benchmark datasets, or severity assessment frameworks. About a year later, Rayhana et al. [10] shifted the survey focus toward robotic inspection platforms, providing a rich taxonomy of mobile systems and their sensing capacities. While this review is valuable for understanding the hardware landscape, it did not delve into the algorithmic aspects of defect detection or condition assessment, nor did it evaluate datasets or model performance metrics. In 2022, Li et al. [11] broadened the review scope, especially on the traditional and deep learning approaches. Although informative, the discussion remained largely descriptive and lacked systematic comparisons of different model architectures, as well as detailed analysis of dataset characteristics and severity rating frameworks, limiting its utility for benchmarking and reproducibility. The 2023 review paper by Sun et al. [12] focused more directly on the practical integration of deep learning with CCTV-based sewer inspection, particularly in the context of urban water management. However, it did not provide comprehensive benchmarking of models and datasets, nor did it extensively explore severity assessment methods. Similar limitations are evident in the survey by Haurum and Moeslund [13], which, despite its technical depth and historical breadth, also omits severity assessment frameworks and standardised benchmarking across datasets.

To address the limitations identified in prior reviews and assist newcomers to the field, this paper presents a comprehensive and technically rigorous synthesis of the state-of-the-art literature on vision-based sewer defect detection and condition assessment. It offers foundation and systematically evaluates CV algorithms—including latest model architectures—across multiple tasks such as classification, detection, and segmentation. It also provides a structured taxonomy of publicly available datasets, detailing their characteristics, annotation standards, and usage in benchmarking. Importantly, the paper presents a dedicated analysis of severity assessment frameworks as well as establishes essential links between the assessment outcome and the decision-making process. By integrating performance metrics, dataset comparisons, and architectural insights, the paper aims to serve as a valuable reference for both researchers and practitioners. For newcomers to the field, the paper offers a clear, reproducible, and up-to-date roadmap to help them in navigating the rather sophisticated landscape of automated sewer inspection.

2. General concepts of computer vision and sewer defect inspection

The utilisation of CV has been widespread in sewer inspection for improving the speed, accuracy and safety for assessing underground pipelines. The advances of vision sensor technology, image processing, and automated analysis now allow defects to be detected, classified and measured with greater precision than traditional manual methods. This section first outlines how CV has developed in structural inspection, with a focus on sewer systems, and then describes how these techniques are applied in complete inspection frameworks that combine data collection, defect detection, and condition assessment to guide maintenance planning.

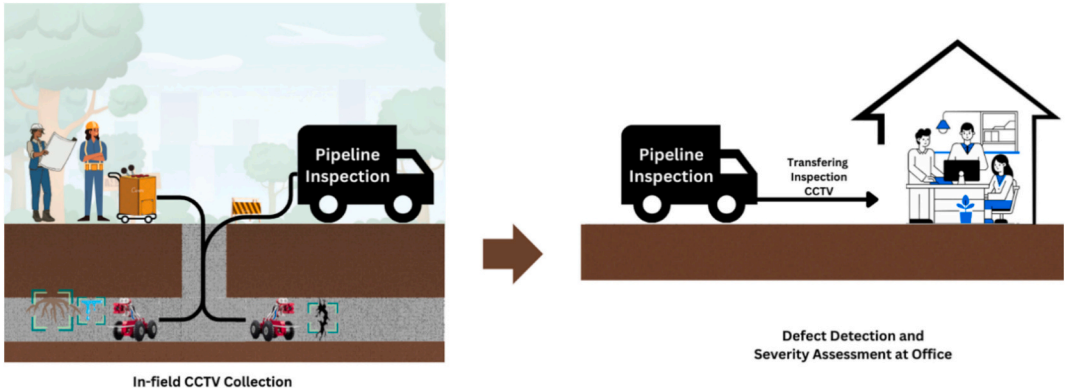


Fig. 1. CCTV-based sewer pipeline inspection with control unit and remote defect inspection.

2.1. Computer vision in structural inspection

Since its inception in the 1960s, CV has been recognised and applied in the field of infrastructure condition assessment, becoming a critical element for automated sewer inspection and monitoring [7]. CV has improved the inspection method by enabling more efficient, accurate and safer assessment processes. High-resolution cameras and advanced image processing algorithms can capture and extract detailed images and videos of the interior of sewer pipelines. These visuals are then analysed to detect and classify defects such as cracks, blockages, and root intrusions, often with greater precision than human inspectors. By automating the inspection process, CV not only reduces the need for human intervention but also enhances the spatial resolution of the inspections, leading to more effective maintenance and repair strategies [7,8,14].

Additionally, CV technology can be integrated with unmanned ground vehicles, allowing for rapid and thorough inspection of extensive sewer networks, thus ensuring the longevity and reliability of critical infrastructure. However, the performance of CCTV sewer inspection method heavily depends on the quality of dataset [15]. Factors, such as camera resolution, lighting conditions, focal length, zoom capabilities, camera stability, and environmental conditions, cause issues in image quality. Consequently, image processing algorithms are required to remove potential noise and enhance the dataset quality. Due to the typical environment of the pipeline systems, which often lack light and have high moisture levels, there are studies that applied pre-processing to their dataset images before conducting the condition assessments [16–18].

During the last two decades, advancements in CV techniques have been driven by ML to identify sewer defects. In the early stage, numerous studies have utilised traditional ML approaches to detect these defects [3,19–21]. By leveraging labelled datasets, these ML models can be trained to recognise and classify various types of defects. Over time, these models improve their detection capability through continuous learning, leading to more reliable and efficient sewer maintenance. A fault detection method using unsupervised ML algorithm was proposed by Xu et al. for anomaly detection in sewer pipeline visual inspection [19]. Similarly, Gedam et al. utilised a linear regression approach for predicting sewer pipe main conditions and applied ML in forecasting sewer pipe conditions [22].

More recently, DL methods have been utilised for sewer inspection to handle more complex and larger datasets, automatically extract features and improve performance. DL techniques such as Convolutional Neural Networks (CNNs) have applied to estimate the water level in sewer pipes showcasing the potential in automating the inspection processes [23]. Yin et al. proposed a CNN-based object detection to automate defect detection in real time, leveraging the advantage of DL algorithms [24]. These technologies not only enhance the detection and classification of sewer defects but also enable predictive maintenance by identifying potential issues before they escalate into major problems. As a result, sewer pipeline inspection has been transformed into a more proactive and data-driven field, ensuring the longevity and reliability of critical infrastructure.

In addition to basic defect inspection, several studies have applied CV techniques, enhanced by DL methods, widely in condition analysis or defect grading for sewer pipelines in recent years [11,19,25,26]. This analysis and grading process is crucial for monitoring the structural integrity and functionality of underground infrastructures. For instance, Wang & Cheng used semantic segmentation with deep dilated CNN for the automatic severity assessment of sewer pipe defects [25]. The integration of dilated convolution and multiscale techniques with recurrent neural network layers has been proposed for the severity assessment of sewer pipeline faults by Xu et al. in 2020 [19]. However, the most recent defect severity assessment mainly focuses on crack or fracture defects, with limited studies addressing other sewer pipeline fault categories.

2.2. CV-assisted sewer defect inspection system

A sewer defect inspection framework with the support of CV involves several stages that contribute to the efficient assessment of sewer pipelines (as shown in Fig. 2). The framework starts with the data acquisition stage, which utilises a crawler robot to enter the pipeline and capture images and videos using sensing techniques, such as CCTV and Light Detection and Ranging (LiDAR) laser [27]. It is followed by transferring data to the head unit for pre-processing images and the augmentation stage of extracting frames from videos, enhancing the quality of images and expanding the dataset. Training detection model for sewer defects is then conducted based on the processed dataset for classifying, detecting and segmenting employing several image processing algorithms and convolutional neural networks. Next, the models are refined and validated to maximise their performance. To examine the severity of defects, object measurement steps are done to obtain the characteristics and dimensions of defects in the post-processing stages. Finally, the framework exports faulty evidence from sewer pipelines as well as a risk assessment to support decision-making in inspectors' maintenance.

The training stage for defect inspection is often the most challenging and resource-intensive part of the framework, encompassing tasks such as image classification, object detection, and object segmentation (Fig. 3). Detecting defects by image classification is the first and fundamental step, where entire images are categorised into predefined classes based on specific rules or ML algorithms. On the other hand, locating defect by object detection is the higher level, which involves identifying and localising multiple patterns using bounding boxes and confidence scores for its defect (Fig. 3). Object detection can be done using one-stage or two-stage algorithms, which will be discussed in the following section. Characterising defects by image segmentation is the most precise task diving images into regions corresponding to predefined labels. The paper discusses three segmentation approaches: morphological, semantic and instance segmentation. Fig. 4 summaries the techniques and algorithms used for each type of sewer defect inspection.

Following the defect inspection stage, condition assessment is conducted to evaluate the structural and operational state of sewer pipelines. This process is critical for maintaining the pipeline integrity and ensuring public safety. It requires several types of defect information from the inspection stage, such as defect identification (e.g. cracks, fractures, roots), characteristics (e.g. length, width, depth, area, orientation), and location. Using this data, the computers analyse the extent and severity of the defect and generating risk assessment reports with corresponding defect grades. This stage plays a vital role in supporting engineers and inspectors in making informed decisions about maintenance and repair priorities based on the potential risks to pipeline operation.

3. Benchmark datasets and tools

Effective sewer defect inspection depends on both the quality of the datasets used for training and the method applied to evaluate model performance. Publicly available datasets vary widely in size, resolution, and defect categories, making dataset benchmarking an important step in comparing inspection algorithms. Preprocessing and augmentation techniques are often required to address issues such as poor lighting, noise, and limited diversity in the data, ensuring models are better able to generalise to real inspection scenarios. Finally, performance metrics tailer to classification, detection, and segmentation tasks provide a standard basis for assessing accuracy, robustness and reliability. This section reviews commonly used sewer inspection datasets, outline preprocessing and augmentation approaches, and summaries the key evaluation metrics used in the field.

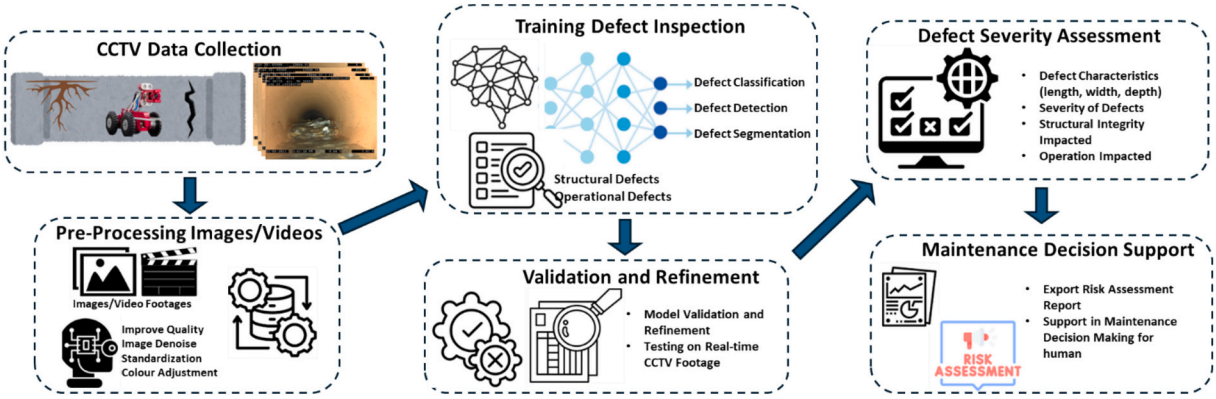


Fig. 2. Framework of sewer defect inspection.

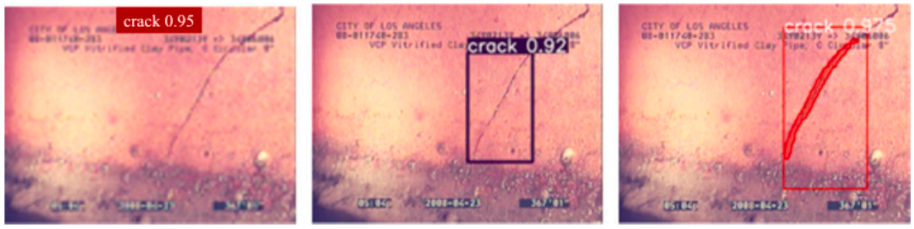


Fig. 3. Example of sewer pipeline crack inspection using image classification (left), object detection (middle) and image segmentation (right) [28].

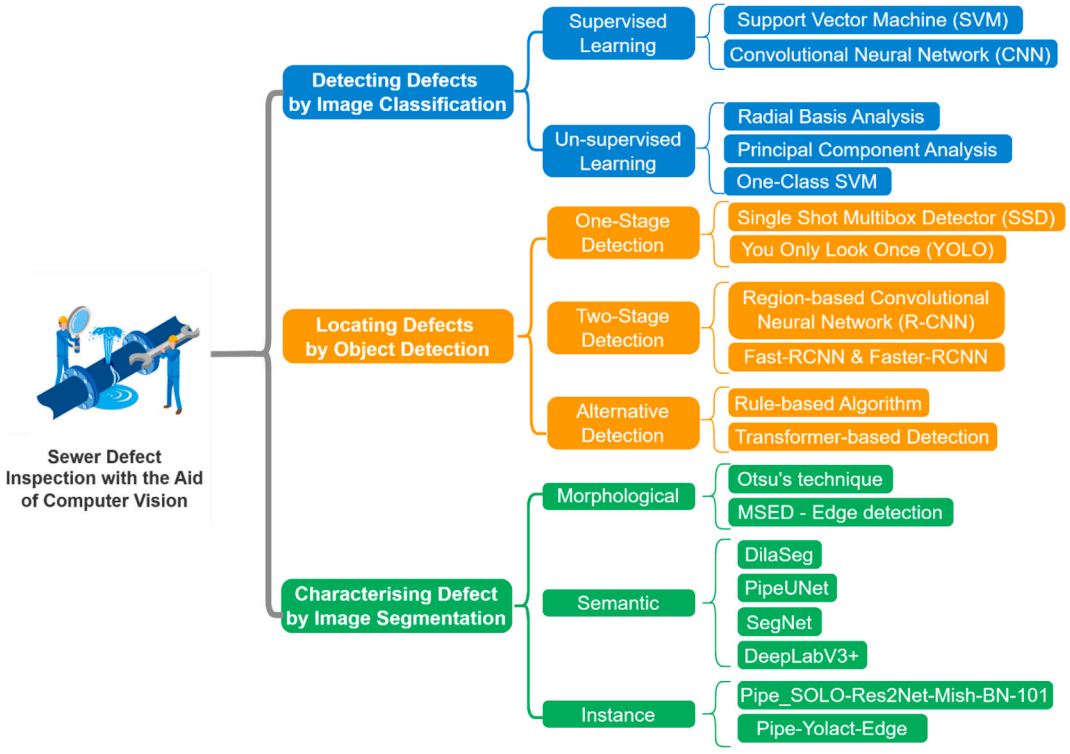


Fig. 4. Hierarchy chart of sewer defect inspection with the aid of computer vision.

3.1. Datasets benchmarking

To compare the performance of sewer defect inspection algorithms, it is essential to benchmark the dataset and evaluate its characteristics such as quality, diversity or robustness for model development and evaluation. Sewer-ML is a widely-used public dataset, with more than 1.3 million images and 18 predefined classes [29]. It was collected by three different Danish companies from 2011 to 2019 and pre-processed by experts before being publicly released online with a wide range of image resolutions from 350x284 to 768x576 pixels. Another large-scale dataset was introduced by Meijer et al. [30] comprising approximately 2.2 million images with 12 different defects, captured at 1040x1040

resolution using 185° angle camera. However, this dataset is unannotated and requires manual labelling, which is labour-intensive. Liu et al. [31] proposed two different video-based datasets, QV-Pipe and CCTV-Pipe, for defect inspection by image classification and object detection, respectively. QV-Pipe was collected by a pole-mounted camera from maintenance holes for rapid anomaly assessment, while CCTV-Pipe involved robotic crawlers capturing high-resolution footage along the pipeline system. Table 1 summarises the available datasets for defects in sewer pipelines from previous studies.

3.2. Image preprocessing and augmentation algorithms

The challenging operating condition of sewer systems, particularly low visibility, poor lighting, harsh and unpredictable conditions, significantly impact the training and performance of object detection. Poor illumination often results in low-quality and noisy datasets where defect like small cracks or subtle root intrusions are difficult to discern. Additionally, uneven lighting caused by water reflection and over-exposed area make an inconsistency in the image that hinders the model's ability to generalise across datasets. These lighting challenges can lead to domain shifts, reducing the model's accuracy when applied to different sewer environments. These issues can be mitigated by the application of image preprocessing techniques and data augmentation algorithms.

Table 1
Benchmark datasets for sewer pipelines defects from previous studies

Ref.	Names/ Authors	Types of Defects	No. of Images (I) /Videos (V)	Resolution
[31]	QV-Pipe	17 defect classes (undisclosed class names)	9601 (V)	Not Specified
[31]	CCTV-Pipe	16 defect classes (undisclosed class names)	575 (V)	Not Specified
[29]	Sewer-ML	Water levels, crack, breaks, collapse, surface damage, production error, and 11 additional defect classes	1,300,201 (I)	From 350x284 to 768x576 pixels
[16]	Hassan et al.	Defect longitudinal, debris silty, joint faulty, joint open, lateral protruding, and surface damage	24,137 (I)	256x256 pixels
[32]	Xie et al.	Normal, deposition, stagger, fracture, high water level, disjunction, and additional defect classes	42,800 (I)	Not Specified
[30]	Meijer et al.	Fissure, surface damage, intruding and defective connection, intruding sealing material, displaced joint, porous pipe, and 5 additional defect classes	2,202,582 (I)	1040x1040 pixels
[33]	Kumar et al.	Root intrusion, deposits, cracks, infiltration, debris, and 3 additional defect classes	12,000 (I)	From 320x256 to 1440x720 pixels
[34]	Li et al.	Deposits settlement, joint offset, broken, obstacles, water level stag, deformation	18,433 (I)	From 296x166 to 1435x1054 pixels
[35]	Dang et al.	Crack, debris silty, faulty and open joint, protruding lateral, surface damage, and pipe broken	38,386 (I)	1280x720 pixels
[36]	Cheng et al.	Root, crack, infiltration, and deposit	1,260 (I)	From 320x256 to 1440x720 pixels
[37]	Chen et al.	Normal, blur, intrusion, deposit, and obstacle	10,000 (I)	Not Specified
[38]	Wang et al.	Crack, deposit and root	1,885 (I)	512x256 pixel

Image preprocessing transforms the original image into a new format suitable for CV models. This process includes several tasks, such as modifying the geometry, colour, noise reduction/filtering, and/or normalisation of the image. The geometry of the original image size is typically adjusted through resizing, cropping and scaling to meet the model input requirements. Colour adjustment algorithms, including normalisation, standardisation, noise reduction, can be applied to images with the main target of minimising colour variation and enhancing image quality. Normalisation [39] plays a vital role in scaling pixel values to a standard range, such as [0, 1] or [-1, 1], to improve convergence of CV algorithms and ensure consistency in the dataset. On the other hand, standardisation [40] adjusts pixel values to have zero mean and unit variance, reducing bias caused by various lighting conditions and enhancing training stability. Additionally, noise reduction and filtering help remove unwanted disturbances while retaining essential features, like edges and texture.

In contrast to preprocessing image algorithms, image augmentation is a technique used in CV to artificially expand the size and diversity of a training dataset by applying a variety of transformations to existing images. These transformations include geometric changes (e.g. rotation, flipping, scaling, cropping, translation), photometric adjustments (altering brightness, contrast, saturation, and hue) and noise introduction (adding Gaussian noise, blur, or perspective distortion). With the expanded dataset, the risk of overfitting can be reduced, while the robustness of the model is improved. An example of image augmentation algorithms is shown in Fig. 5. The key difference between image preprocessing and image augmentation lies in their purpose and application. Image preprocessing is applied to the entire dataset and thus its effects are reflected across all subsets, training, validation and testing. On the contrary, image augmentation is only applied to the training subset and is specially aimed at improving the generalisation ability of the model by exposing it to a wider range of simulated scenarios and environmental conditions.

3.3. Performance metrics

To evaluate the model performance in sewer pipeline inspection, various quantitative metrics are used to ensure reliable and accurate defect detection, against various challenges including the presence of imbalanced datasets [41]. Different tasks—such as image classification, object detection, and segmentation—require specific metrics that highlight different aspects of model effectiveness. This section reviews the key metrics such as Precision, Recall, Accuracy, F1 score, Intersection over Union (IoU) [42], Average Precision (AP), Pixel Accuracy (PA), and Area Under the Receiver Operating Characteristic curve (AUC-ROC) [43,44]. Tools such as confusion matrix, Receiver Operating Characteristic curve (ROC) [45], and Precision-Recall Curve (PRC) [41], used for obtaining performance metrics are also discussed.

The confusion matrix is a key tool for evaluating performance in ML, showing counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), as illustrated in Fig. 6. In sewer pipeline inspection, it helps assess model accuracy in defect identification. For example, in the crack class, true positives occur when both prediction and ground truth indicate a crack, while true negatives reflect agreement on non-crack defects. False positives arise when a crack is incorrectly predicted, and false negatives when an actual crack is missed.

3.3.1. Metrics for classification tasks

Precision and recall are essential metrics for evaluating a classification performance, particularly with imbalanced datasets. As shown in Fig. 6, precision measures the accuracy of positive predictions and is calculated based on elements of the confusion matrix, whereas recall, or sensitivity, measures the model's ability to identify all actual positives,

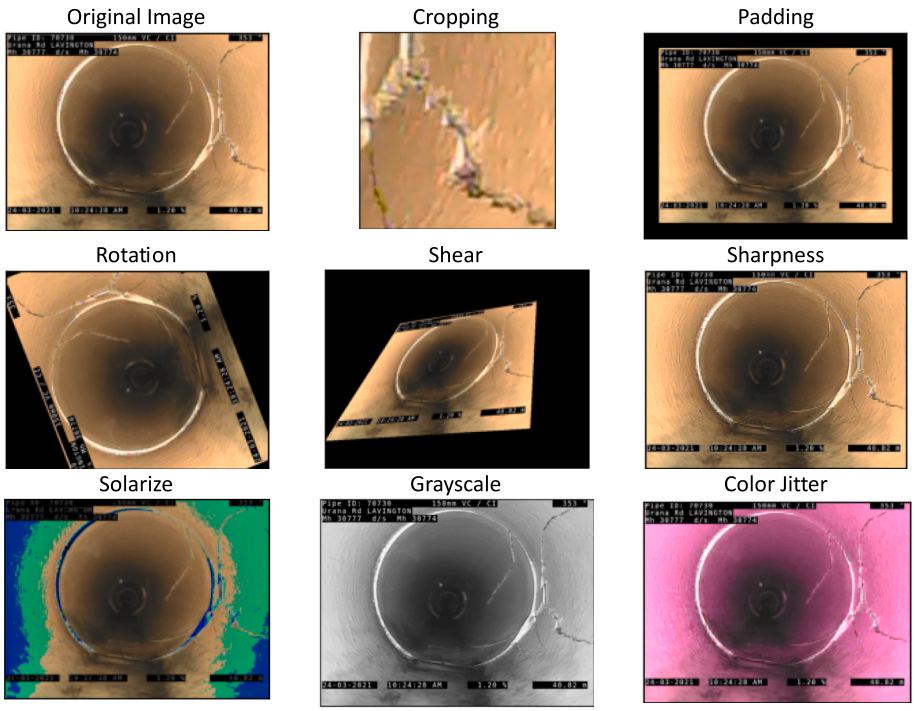


Fig. 5. Examples of different augmentation algorithms

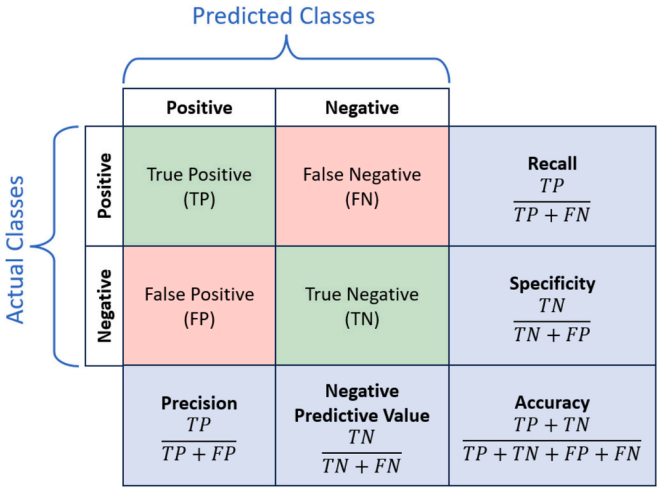


Fig. 6. Confusion matrix tools to obtain performance metrics

calculated as the ratio of true positives to the total actual positives.

The F1 score, the harmonic mean of precision and recall, provides a balanced measure of the model performance, especially when both precision and recall are equally important. As shown in Eq. (1), F1 score is calculated based on the proportion of actual positives relative to all predicted and actual positive labels. Ranging from 0 (poor) to 1 (excellent), a higher F1 score indicates a better balance between precision and recall.

$$F1\ Score = 2^{*}\left(\frac{Precision^{*}Recall}{Precision + Recall}\right) \tag{1}$$

The ROC curve and its associated metric, AUC-ROC, are common tools for evaluating binary classifiers. The ROC curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR) across thresholds. TPR, or recall, measures the proportion of correctly identified positives, while FPR quantifies the proportion of false alarms among actual negatives. AUC-ROC quantifies the overall discriminatory ability of models, with values closer to 1 indicating superior performance.

$$TPR = \frac{TP}{TP + FN} \tag{2}$$

$$FPR = \frac{FP}{FP + TN} \tag{3}$$

For multi-class classification problems, performance metrics can be extended by employing macro- and micro-averaging strategies. Micro-averaging calculates metrics globally by summing the true positives, false positives, and false negatives across all classes, giving more weight to larger classes. In contrast, macro-averaging computes the metric for each class independently and then averages these per-class results, treating all classes equally regardless of their size, which is particularly useful for imbalanced datasets common in defect detection. These strategies also apply to ROC and AUC-ROC evaluations, where One-vs-Rest (OvR) is used to compute a binary ROC curve per class. The macro-AUC is then obtained by averaging the per-class AUCs, while micro-AUC considers all predictions collectively. Sokolova and Lapalme [46] reviewed averaging strategies for precision, recall, and F1-score in multiclass, multilabel, and hierarchical classification settings. Their work provides a clear framework for selecting appropriate evaluation metrics, making it a valuable reference for newcomers to classification tasks.

3.3.2. *Metrics for object detection tasks*

For object detection tasks, Intersection over Union (IoU), also known as Jaccard Index, measures the overlap section between the predicted and ground truth labels. It is calculated as the ratio of the intersection (the overlap area) to the union (combined area) of the predicted and actual segments. As shown in Equation (4), an IoU closer to 1 indicates a better match.

$$IoU = \frac{\sum_{j=1}^{k} n_{jj}}{\sum_{j=1}^{k}\left(n_{ij} + n_{ji} + n_{jj}\right)} \tag{4}$$

where n_{jj} is the number of pixels correctly classified as a class j. n_{ij} is

the number of pixels, which are labelled as class i, but classified as class j. Similarly, n_{ji} is the total number of pixels labelled as class j, but classified as class i. In multiclass problems, mean Intersection over Union (mIoU) is widely used to evaluate performance across all classes [47]. It is computed by averaging the IoU values for each class, providing an overall measure of segmentation quality regardless of class imbalance.

AP is a widely used metric in object detection that summarises the Precision–Recall (PR) curve into a single scalar value. It corresponds to the area under the PR curve, where precision is plotted against recall across varying confidence thresholds. Higher AP values indicate better detection performance. For multiclass object detection, the standard evaluation metric is mean AP (mAP), calculated by averaging the AP scores across all object classes. This provides a comprehensive measure of overall detection performance across all defect types.

3.3.3. Metrics for segmentation tasks

PA is commonly used for semantic segmentation to measure the percentage of correctly classified pixels in the entire image. It is computed as the ratio of correctly predicted pixels to the total number of pixels, as shown in Eq. (5). While PA offers a general measure of performance, it may be less informative for imbalanced classes, where smaller classes may be underrepresented. To address this limitation, mean PA (mPA) is used, which computes the pixel accuracy separately for each class and then averages the results [47], as shown in Eq. (6). This ensures that all classes contribute equally to the final score, making it more reliable for evaluating performance in datasets with class imbalance.

$$PA = \frac{\sum_{j=1}^{k} n_{jj}}{\sum_{j=1}^{k} t_j} \tag{5}$$

$$mPA = \frac{1}{k}\sum_{j=1}^{k}\frac{n_{jj}}{t_j} \tag{6}$$

where n_{jj} is the number of pixels correctly classified as a class j and t_j is the total number of pixels labelled as class j. In term of the confusion matrix's four elements, n_{jj} corresponds to the true positive for class j and PA can also be calculated by the same equation for accuracy in Fig. 6.

Similar to object detection tasks, segmentation tasks also use Intersection over Union (IoU) and mean IoU (mIoU) as key evaluation metrics. These metrics can be further extended to frequency-weighted IoU (FwIoU), which incorporates the relative frequency of each class in the dataset. By weighting each class's IoU by the number of ground truth pixels it contains, FwIoU provides a more representative evaluation when the dataset contains classes with significantly different pixel counts, such as background versus rare defect types in sewer inspection tasks [47]. FwIoU can be calculated by multiplying the IoU for each class by the number of pixels in that class and then summing them up as Eq. (7).

$$FwIoU = \frac{1}{\sum_{j=1}^{k} t_j}\sum_{j=1}^{k} t_j \frac{n_{jj}}{n_{ij} + n_{ji} + n_{jj}} \tag{7}$$

4. Detecting defects by image classification

Detecting sewer defects by image classification is considered as the first level of defect inspection. The main purpose of this method is to identify and categorise the image under a specific label with the support of advanced technologies such as CV and learning algorithms. Classification algorithms are commonly divided into three categories, including unsupervised, semi-supervised and supervised learning algorithms.

4.1. Supervised learning algorithms

Supervised classification algorithms are trained with a labelled dataset categorise the image to a predefined label. Using image-label pairs, the models learn to recognise and remember the patterns and features of sewer defects. Common algorithms include Support Vector Machines (SVMs), which find optimal boundaries to separate defect categories; and Convolutional Neural Networks (CNNs), which excel at recognising patterns in visual data. Defect classification is typically approached as either binary (defect vs. no defect) or multi-class (e.g., crack, root, deposit, blockage).

The SVM is a popular supervised learning approach for image classification, particularly from late 1990s to early 2010s. This algorithm works by finding an optimal hyperplane that separates classes in the feature space, maximising the margin between them [48]. The hyperplane's form varies with the number of features or classes, ranging from a line to a multi-dimensional plane.

Ye et al. [49] applied SVM algorithm to classify 1045 CCTV images from seven sewer defect types using four feature extraction methods, Daubechies (DBn) wavelet transform, Hu invariant moment, lateral Fourier transform and texture features. The SVM algorithm achieved an average accuracy of 0.84 for all types of defects, with 0.99 for settled. Yang et al. [50] compared SVM with Radial Basis Network (RBN) and Back-Propagation Neural Network (BPN) and two experiments with four defect patterns showed SVM achieving an accuracy of 60%. Zou et al. [51] used SVM to classify three crack types—longitudinal, circumferential and multiple cracks—and integrated a histogram of oriented gradients (HOG) for feature extraction. Their model demonstrated a robust performance, with accuracy exceeding 90%.

With the advent of DL, Convolutional Neural Networks (CNNs) were developed from Artificial Neural Networks (ANN) and have become some of the most powerful and widely-used algorithms in CV tasks [52]. CNNs are designed to automatically learn hierarchies of features from raw image information through backpropagation. Their ability to handle complex visual data and support end-to-end learning has led to successful applications in fields such as medical imaging, agriculture, and structural inspection. Due to their high accuracy and robust performance. Due to superb performance, CNNs have become a standard tool for image classification and object detection.

Chen et al. [37] proposed a CNN-based sewer defect detection system combining SqueezeNet [53] and InceptionV3 [54]. After augmentation, SqueezeNet first identified abnormal images, which were then passed to InceptionV3, to detect deposition, obstacles, blur and intrusions. This two-stage model outperformed SVM in detection accuracy and handled natural pipeline scenes effectively. Similarly, Li et al. [34] introduced a hierarchical CNN model using ResNet18 for feature extraction, tested on an imbalanced dataset. It first classified images as defective or not, then further categorised defects into one of the total seven types, achieving a defect detection accuracy of 83.2%. In contrast, Kumar et al. [33] developed a binary CNN classification framework (Fig. 7) to categorise three defect types—crack, deposit and root intrusions—using a large sewer defect dataset of 12,000 images. The framework uses binary classification stages, where the number of stages depends on the number of defect types. Each stage produces two sub-datasets, refining the classification process. The highest accuracy achieved by this framework was 86.2%.

4.2. Unsupervised and Semi-supervised Learning Algorithms

Unsupervised and semi-supervised algorithms, unlike supervised algorithms, utilise the unlabelled or partially labelled dataset for feeding the models, respectively. Unsupervised learning focuses on detecting the object patterns and groups that all similar instances share together into a class. This method eliminates the requirement of a human in the labelling process, which is affected negatively by several factors (e.g. tiredness). Semi-supervised algorithms combine supervised and

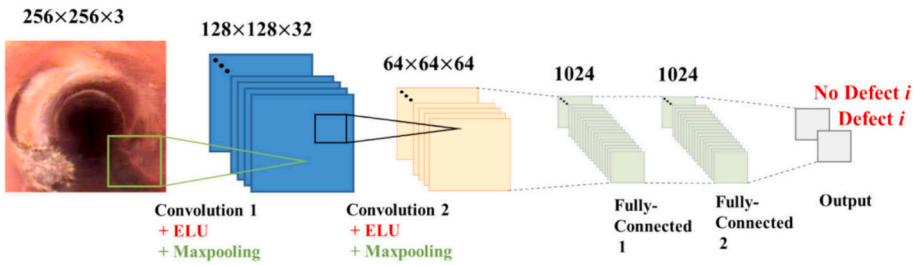


Fig. 7. Proposed CNN for binary classification of Kumar et al. [33].

unsupervised learning by learning from a labelled dataset and exposure to unlabelled information. Yang et al. [50] examined the performance of a RBN-based classification model, which combined both supervised and unsupervised algorithms for classifying pipe defect patterns. The RBN transforms the input image information through a hidden layer with a Gaussian activation function. Evaluated against SVM and BPN, RBN stood out for its exceptional computation efficiency and robust classification accuracy, particularly in the class of broken pipe defects. While not achieving the best overall accuracy, RBN, with its combination of supervised and unsupervised algorithms, provided a highly effective and efficient solution for pipe defect classification.

A few other unsupervised learning algorithms were also used in sewer defect detection contexts. For instance, Principal Component Analysis (PCA) an algorithm well-known for dimensionality reduction of image data to facilitate exploration. Meijer et al. [55] introduced a three-part PCA-based framework for sewer image anomaly detection, responsible for feature decomposition and partial reconstruction. The trained model achieved an AUC-ROC of 0.946 on a smooth dataset and 0.714 on coarse one. Fang et al. [19] found that the performance of the One-Class Support Vector Machine (OC-SVM) algorithm in sewer pipeline fault detection was consistently lower than that of other anomaly detection methods evaluated in the study. Across multiple datasets and feature combinations, OC-SVM demonstrated limited effectiveness, particularly in handling noisy data and complex fault patterns, with the

Table 2
Previous studies about detecting sewer defect by image classification

Ref.	Year	Algorithm/Model	Dataset Size	No. of Classes	Performance	Comments
Supervised Learning						
[50]	2008	SVM	291 Images	4 Classes	Average accuracy = 0.6	- Able to classify structural defects - Low classification accuracy
[56]	2018	Multiclass Random Forest	2,424 Images	11 Classes	Overall accuracy = 0.71	- Able to classify defects in CCTV videos - Applicable for real-time inspection
[37]	2018	CNN – SqueezeNet + Inceptionv3	10,000 Images	4 Classes	AUC-ROC = 0.93 Average accuracy = 0.81	- Detected only obvious feature defects
[33]	2018	Binary CNN	12,000 Images	3 Classes	Average accuracy = 0.86	- Image resolution from 320x256 to 1440x720 pixels - Unable to classify sub-pattern of main defect
[32]	2019	Binary and Multiclass CNN	42,800 Images	6 Classes	Average accuracy = 0.95 for both binary and multiclass	- High classification accuracy - Applicable for real-time CCTV videos - High classification accuracy for multiple classes
[34]	2019	Multiclass CNN – Resnet18	18,333 Images	7 Classes	Average accuracy = 0.83	- Image resolution from 296x166 to 1435x1054 pixels - Improper labels and imbalanced dataset - Two levels classification with high- and low-level categories
[49]	2019	SVM	1,045 Images	7 Classes	Average accuracy = 0.84	- Difficulty in detecting collapse and joint damage - Low classification accuracy for structural defects
[51]	2020	SVM	1,001 Images	3 Classes	Recall ≈ 0.9	- Image resolution of 320x240 pixels - Able to classify crack sub-category in sewers - High inference speed and recall, applicable only to crack pattern
[58]	2020	CNN	800 Images	2 Classes	Accuracy = 0.47 – 0.96	- Low resolution and noisy images, imbalanced dataset
[35]	2021	CNN – Fine-tuned VGG19	38,386 Images	8 Classes	Accuracy = 0.98	- Fine-tuned 19-layers CNN delivers - High classification accuracy - Robust in noisy and highly imbalanced dataset.
[59]	2023	CNN – RegNet+	12,000 Images	20 Classes	F1-score = 0.98 Accuracy = 0.98	- Image resolution of 1280x720 pixels - High accuracy across wide range of defects
Unsupervised and Semi-supervised Learning						
[50]	2008	RBN	291 Images	4 Classes	Overall accuracy = 0.54	- Low classification accuracy with short computation time
[55]	2010	PCA	684 Images - Smooth 698 Images - Coarse	N/A	AUC-ROC = 0.946 – Smooth AUC-ROC = 0.714 – Coarse	- Able to classify only smooth and coarse images - Unable to detect sewer defects.
[57]	2018	OC-SVM	7,842 Images	14 Classes	Accuracy = 0.75 – Images Accuracy = 0.85 - Video AUC-ROC = 0.76	- Applies to both images and video for classification
[19]	2020	OC-SVM	~11,000 Images	N/A	Accuracy = 0.471 (lowest)	- Poor and highly varied performance

lowest accuracy only at 0.471.

4.3. Discussion

Table 2 summarises several detecting sewer defects by image classification studies from 2008 until now with many algorithms and techniques. Before 2018, most papers [19,50,56,57] utilised traditional ML algorithms (such as SVM, RBN and Random Forest) to classify sewer defects with low training accuracy, ranging from 0.54 to 0.75. The accuracy metric was improved significantly up to 0.9 when the classification model shifted towards more complex models such as multiclass Random Forest and various CNN architectures. The low accuracy of traditional ML models such as SVM and Random Forest can be explained by their reliance on handcrafted feature extractions. These manually engineered features often struggle to capture the complex and varied nature of sewer defects, particularly in low-quality CCTV datasets. Moreover, this dependence on manual feature design limits the scalability of such models, making it challenging to adapt them to large-scale and diverse datasets.

CNN models have become dominant in recent research with their superior ability to handle large and diversified datasets. For instance, the VGG19 model, with the 19-layer deep network, was fine-tuned by Dang et al. [35] to effectively capture intricate features of more than 38,000 sewer defect images and achieve a near-perfect validation accuracy of 0.98. This depth allows the model to learn complex, hierarchical feature representations, leading to high classification accuracy. Similarly, newer architectures such as RegNet+ have demonstrated comparable accuracy (up to 0.987) while addressing more diverse classification tasks involving up to 20 defect classes, making them suitable for real-time sewer inspection applications.

In terms of learning types, DL models with supervised learning often outperform unsupervised and semi-supervised learning algorithms because they leverage labelled datasets to learn from specific examples of defects. Besides accuracy, AUC-ROC was also used to evaluate the performance of classification models when dealing with imbalanced datasets in several studies [37,55,58]. CNN models that combine architectures such as SqueezeNet and InceptionV3 achieve high AUC-ROC values of up to 0.93, indicating strong effectiveness in distinguishing between defect classes. In contrast, unsupervised models like OC-SVM and PCA are simpler algorithms and do not benefit from labelled training data. As a result, they are less effective in handling fine-grained defect patterns (such as cracks or roots) and low-quality images, which was reflected in their lower AUC-ROC scores of 0.76 and 0.714, respectively.

Besides the development of model architecture, researchers also paid attention to enhancing quality and expanding datasets—from just 291 images in an early study [50] to 42,000 images in a more recent one [32]. Additionally, the number of sewer defect classes was increased from 2 (binary classification) to 20 classes, allowing for more fine-grained categorisation. These enhancements have significantly contributed to improved model performance and generalisability. For example, the CNN–VGG19 model [35], trained on a dataset of 38,386 images across eight classes, achieved near perfect validation accuracy of 0.98. Similarly, another CNN-based study [59], trained on a 12,000 image dataset with 20 defect classes, reported an accuracy of 0.987.

For new researchers seeking foundational studies in sewer defect detection by image classification, it is suggested to review the binary and multiclass CNN models, proposed by Xie et al. in 2019 [32], which focused on automatic detection and classification of sewer defects. These models were developed using a hierarchical DL approach to learn features progressively from general to specific, starting with binary classification and advancing to detailed categorisation. Additionally, the models were trained and validated on a large dataset containing 42,800 sewer pipeline images across six different classes, offering a rather comprehensive detection application. Another key study to consider is the fine-tuned VGG19 CNN model by Dang et al. [35], which stands out as one of the highest-performing models achieving an impressive accuracy of 0.98. Based on the well-known VGG19 architecture, the study used another large dataset with over 38,000 images across 8 classes.

5. Locating defects by object detection

While image classification is concerned with assigning a single label to an entire image, object detection advances further by identifying and localising multiple objects within an image, thereby enabling the defects to be located based on pre-defined classes. The detection outputs include bounding boxes, the class labels and confidence scores. Recent progresses in DL has rapidly advanced object detection, making it a major research focus. Applications span various fields, including security cameras [60–62], self-driving vehicles [63–65], pest detection [66–68], and healthcare [69–71]. Object detection is complex due to the need for both categorisation and localisation, leading to the development of one-stage and two-stage detection architectures. One-stage detection models, such as Single-Shot Multibox Detection (SSD) [72], You Only Look Once (YOLO) [73], RetinaNet [74], and CornerNet [75] perform classification and localisation in a single step, offering faster inference suitable for real-time use. Two-stage models like Region-based Convolutional Neural Network (R-CNN) [76], Fast R-CNN [77] and Faster R-CNN [78] first generate region proposals, then refine and classify them, achieving higher accuracy but at slower speed. This section reviews most notable studies in both approaches.

5.1. Single Shot Multibox Detector (SSD)

In 2016, Liu et al. [72] introduce introduced SSD, as a new object detection method using a single deep neural network. This version of SSD employed VGG-16 as the backbone for initial feature extraction (as Fig. 8), followed by auxiliary convolutional layers that capture multi-scale features through progressive downsample. This structure enables efficient detection of objects with varying sizes and aspect ratios. Predictions are made at the detection layers, returning bounding boxes and confidence scores, refined using Non-Maximum Suppression (NMS). T SSD's inference speed and performance have been validated on several public datasets, such as PASCAL VOC and COCO, showing strong results.

In sewer pipeline inspection, Kumar et al. [9] applied SSD to detect defects in 3,800 CCTV images, covering eight defect types with resolutions ranging from 720x576 to 1,507x720. The SSD model has been compared with other DL models, including YOLOv3 and Faster-RCNN. To improve detection speed, the authors replaced the original VGG-16 backbone with MobileNet, following the approach by Howard et al. [79], which slightly reduced accuracy. The modified SSD model achieved a detection speed of 33 ms and a mean average precision (mAP) of 54.4%, both lower than the other models. Similarly, Wang et al. [80] examined SSD on a smaller dataset, with image processing applied before training, and also found its performance inferior, reinforcing the findings by Kumar et al. [9].

In a 2023 study, Shen et al. [81] proposed an improved object detection algorithm for sewer pipeline inspection called Enhanced Feature Extraction SSD (EFE-SSD), targeting four defect types. Built on the original SSD with a VGG-16 backbone, the model integrates a Receptive Field Block (RFB) to improve feature extraction, an enhanced ECA attention mechanism to adjust channel weights, and replaces the cross-entropy loss with Focal Loss to address class imbalance during training. EFE-SSD achieved a mean average precision (mAP) of 92.2%, outperforming several state-of-the-art models, including Faster R-CNN, YOLO, and RetinaNet.

5.2. You Only Look Once (YOLO)

Similar to SSD, YOLO is an object detection algorithm that frames detection as a single regression problem, straight from image pixels to bounding box coordinates and class probabilities just by a single pass

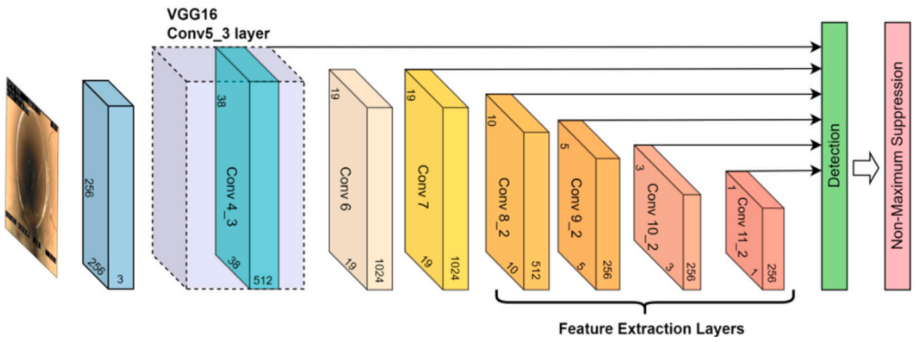


Fig. 8. Single shot multibox detector architecture.

[73]. It is a pioneering DL framework designed with an outstanding efficiency and speed in detecting objects within images and videos [82]. YOLO accomplishes this by dividing the media information into a grid and predicting bounding boxes and class probabilities directly without the need for region proposals and post-processing steps. Over time, YOLO has evolved through eleven versions, each introducing enhancements to improve performance and showcasing the model's iterative progression and continuous refinement [83,84]. Several variants have also been developed, such as YOLO-LITE optimised for non-GPU device [85] and YOLOv3-SPP with the addition of a Spatial Pyramid Pooling (SPP) layer to improve the detection of objects at different scales [86]. Among the many versions, YOLOv2, YOLOv3, and YOLOv5 are most commonly used in sewer pipeline inspection due to their robust performance and adaptability to diverse defect types.

5.2.1. YOLOv2

YOLOv2 is a fully convolutional network that processes input images and reduces them to an output grid with a resolution that is 32 times smaller than the original input resolution [87]. The model uses Darknet-19 as its backbone, a deep convolutional neural network consisting of 19 convolutional layers and five max-pooling layers. YOLOv2 utilises anchor boxes for improved bounding box prediction and applies batch normalisation to all convolutional layers to enhance training stability and performance.

Zhou et al. [88] evaluated YOLOv2 for automated locating sewer defects, comparing its accuracy and inference speed with the Faster R-CNN model, using the same dataset, system, and hyperparameters. Despite stable training times as dataset size increased from 10% to 100%, YOLOv2 achieved lower accuracy, highlighting the trade-off between detection accuracy and computational efficiency. In another study, Situ et al. [89] applied transfer learning to YOLOv2 by replacing the Darknet-19 (Fig. 9) by eleven pre-trained CNNs. The study [89] found that InceptionV3 delivered the best performance with a mean average precision (mAP) of 0.71, while InceptionResNetV2 yielded the lowest precision and speed. The author concluded that CNNs with fewer convolutional layers and parameters achieved better performance, highlighting the importance of selecting appropriate feature extraction models for an early version of YOLO.

5.2.2. YOLOv3

YOLOv3 represents a significant evolution from its predecessors, introducing notable improvements in object detection capabilities. Unlike YOLOv2, it adopts a residual network architecture based on Darknet-53 [90]. This increases depth and complexity through 53 convolutional layers and residual connections (often referred to as cross-layer summation). A key enhancement is the integration of a multi-scale prediction mechanism, similar in principle to a Feature Pyramid Network (FPN), enabling predictions at three different scales (as show in Fig. 10). This is achieved by combining semantically rich, low-resolution features from deeper layers with high-resolution, detailed ones from shallower layers through upsampling and concatenation [91]. YOLOv3 also employs a total of nine pre-determined anchor boxes, which are strategically assigned to these three different prediction feature maps based on their sizes, improving detection accuracy for large, medium, and small objects.

Based on YOLOv3, several studies have proposed automated sewer pipeline object detection for use with CCTV images and video footage [9,24,92]. Kumar et al. [92] have proposed a YOLOv3 framework with the integration of 5 CNN layers before the detection stage for classification. A dataset with 1800 images were sorted out for the training of locating defect model using the YOLOv3 model, yielding a high average precision of 71% with an IoU threshold of 0.2. In another study [9], Kumar et al. employed the same YOLOv3 model without adding 5 CNN layers for classification. The authors used an extension dataset of 3800 images and achieved a higher mean average precision of 75.2% on the validation dataset with the same IoU threshold. Furthermore, a range

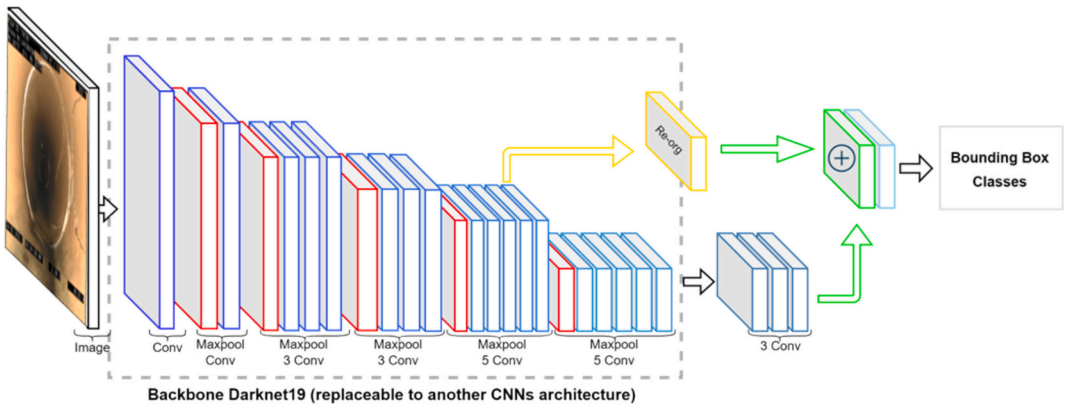


Fig. 9. Architecture of YOLOv2.

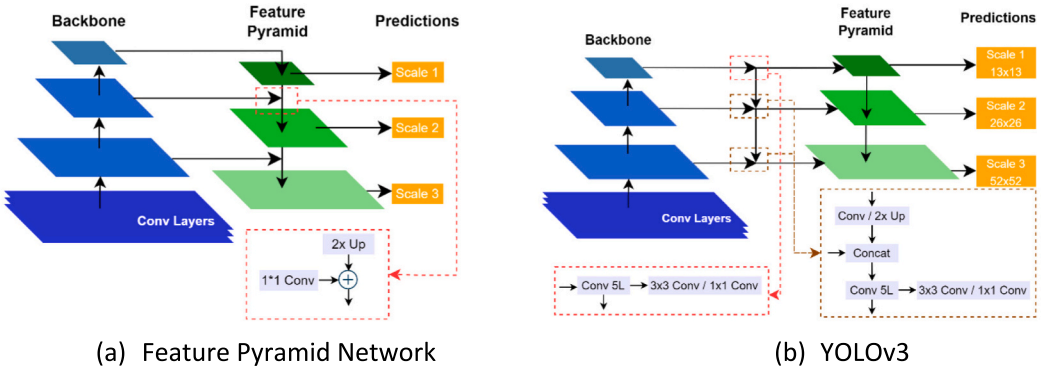


Fig. 10. Feature pyramid network and YOLOv3 architectures.

value of IoU thresholds, from 0.2 to 0.5, were also experimented using the same models and found that the AP of the YOLOv3 model decreased dramatically from 70.8% to 58.9% for root defect class. In comparison, Yin et al. [24] applied the original YOLOv3 model and achieved a higher mAP of 85.37% across seven sewer defects and an inference speed of 33 frames per second (or 30 ms/image) for real-time applications. Later, Tan et al. [28] modified YOLOv3 method by improving the loss function to Generalised Intersection Over Union (GIoU) and incorporating mosaic augmentation and refined bounding box prediction, achieving faster training convergence and peak mAP of 92% at an IoU threshold of 0.6. The improved YOLOv3 also demonstrated an efficient inference speed of 5.7 ms/image.

5.2.3. YOLOv5

YOLOv5, introduced in 2020 by Ultralytics, significantly advanced the YOLO series through an end-to-end solution [95]. This version is known for its lightweight architectures and ease of implementation, with configurations from small to extra-large models. The model employs a robust Darknet-based feature extraction network, integrating combined elements of the Cross Stage Partial Path Aggregation network (CSP-PAN) [93,94] to form CSP-Darknet53 backbone. CSP-PAN improves gradient flow and reduces computational complexity by dividing and merging feature maps through a cross-stage hierarchy. Additionally, Spatial Pyramid Pooling Fast (SPPF) aids the model in handling varying input image sizes and scales [95]. YOLOv5 also incorporates several modern techniques, such as auto-learning bounding box anchors, mosaic data augmentation, and hyperparameter evolution, boosting object detection precision.

Zekuan et al. [96] trained and tested the original YOLOv5 with 4660 images across five sewer defects. It achieved an mAP of 0.87 at an IoU threshold of 0.5, though this dropped to 0.69 when the threshold varied from 0.5 to 0.9. The authors also proposed a modified YOLOv5 incorporating multiple attention mechanisms into the backbone and replacing the neck network by a Weighted Bi-directional Feature Pyramid Network (BiFPN) [97]. The modified model has shown its advantages by increasing mAP to 0.88 at an IoU threshold of 0.5, and achieving an mAP of 0.72 across IoU thresholds ranging from 0.5 to 0.9. In another study [98], Situ et al. developed a real-time YOLOv5 model, employing transfer learning and channel pruning techniques to reduce computational complexity and memory usage. Channel pruning reduced the model parameters by 81% and operations by 48.8%. Despite these reductions, the mAP for sewer defect inspection remained high at 92.3% for the small YOLOv5 and 91.8% for the pruned small YOLOv5 models. Additionally, the inference speed significantly improved, decreasing from 7.9 to 5.2 ms/image.

5.2.4. YOLOv7 to YOLOv11

With the new architectural innovations, YOLOv7 marked a significant milestone in the YOLO family by enabling faster inference and higher accuracy than earlier versions. It utilised an extended Efficient Layer Aggregation Network (ELAN) to restructure the computational block and optimise the learning ability without increasing computational resources [99]. This version also incorporates model scaling, re-parameterisation techniques and auxiliary head learning to enhance performance in real-time object detection tasks. YOLOv8, developed by Ultralytics in 2023, introduced a complete redesign with a focus on modularity, anchor-free detection, and an updated loss function. It also offers cutting-edge accuracy and detection speed with new features and optimisations suitable for various detection tasks. In 2024, several performance tests were conducted for these YOLO models or their modified architectures (e.g. lightweight versions) using sewer datasets [100–102].

YOLOv9 built upon these foundations by integrating advanced backbone architectures and introducing two key innovations – Programmable Gradient Information (PGI) and Generalised Efficient Layer Aggregation Network (GELAN) – to improve training effectiveness and reduce information loss across network layers [103]. PGI preserves input information for more reliable gradient computation, while GELAN provides a lightweight yet powerful architecture using only standard convolutions. Building on this, YOLOv10 addresses shortcomings in pre-processing and model architecture while enhancing performance for edge deployment by significantly reducing the model size and computational cost without sacrificing accuracy. YOLOv11 continues this evolution by introducing a more efficient architecture with improved feature extraction and attention mechanisms such as Cross Stage Partial with Kernel Size 2 (C3K2), Spatial Pyramid Feature Fusion (SPFF) and Cross-Stage Partial Self-Attention (C2PSA). It integrates dynamics computation strategies and improved quantisation support, making it well-suited for real-time applications on resource-constrained hardware while maintaining competitive detection accuracy. Additionally, YOLOv11 achieves higher accuracy with fewer parameters and supports a broad range of tasks across diverse deployment environments.

5.3. Two-stage detection

Advancements in this area were mainly built upon the foundational R-CNN model by Ross et al. [76]. R-CNN initially utilised CNNs for detection via selective search for region proposal and feature extraction, but its multi-stage pipeline was computationally inefficient. Fast R-CNN improved this by introducing Region of Interest (RoI) pooling and integrating classification and regression into a single network [77]. Faster R-CNN improved upon Fast R-CNN by incorporating a Region Proposal Network (RPN) directly into the Fast R-CNN framework [78], as shown in Fig. 11. As a result, it can significantly accelerate the detection process and enable end-to-end training. This integration enables end-to-end training and significantly accelerates the detection process. Input images are first passed through a backbone CNN to extract features and generate a feature map. The RPN then slides a small

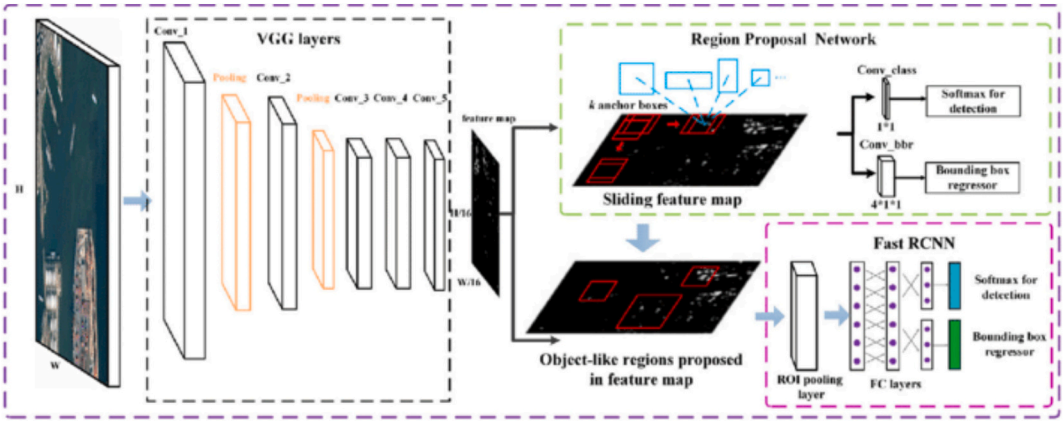


Fig. 11. Architecture of faster R-CNN [78].

network over this map to generate region proposals and objectness scores, before the proposals are refined through fully connected layers for classification and localisation. This two-stage architecture allows Faster R-CNN to achieve high accuracy, making it a preferred choice over single-stage detectors like SSD or YOLO when accuracy is more critical than speed.

Faster R-CNN was evaluated in a few comparative studies including [9,88]. In [9], this model was trained with 3420 images and compared with YOLOv3 and SSD achieving the highest accuracy with $mAP_{0.2}$ of 76.2% on validation and 71.8% with testing dataset. Zhou et al. [88] used a smaller dataset of 610 images covering five sewer defect types and found that Faster R-CNN achieved outperformed YOLOv2 in accuracy, though with slower detection speed. Another study [36] explored the impact of dataset size, architecture, and hyperparameters on performance. A modified Faster R-CNN trained on a four-defect dataset from CCTV videos achieved an mAP of 83%, a 20% improvement over the original setup, with an inference speed of 9.434 frames per second (or 106 ms/image).

5.4. Alternative algorithms

Before the advent of CNN for object detection, manual image processing algorithms —such as feature extraction and rule-based logic—were used to detect and classify objects. These rule-based algorithms relied on predefined criteria rather than learning from data, involving steps like noise reduction, thresholding, and edge detection for segmentation. While simple and interpretable, they lacked scalability and adaptability. In 2014, Halfawy et al. [104] developed a rule-based model to identify root intrusion in sewer pipes, achieving 86% accuracy. Another study [105] proposed an eight-step method to detect sewer cracks on low-resolution images, using threshold optimisation by image processing and crack detection.

In 2020, Facebook introduced Detection Transformer (DETR) [106] a novel object detection model that rely on complex anchor mechanisms and post-processing steps, this model leverages the power of the transformer from common natural language processing (NLP). The DETR utilises the CNN network as a backbone for the encoder section to process the feature maps, capturing global context and relationships across the entire image. On the other hand, its decoder used learned object queries to attend to these encoded features, iteratively refining prediction for bounding boxes and class labels. With the appearance of DETR, Dang et al. [107] have modified and developed a DefectTR to localise and classify sewer defects on 47,100 images covering 10 defect types. The authors have found that the original DETR model achieved a mAP of 56.2% with an inference time of 83 ms/image, while their modified DefectTR model achieved a mAP and detection speed of 60.2% and 85 ms/image, respectively.

5.5. Discussion

Table 3 summaries of previous studies and research about locating defect by object detection in sewer pipeline systems. The reviewed papers and research have been generated in the last decade using two main methods, which are traditional and DL algorithms. In comparison to DL methods, traditional locating defect algorithms [104,105,108] from 2014 to 2019 achieved lower detection accuracy (0.84 – 0.89) with slower speeds of one frame per second (or 1000 ms/image). In contrast, DL models (SSD, YOLO, Faster R-CNN and transformer) from 2018 onward illustrated significant improvements in both detection accuracy (consistent mAP score above 0.9) and speed (average of 58.3 ms/image), which can be explained by the ability of deep neural networks to automatically learn complex features, as opposed to manual features used in traditional methods. Mainly, the convolutional layers in DL are utilised to identify spatial features in the images for better feature extraction.

Kumar et al. paper [9] in 2018 found that the Faster R-CNN model achieved higher accuracy than one-stage detection models (SSD and YOLO) by 22%. This is because Faster R-CNN first generates a region proposal, then refines its proposal and classifies it into objects. On the other hand, a one-stage model makes prediction of object location and classification directly in one stage, which sacrifice prediction accuracy for inference speed. However, over time, the modified one-stage detection Sewer-YOLO-Slim model [81,100] in 2024 has outperformed the Faster R-CNN model with a mAP of 0.93. That accuracy value was almost achieved by the EFE-SSD [81] and Pruned YOLOv5s [98] with a fast detection speed. YOLOv3 with GIoU and Mosaic [28] also achieved the third fastest detection speed in the comparison of 5.7 ms/image, which is about 200 times faster than the traditional methods and about 20 times faster than some of the earlier DL models. The most recent YOLO models [101,102] also achieved good detection accuracy of 0.86 in average, with one model also delivering the fastest detection speed at 3.83 ms/image for sewer pipeline defects On the other hand, there is a trade-off between mean average precision and inference speed, such as DefectTR [107] with a lower mAP of 0.6 and a fast inference time of 55 ms/image.

Dataset and hyperparameter differences are also factors that affect the performance of locating sewer defect. According to Table 3, the CCTV footage and image resolution vary widely across different studies, from 224x224 pixels to 1500x720 pixels. The higher resolution images contain more pixels, allowing for capture of finer features, which is crucial for detecting small defects, such as cracks or roots. As a result, the higher resolution trends of recent studies helped the detection model in feature extraction, defect localisation, and potentially higher detection accuracy.

In all DL models, configuration hyperparameters play a vital role in

Table 3
Previous studies about identifying and locating sewer defect by object detection

Ref.	Year	Algorithm/Model	Dataset Size	Hyperparameter	Performance	Comments
Traditional Object Detection						
[104]	2014	SVM with Histogram of Oriented Gradients (HOG)	1,000 Images	Not Specified	Accuracy = 0.86 Speed = 1 s/image	- Image resolution of 320x240 pixels. - Model detected only root defects
[108]	2018	Image Processing Algorithms	2 CCTV Videos	Not Specified	Accuracy = 0.84-1 Speed = 1 s/image	- Inconsistent accuracy - Low performance metrics
[105]	2019	Rule-based and Image Processing Algorithms	200 Images	Not Specified	Accuracy = 0.89 Speed = 1 s/image	- Image resolution of 240x320 pixels - Detected only crack defects. Capable of defining crack characteristics
Deep Learning Models without Modification						
[36]	2018	Faster R-CNN with ZF network as backbone	3,000 Images	Not Specified	mAP = 0.83 Speed = 106 ms/image	- Image resolution of 224x224 pixels - Detected multiple defects with high accuracy - Slow inference speed
[92]	2019	YOLOv3 with additional of 5 CNN layers	1,800 Images	Not Specified	mAP = 0.71 @ IoU = 0.2 Speed = 28 ms/image	- Image resolution of 512x512 pixels - Integrated 5 CNN layers for classification prior to detection. Applied to fracture defects with a low confidence threshold (0.2)
[9]	2020	SSD	3,800 Images	Not Specified	mAP = 0.544 @ IoU = 0.2 Speed = 33 ms/image	- Image resolution of 1500x720 pixels - Applied only to intrusion root and deposit defects, not structural defects. - Validated and tested on real-time CCTV video - Detected defects with a high accuracy rate (51/56)
		YOLOv3			mAP = 0.745 @ IoU = 0.2 Speed = 57 ms/image	
		Faster R-CNN			mAP = 0.762 @ IoU = 0.2 Speed = 110 ms/image	
[24]	2020	YOLOv3	3,664 Images	Batch size of 64 Learning rate of 0.0001	mAP = 0.85 Speed = 30 ms/image	- Image resolution of 416x416 pixels - Detected multiple defects with high average accuracy. Lowest detection accuracy with crack and root defects
[88]	2022	Faster R-CNN YOLOv2	610 Images	SGDM optimizer Batch size of 8 Learning rate of 0.001	Not Specified	- Image resolution of 256x256 pixels - Able to detect multiple defects - Faster R-CNN achieved higher prediction accuracy than YOLOv2
Deep Learning Models with Modification						
[109]	2021	Strengthened Regional Proposal Network with VGG16 as backbone	2,000 Images	SGDM optimizer, Batch size of 8 Learning rate of 0.0005	mAP = 0.51 Speed = 153 ms/image	- Image resolution of 600x480 pixels - Low mean average precision and slow inference speed for detecting multiple sewer defects
[28]	2021	Improved YOLOv3 with GIoU and Mosaic	3,000 Images	ADAM optimizer Batch size of 16 Cosine learning rate	mAP = 0.92 @ GIoU = 0.92 Speed = 5.7 ms/image	- Image resolution of 416x416 pixels - Improved architecture aided faster converge and reduced training time + High accuracy with fast detection for four common defects.
[96]	2022	YOLOv5 – TB	2,333 Images	Batch size of 32 Learning rate of 0.01	mAP = 0.88 @ IoU = 0.5	- Image resolution of 640x640 pixels - Modified YOLOv5 improved detection accuracy
[89]	2023	YOLOv2 with pre-trained CNN as backbone	1,200 Images	SGM optimizer Batch size of 3 Learning rate of 0.001	Best mAP = 0.71 for InceptionV3 backbone	- Image resolution of 256x256 pixels - Compared detection accuracy of several CNN backbones for YOLOv2 in transfer learning. - Modified backbone achieved better accuracy than other models
[81]	2023	EFE-SSD	4,000 Images	Not Specified	mAP = 0.92 @ IoU = 0.5	- Image resolution of 300x300 pixels - Achieved higher accuracy compared to the original SSD and mainstream networks
[98]	2024	Pruned YOLOv5s	2,000 Images	ADAM optimizer Batch size of 16 Learning rate of 0.001	mAP = 0.918 Speed = 5.2 ms/image	- Image resolution of 640x640 pixels - Achieved high mean average precision with a 20.6% reduction in model size
[102]	2024	YOLOv8	5,000 Images	SGM optimizer Momentum of 0.9 Learning rate of 0.001	mAP = 0.85	- Image resolution of 608x608 pixels - Fused recursive feature boosting and squeeze-and-excitation attention improved accuracy and stability though not significantly compared to original YOLOv8
		YOLOv8 ++			mAP = 0.87	
[100]	2024	Sewer-YOLO-Slim	6,368 Images	Batch size of 16 Learning rate of 0.001 Weight decay of 0.0005	mAP = 0.93 @ IoU = 0.5	- Image resolution of 416x416 pixels - Reconstructed lightweight model from the YOLOv7-tiny with 60.2% and 60.0% reduction in model size and parameters, respectively - Deployed on edge devices aided by TensorRT achieving 15.3 ms/image speed
[101]	2024	RLL-YOLOv8	4,030 Images	SGD optimizer Batch size of 32 Momentum of 0.937 Learning rate of 0.01	mAP = 0.862 @ IoU = 0.5 Speed = 3.83 ms/image	- Image resolution of 352x228 pixels - Lightweight model with enhanced feature extraction (including handling multi-scale features)

(*continued on next page*)

Table 3 (*continued*)

Ref.	Year	Algorithm/Model	Dataset Size	Hyperparameter	Performance	Comments
Transformer-Based Object Detection						
[107]	2022	DefectTR	47,100 Images	LaProp optimizer Batch size of 4 Learning rate of 0.0001	mAP = 0.6 Speed = 85 ms/image	- Outperformed other networks (SSD, YOLOv4, Faster R-CNN) in mAP - Achieved higher inference speed than SSD and YOLOv4 models.

controlling the process of training and validation. Alongside the dataset and the model's architecture, hyperparameters are one of the most critical factors that critically influence the model's performance. Hyperparameters consist of several variables, such as learning rate, momentum, epochs, batch size, optimiser, activation functions, L1/L2 regularisation, learning rate schedule, etc. However, this section only discusses batch size, learning rate, and optimiser, which have been published in reviewed papers and summarised in Table 3. Firstly, the batch size is the number of training data samples processed together in one iteration of the training process. Particularly, the iteration is a single pass of all samples in the forward or backward directions. In the reviewed papers, the batch size varies from 3 to 64, which is based on the computing resources, as a larger batch size will require more memory. However, in the most recent studies from 2021 until now, researchers intentionally choose small batch size values (ranging from 3 to 16), which is explained by a more complex model's architecture or high-resolution image dataset.

Secondly, the learning rate is a hyperparameter that regulates the step size at each data sample pass of the optimisation process. It determines how quickly and slowly the neural network weights are adjusted concerning the loss gradient. A high value of learning rate results in significant updates to the weights and allows faster learning with a risk of overshooting or missing features. Conversely, a low value of the learning rate can make the model converge more precisely, but it requires more training time and computing resources to reach an optimal solution. Similar with the batch size, the chosen learning rate of reviewed studies in Table 3 shows some significant differences, ranging from 0.0001 to 0.01, with the most popular learning rate of 0.001 in 3 papers [88,89,98]. Instead of a predefined learning rate, there is a DL model proposed by Tan et al. [28] with a cosine learning rate schedule. In detail, their model started with a very large learning rate and then decreased dramatically to a value near 0 before increasing the learning rate again.

Optimiser selection is very important in DL, because it fine-tunes the neural network parameters such as weights and learning rate during the training process to reduce the losses. Optimisers are divided into two main types of non-adaptive and adaptive. Non-adaptive includes Stochastic Gradient Descent (SGD) and Momentum, which adjust the weights of model with the fixed learning rate from the initial to the end of the training, so the weight's step size is constant for the whole training process. On the other hand, the learning rate of adaptive optimisers is scheduled and adjusted based on the training process, resulting in more efficient training and better performance in the DL model. The typical adaptive optimiser includes Adaptive moment Estimation (Adam), Root Mean Square (RMSprop), and Adaptive Gradient Descent (Adagrad). The most popular optimisers for sewer defect inspection are SGD and Adam based on several studies from 2020-2024. However, in the recent study of Dang et al. [103], the authors utilised the LaProp optimiser with the separation of momentum and adaptivity and returned a faster training speed and better stability than the Adam optimiser.

To compare the performance of YOLOv8 models, Lv et al. [101] selected a portion of the Sewer-ML dataset for training and initial evaluation. The models range in size from Nano to Extra-Large, corresponding to parameter sizes from approximately 3.2 million to 68.2 million. Table 4 illustrates the trade-off between model size, accuracy (mAP and F1 score) and computational cost (Floating-Point Operations Per Second - FLOPs). The YOLOv8 Nano model, with the smallest number of parameters and lowest computational demand (8.9 GFLOPs), achieved mAP of 82.1 % and F1 score of 84.3%, making it well-suited for deployment in highly resource-constrained environments such as embedded edge devices. The Small and Medium variants offered modest improvements in accuracy (mAP of 83.3% and 83.1%, respectively) and may be suitable for systems with moderate resources, balancing efficiency and performance. YOLOv8 Large and Extra-Large models achieved the highest detection accuracy. However, these gains came at a significant increase in computational cost, particularly for Extra-Large model, which required 257.8 GFLOPs. Such models are more appropriate for high-performance computing environments where maximising detection accuracy is priority. Overall, this comparison highlights that while larger models can yield slightly better performance, the improvements may not justify the increased computational burden in many practical scenarios.

Table 4
Examine the performance of all YOLOv8 models on the portions of Sewer-ML dataset [101]

Model	Parameters (M)	mAP_{50}	F1	FLOPs (Giga)
YOLOv8 Nano	3.2	82.1	84.3	8.9
YOLOv8 Small	11.2	83.3	82.7	28.6
YOLOv8 Medium	25.9	83.1	84.1	79.3
YOLOv8 Large	43.7	82.8	82.8	165.2
YOLOv8 Extra-large	68.2	83.9	83.3	257.8

Table 5 compares a comparative study by Liu et al. [100] of recent lightweight YOLO models—including tiny or nano variants from YOLOv7 to YOLOv11. The comparison reflects the growing trend and demand for deploying models on embed systems, where computational efficiency and smaller model sizes are crucial. YOLOv9 small achieves the highest detection accuracy with a mAP of 93.4%, but at the cost of increased parameters (7.17 M) and computation (26.7 GFLOPs), making it less ideal for real-time embedded applications. In contrast, YOLOv11 Nano and YOLOv10 Nano represent a strong trade-off between performance and efficiency, with mAP scores of 90.5% and 91.1%, respectively, while maintaining very low parameter counts (2.58M and 2.27M) and minimal computational demands (6.3 and 6.5 GFLOPs). These models are particularly well-suited for deployment on embedded or edge devices, where resources are limited but real-time defect detection remains critical. This progression illustrates a clear focus in recent YOLO versions toward optimising models for practical, resource-aware application in fields.

For the beginners in object detection with sewer defects, the study by Kumar et al. [9] is recommended, as it compares three popular models:

Table 5
Comparison of lightweight YOLO models in newest version on Sewer Defect Image [100]

Model	Parameters (M)	*mAP*	FLOPs (G)
YOLOv7 Tiny	6.03	92.0	13.2
YOLOv8 Nano	3.00	92.8	8.2
YOLOv9 Small	7.17	93.4	26.7
YOLOv10 Nano	2.27	91.1	6.5
YOLOv11 Nano	2.58	90.5	6.3

SSD, YOLOv3, and Faster R-CNN, offering a clear introduction to object detection and their speed-accuracy trade-offs. Following this, Situ et al. [89] provides insights into transfer learning-based YOLO networks and the feature extraction processes of various CNN backbones, aiding in identifying optimal feature extraction CNNs for sewer defect localisation. For instance, deeper backbones like ResNet are highly effective for detecting fine-grained defects such as cracks. Their depth, enhanced by residual connections, allows them to capture subtle patterns and learn complex representations. ResNet also generalises well across diverse, noisy, or low-visibility datasets, further boosting crack detection performance [110]. In contrast, simpler backbones like **MobileNet** prioritise efficient processing, making them ideal for real-time tasks or resource-constrained environments. This efficiency enables quick detection of larger, less detailed objects like root intrusions, which don't require the fine-grained analysis needed for cracks. Darknet, the backbone used in YOLO models, strikes a balance between speed and feature extraction capability [83,111]. Optimized for real-time performance with reasonable accuracy, it excels at detecting larger, irregular objects like root intrusions where rapid processing is key. While its shallower depth compared to ResNet allows faster image processing, it may sacrifice fine-grained detail, making it less ideal for detecting small, subtle defects like cracks. Nevertheless, Darknet's speed and balanced accuracy make it a strong choice for systems requiring quick decisions in dynamic, real-time sewer inspections.

For studies focusing on specific defect types, newcomers can explore research tailored to particular objects. For structural defects like cracks or deformations in sewer systems, fine-tuned models with high-resolution feature mapping, such as Faster R-CNN or U-Net, are ideal as they excel at detecting narrow, elongated structures. Conversely, operational defect detection (e.g., root intrusion, deposit, settlement) might benefit from real-time approaches like YOLO, which effectively handle irregularly shaped and dynamic objects. Exploring these resources helps beginners understand which models best address the distinct challenges of different sewer defect types.

6. Characterising defects by image segmentation

Characterising defects by image segmentation algorithms plays a crucial role in predicting defect categories and providing pixel-level location information with precise shapes, which is essential for sewer condition assessment. The complexity of image segmentation arises from the need to accurately classify each pixel while also distinguishing between different instances of the same defect type. Researchers have developed several segmenting models for sewer pipeline defects based on different algorithms, including morphological segmentation [112], semantic segmentation [113], and instance segmentation [114]. Morphological segmentation, an early computer vision technique, leverages image processing and mathematical morphology to quickly isolate and enhance features. While efficient and relatively simple, its rigid nature limits its effectiveness with complex and varied image structures. In contrast, deep learning-based segmentation (semantic or instance) provides superior accuracy and adaptability. By learning from extensive labelled datasets, these methods usually excel at handling intricate and diverse features, making them more robust for modern imaging challenges.

6.1. Morphological segmentation

Morphological segmentation is a technique in image processing that utilises mathematical morphology to analyse object structure, primarily manipulating shapes and features in binary and grayscale images [115]. Key operations include erosion (shrinking objects), dilation (expanding objects), and their combinations: opening (erosion then dilation) and closing (dilation then erosion). The top-hat transform (original minus opened) highlights small bright elements, while the bottom-hat transform (closed minus original) emphasises small dark elements [116,117]. In 2009, Yang et al. integrated the opening operation with Otsu's technique for automated sewer defect diagnosis [118]. However, the method proved ineffective, accurately segmenting only 62 of 291 defects. Poor performance was attributed to issues like camera pose, lighting, sewage, and unsmoothed CCTV footage notations.

In another research, Su et al. [112] proposed MSED, a morphological segmentation model for sewer pipe defects based on edge detection, which offers more precise segmentation. MSED was tested against a diverse dataset of pipe defects (e.g., fractures, debris, holes) and outperformed the opening top-hat operation (OTHO) for specific defects like broken pipes and holes. Later on, Su et al. [117] compared MSED, OTHO and closing bottom-hat operation (CBHO) on 20 vitrified clay pipe images (from 10-minute CCTV video) showed MSED excelled at crack detection, while OTHO outperformed it for open joint defects.

6.2. Semantic segmentation

Semantic segmentation is a core CV task that involves assigning a semantic label to every image pixel, providing a detailed, pixel-level understanding of a scene. Unlike image classification or object detection, which offer broad labels or bounding boxes, semantic segmentation precisely delineates distinct objects and regions. Several segmentation architectures have been adapted for sewer defect inspection, including the Fully Convolutional Network (FCN), DilaSeg, U-Net, SegNet, and Deeplab, each with unique characteristics.

FCN [119] was foundational, transforming the traditional CNN architecture to produce dense, pixel-wise predictions by replacing fully connected layers by convolutional ones and and utilising skip connections to combine multi-resolution features. DilaSeg is, on the other hand, a deep convolutional neural network developed for semantic segmentation that uses dilated convolutions and multi-scale dilated convolutions to address spatial information loss and improve feature map resolution, especially for objects of varying scales. Wang et al. [25] compared the performance of these two models on sewer defects, and found that FCNs yielded 18% lower mean Pixel Accuracy (mPA) and 22% lower mean Intersection over Union (mIoU) than DilaSeg. Subsequently, Wang et al. [38] proposed DiLaSeg-CRF, integrating a dense Conditional Random Field (CRF) module. By converting recurrent neural network (RNN) layers into CNN operations, CRF improved the accuracy and inference speed of the standard DilaSeg model (Fig. 12). This modification improved standard DilaSeg's accuracy and inference speed, increasing mIoU by 32% over FCNs and 20% over DilaSeg in pipe defect datasets.

Based on the FCN concept, Ronneberger et al. [120] developed U-Net in 2015 for biomedical image segmentation, featuring a symmetric encoder-decoder architecture with skip connections. These connections link corresponding encoder and decoder layers, preserving spatial information and fine details. For sewer defect characterisation, Pan et al. [113] introduced PipeUNet in 2020, using U-Net as a backbone due to its rapid convergence. They enhanced it with a Feature Reuse and Attention Mechanism (FRAM) block in skip connections and focal loss to handle imbalanced datasets. To improve the feature extraction process, the FRAM block is located before the skip connection between the encoder and decoder parts. In 2024, Li et al. [121] developed PipeTransUnet, a modified U-Net architecture, for semantic segmentation and severity quantification in sewer pipes. This method incorporated a hybrid Transformer model and a ResNet50 backbone for efficient feature extraction, along with Channel Attention Module (CAM) and Position Attention Module (PAM) [122]. Both PipeUNet and PipeTransUnet demonstrated significant advantages over the base U-Net, achieving 5.95% and 45.95% higher mIoU, respectively.

With a similar encoder-decoder architecture, SegNet was designed for semantic segmentation for autonomous driving and medical imaging by Badrinarayanan et al. [123]. Its encoder uses 13 convolutional layers from VGG16 for feature extraction. The decoder upsamples features to full resolution using pooling indices from the encoder, maintaining

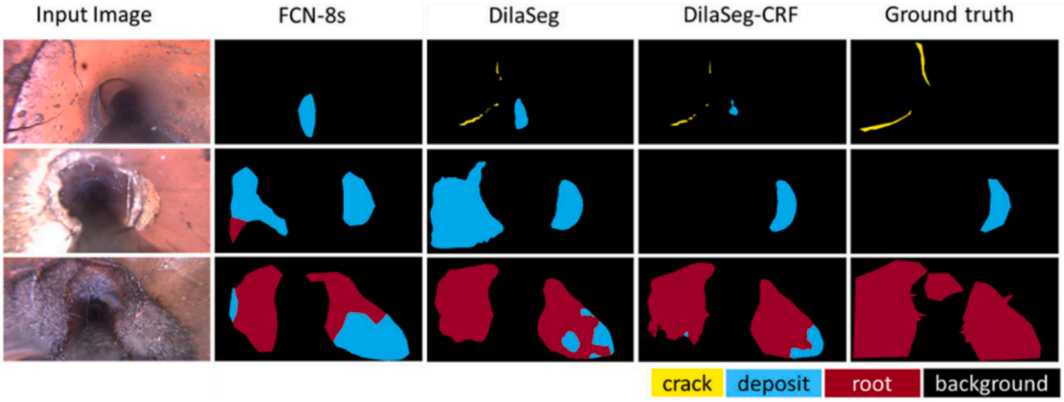


Fig. 12. Example of defect characterisation by Wang et al. [38].

spatial information and improving computational efficiency. He et al. [124] examined SegNet with a VGG16 backbone for automated sewer pipeline defect classification and segmentation. They employed histogram equalisation, weighting, and augmentation to improve accuracy and address dataset imbalance. The model achieved PA of 80.89% and mean IoU of 0.68.

DeepLab, developed by Chen et al. [125], is a notable semantic segmentation model combining atrous convolution, deep CNNs, and a fully connected CRF. Atrous convolution expands the receptive field without losing resolution, capturing multi-scale context effectively. The CRF refines output boundaries. Subsequent improvements of DeepLab led to DeepLabV3 and DeepLabV3+ [126]. These versions refined the encoder-decoder framework with Atrous Spatial Pyramid Pooling (ASPP), atrous separable convolutions, and batch normalisation, while removing CRFs. This significantly boosted accuracy and multi-scale context capture on datasets like PASCAL VOC 2012. DeepLabV3+ models have been applied in two sewer defect inspection studies [26,127]. These studies tested the model on diverse sewer datasets, consistently achieving high pixel accuracy exceeding 90%. Dang et al. [127] also investigated the impact of various backbones on DeepLabV3+ performance, identifying ResNet152 as as the optimal backbone for sewer datasets.

6.3. Instance segmentation

Instance segmentation extends semantic segmentation by not only categorising each pixel but also differentiating individual instances within the same class; each object of a predefined class is uniquely segmented (Fig. 13). Algorithms are typically one-shot or detection-based. One-shot methods, like SSDs in object detection, identify and segment instances in a single network pass, offering faster training for real-time applications. Popular one-shot networks include Mask R-CNN and YOLACT. Mask R-CNN [128] extends Faster R-CNN with a mask prediction branch, detecting objects via bounding boxes and then segmenting their boundaries. YOLACT [129] focuses on real-time segmentation by combining one-shot methods with a novel mask generation approach. Ma et al. [130] utilised YOLACT to propose Pipe-Yolact-Edge, a real-time instance segmentation system for sewer pipeline defects. Trained on 1,403 images, the model achieved a high mean Average Precision (mAP) of approximately 92% on both a high-performance server and an embedded Jetson TX2 device for on-site inspection. While accuracy remained high, inference speed on the embedded device was significantly lower than the server.

Detection-based instance segmentation operates in two stages: first detecting objects, then segmenting them, often utilising separate networks for each task. For example, SOLOv2 simplifies the segmentation task by converting it into a classification problem and segments the instance based on its spatial location within the grid [131]. Li et al. [114] proposed Pipe-SOLO to enhance SOLOv2 for underground sewer defects. This robust model integrates a Res2Net-Mish-BN-101 module into its backbone and EBiFPN in its neck section. Tested on a dataset of 3,888 images across six types of defects, Pipe-SOLO outperformed existing methods like SOLOv2, Mask R-CNN, and MS R-CNN.

6.4. Discussion

Table 6 summarises eleven studies from 2009 to date on sewer pipeline defect characterisation using image segmentation techniques. These models share some hyperparameters similar to object detection models including learning rate, optimiser, and batch size, along with additional ones like momentum and weight decay. Learning rates typically range from 0.0001 to 0.01, with 0.01 being the most common, though recent papers lean towards smaller values for fine-tuning and preventing missed optimal model features. Batch sizes vary from 4 to 16, with 8 being prevalent; larger batch sizes require more memory but yield more stable gradient estimates. Like in object detection, SGDM and Adam are common optimizers in image segmentation. SGDM introduces Momentum (consistently 0.9 in reviewed papers), which accelerates training and affects convergence speed and stability. High Momentum often necessitates a lower learning rate and larger batch size. To prevent overfitting, L2 regularisation (weight decay) was utilised for a few models [25,113,130] with two values of 0.9 and 0.005. This adds a proportional sum of squared weights to the loss function, directly controlling the strength of this penalty to help the model generalise better by discouraging overly large weights.

The datasets used in sewer pipeline defect characterisation studies show a clear trend: they have grown significantly in size and diversity. Datasets now range from as few as 100 images with two defect types

Fig. 13. Example of semantic and instance segmentation.

Table 6
Previous studies on characterising sewer defect using image segmentation

Ref.	Year	Algorithm/Model	Dataset Size	No. of Classes	Hyperparameter	Performance	Comment
Morphological Segmentation							
[118]	2009	Otsu's technique	291 Images	3 Classes	N/A	69/291 of successful segmented	- Low accuracy and not effective for detecting fractures
[117]	2014	MSED – Edge detection	100 Images	2 Classes	N/A	N/A	- Only robust in detecting open joints and cracks.
Instance Segmentation							
[114]	2022	Pipe_SOLO-Res2Net-Mish-BN-101	3,888 Images	6 Classes	SGDM optimizer Learning rate – 0.01 Momentum – 0.9 Batch size - 8	mAP = 0.593 Min loss = 0.11 Speed = 66.7 ms/image	- Size from 640x480 to 1280x720. - Lower average precision due to low quality dataset. - Unable to test on video or real-time CCTV
[130]	2022	Pipe-Yolact-Edge	4,209 Images	3 Classes	Learning rate – 0.0005 Momentum – 0.9 Weight decay – 0.005	mAP = 0.926 Speed = 24.2 ms/image	- Image resolution of 320x256 to 1440x720. - High prediction accuracy and fast detection speed. - Stimulation severe environmental conditions
Semantic Segmentation							
[25]	2019	DilaSeg	1,510 Images	N/A	Learning rate – 0.01 Momentum – 0.9 Weight decay – 0.005 Iteration – 50,000	mAP = 0.81 mIoU = 0.74 FwIoU = 0.92 Speed = 270 ms/image	- High detection accuracy but slow in detection speed. - Ability to detect on real-time CCTV inspection
[38]	2019	DilaSeg-CRF	1,885 Images	3 Classes	SGD optimizer Learning rate – 0.001 Batch size – 6	mPA = 0.92 mIoU = 0.85 FwIoU = 0.97 Speed = 107 ms/image	- Higher detection accuracy and inference speed than original DilaSeg model.
[113]	2020	PipeUNet	3,654 Images	4 Classes	Adam optimizer Learning rate – 0.0001 Weight decay – 0.9 Epoch - 200	mIoU = 0.76 Speed = 31.25 ms/image	- Image resolution of 256x256 pixels. - Better accuracy for single defect characterisation than multi-defect
[124]	2022	SegNet – backbone VGG16	700 Images	7 Classes	Batch size - 4 Learning rate – 0.01 Momentum – 0.9 Epoch - 90	mPA = 0.8 mIoU = 0.61 BFScore = 0.73	- Image resolution of 360x480 pixels. - High mean pixel accuracy but low in detection accuracy results.
[26]	2022	DeeplabV3+ – Resnet50	600 Images	5 Classes	Batch size – 8	PA = 0.9 mIoU = 0.53 FwIoU = 0.84 F1 = 0.55	- Image resolution of 512x512 pixels. - High pixel accuracy and able to integrate with defect severity analysis, but low detection accuracy
[127]	2023	DeeplabV3+ – Resnet152	3,699 Images	10 Classes	Batch size – 16 Learning rate – 0.01 Momentum – 0.9 Epoch – 80	mPA = 0.98 mIoU = 0.69 Speed = 38.5 ms/image	- Image resolution of 512x512 pixels. - High mean pixel accuracy and fast detection speed, while the detection accuracy is low.
[121]	2024	PipeTransUNet –Resnet50	1,700 Images	8 Classes	Batch size – 12 Learning rate – 0.001 Epoch – 1,500	mIoU = 0.72 mPA = 0.85 FwIoU = 0.91	- Image resolution of 224x224 pixels. - Outperformance other CNN semantic segmentation. Integrated with defect severity assessment
[132]	2024	Enhance Feature Pyramid Network	6,300 Images	9 Classes	Adam optimizer Learning rate – 0.001	mIoU = 0.77 FwIoU = 0.78 F1 = 0.86	- Trained on imbalance-aware dataset - mIoU was improved by 13.8%, while the model parameter is reduced by 96.04%
[133]	2025	Improved DeeplabV3+ –MobileNetv2	1,795 Images	5 Classes	Batch size – 4 Epoch – 300	mIoU = 0.86 mPA = 0.92 Speed = 79.7 ms/image	- Image resolution of 512x512 pixels - Using StyleGAN3 for enhancing sewer defect dataset - Modified model increased the segmentation speed and accuracy of 31% and 20%, respectively, compared to original model

[117] to up to 6,300 images [132]. Notably, the number of distinct defect classes has expanded from 2-3 to 10, a comprehensive scope introduced by Dang et al. in 2023 [127] which is also one of the most diverse datasets to date. Overall, this expansion indicates a move towards models that can handle a wider array of real-world sewer defects. Nevertheless, image resolutions varied widely, from 224x224 to 1280x720 pixels, reflecting a lack of standardisation.

Sewer defect segmentation models commonly use mIoU to assess the overlap between predicted and ground truth masks, with higher values indicating more precise boundaries. Table 6 shows mIoU ranging from 0.53 to 0.86. The highest mIoU (0.86) was achieved by integrating StyleGAN3, an advanced generative adversarial network, with DeeplabV3+ and leveraging MobileNetv2 [133], highlighting the benefit of combining models for lightweight and accurate defect recognition. The improved DilaSeg-CRF achieved the second-best mIoU of 0.85 in conjunction with a rapid inference speed of 107 ms/image. The addition of the CRF module minimised limitation of the original model, particularly in the difficult sewer environments with unclear defect boundaries. DeepLabV3+ with a Resnet50 backbone achieved the lowest mIoU value, highlighting its limitations in detecting smaller, less frequent

defects and handling imbalanced datasets.

For semantic segmentation, Frequency Weighted IoU (FwIoU) was utilised to represent the importance of each class based on its pixel frequency in the dataset, thus emphasising classes with higher pixel counts. Subsequently, in imbalanced dataset, it adjusts the contribution of each class based on their frequency for a more balanced model evaluation. This metric was reported in five papers [26,38,108,121,132] with high FwIoU values ranging from 0.78 to 0.97, indicating a good performance for common defect classes (such as large cracks or root intrusions). Some models achieved high FwIoU, such as DilaSeg-CRF and PipeTransUNet, generalise better across different defect classes, maintaining high accuracy even in the face of dataset imbalances due to their multi-scale processing capabilities.

Another essential metric in comparing segmentation models is inference speed. The fastest models, with inference speed of less than 40 ms/image, are Pipe-Yolact-Edge, PipeUnet and DeeplabV3+, which might be suitable for real-time capability. These models strike a balance between speed and accuracy, with Pipe-YOLACT-Edge achieving a high mAP of 0.926 and PipeUNet reaching a respectable mIoU of 0.76. In contrast, DilaSeg and DilaSeg-CRF in 2019 are the two slowest models, with inference speeds of 3.70 and 9.35 images/second, respectively. The increased complexity in their feature extraction processes and post-processing explains this slower inference time, making them better suited for offline analysis rather than real-time application.

Pixel accuracy (PA) metric, was another metric used to evaluate segmentation accuracy, demonstrating the corrected classified pixels to the total number of pixels in an image. As shown in the table, all reported PA values exceed 0.8, indicating consistently high pixel-level accuracy across sewer defect segmentation models. However, there is a limitation with the PA metric, which does not consider the class imbalance of the dataset, so it is challenging to depend on this metric to compare the models.

For newcomers to this research area, the PipeUNet model proposed by Pan et al. in 2020 [113] is a strong suggestion for exploration. It has a relatively simple architecture based on U-Net, which is the most popular model for image segmentation due to its effectiveness and ease of implementation. Furthermore, the model achieves a good balance between segmentation accuracy and inference speed, which demonstrates practical usability without requiring extensive computational resources, making it accessible for newcomers. PipeUNet provides a practical introduction to segmentation model in sewer inspection, allowing a new user to gain confidence with straightforward model before exploring more complex architectures, such as SegNet with backbone of VGG16 network [124] and DilaSeg [25].

7. Severity assessment and decision-making process

Severity assessment is a crucial step in sewer inspection, providing early maintenance insights after defect characterisation. Traditionally reliant on expert judgment, this process is slow and error prone. Recent studies have introduced ML methods to automate severity evaluation based on country-specific standards.

The Pipeline Assessment and Certification Program (PACP), developed by NASSCO, is a widely adopted standard for assessing sewer pipeline defects [134]. It standardises sewer condition assessment and reporting, ensuring consistent inspections and providing reliable criteria and information for management and maintenance decisions. PACP utilises a comprehensive grading system to classify two main defects, including structural defects and operational and maintenance (O&M) defects. The severity conditions of defects are evaluated based on scale from 1 to 5, meaning minor to major defects (as shown in Table 7). Several studies [80,107,127,135] have utilised NASSCO's PACP for their guidance in defect severity examination in a sewer pipeline system. Dang et al. [107] proposed an automated framework for transformer segmentation of sewer defects and severity assessment. The assessment decision was made based on the zone of influence (ZOI) with the mean attention weight from the feature map combined with defect scores from PACP. In another study, Dang et al. [127] also applied PACP to grading the severity of the defects. Additionally, their system has the ability to count the number of sewer faults and locate their coordinates along the pipeline.

Table 7
NASSCO's PACP defect grade categories

Grading Level	Severity	Actions
1	Minor defects	No immediate action required
2	Minor defects	Minimal deterioration, unlikely to worsen
3	Moderate defects	Deterioration or failure may occur in the future
4	Significant defects	Likely to deteriorate further and require attention
5	Severe defects	Immediate attention required

Besides the PACP standard, another popular standard for visual inspection and condition assessment of stormwater and sewer systems is German EN 13508-2 for European countries. Daher et al. [135] introduced a defect-based condition assessment model using fuzzy hierarchical evidential reasoning, combing standards, including NASSCO's PACO and German EN 13508-2. There are two datasets with four and seven defect types utilised in this study. The overall condition of a defect is evaluated scored defect severity on a scale from 1 (minor) to 5 (severe) using fuzzy membership function and hierarchical reasoning. The model achieved a mean absolute error of 0.643, reflecting good overall predictive accuracy and low average error across both datasets, though some inaccuracy remained in precise severity prediction.

For the pipeline system in China, Jia et al. [136] proposed a defect condition assessment model (DSA-APC) with two diverse defect datasets based on automated pipe calibration and aligned with the Mainland China CJJ 181-2012 standard, which employs a severity scoring system from 1 to 4. The condition of a defect is evaluated based on the area ratio between the defect and the cross-sectional area of pipeline. Depending on the defect types, structural or operational, the severity level and scores are different and matched based on the area ratio (as Table 8). The proposed model has been tested on a ten-defects dataset and achieved high accuracy and mean absolute deviation of 86.73% and 2.008, respectively.

Table 9 summarises research into defect severity assessment for sewer pipelines in recent years. Most studies rely on visual inspection data and apply factors such as defect area, pixel count, fuzzy logic, or attention mechanisms to estimate severity levels. Widely adopted standards like PACP, EN 13508-2, and CJJ 181-2012 form the basis for grading and categorising defect severity, typically on scales ranging from 1 (minor) to 5 (severe).

Several studies in the table use area ratios and pixel measurements to quantify defect severity. While this approach provides a simple and direct way to assess the physical extent of defects, it lacks the ability to account for structural significance, defect orientation, or spatial context. Consequently, these methods often miss minor but critical defects or misclassify severity, especially for complex conditions like cracks or deformations.

8. Research challenges and future directions

This section first summarises common challenges of researching into sewer defect inspection using vision data such as CCTV. It then outlines future directions in the field, offering potential solutions to current challenges and ways to enhance inspection efficiency.

8.1. Research challenges

First, various datasets have been used as described in the previous sections but challenges remain due to their limited public availability

Table 8
Defect severity assessment table proposed by Jia et al. [136]

Structural Defects	P_{loss} (%)	Level	Score	Operational Defects	P_{loss} (%)	Level	Score
Crack (CK)	≤ 10	1	0.5	Deposit (DP)	≤ 20	1	0.5
	$(10, 25]$	2	2		$(20, 40]$	2	2
	$(25, 60]$	3	5		$(40, 60]$	3	5
	> 60	4	10		> 60	4	10
Deformation (DF)	≤ 10	1	0.5	Obstacle (OBS)	≤ 15	1	0.5
	$(10, 25]$	2	2		$(15, 25]$	2	2
	$(25, 60]$	3	5		$(25, 50]$	3	5
	> 60	4	10		> 50	4	10
Corrosion (CR)	≤ 10	1	0.5	Root (RT)	≤ 15	1	0.5
	$(10, 50]$	2	2		$(15, 25]$	2	2
	> 50	3	5		$(25, 50]$	3	5
	-	-	-		> 50	4	10
Side branch (SB)	≤ 10	1	0.5	Encrustation (ER)	≤ 15	1	0.5
	$(10, 30]$	2	2		$(15, 25]$	2	2
	> 30	3	5		$(25, 50]$	3	5
	-	-	-		> 50	4	10
Penetration (PT)	≤ 10	1	0.5	Broken wall (BW)	≤ 15	1	0.5
	$(10, 30]$	2	2		$(15, 25]$	2	2
	> 30	3	5		$(25, 50]$	3	5
	-	-	-		> 50	4	10

Table 9
Previous studies/research about sewer defect severity assessment

Ref.	Year	Factors of Severity Assessment	Standard	Dataset	Grading	Performance	Comments
[137]	2018	- Probability of failure based on CCTV inspection report and Dynamic Bayesian Belief Network (DBN) - Consequence of failure based on total costs resulting from failures	Fuzzy inference system of Sugeno	Not Specified	Minor – Major: 1 – 3	Not Specified	The maintenance decision will be made based on the cost benefit analysis and risk analysis matrix.
[135]	2021	- Fuzzy membership functions - Hierarchical evidential reasoning	NASSCO – PACP[1], German EN 13508-2	Dataset 1: 4 defects Dataset 2: 7 defects	Best – Worst: 1 – 5	MAE[2]= 0.643	Inaccuracy prediction severity of defects
[80]	2021	- Area ratio between defect and cross section area - Severity scores	NASSCO – PACP	Not Specified	Best – Worst: 1 – 5	Deviation = 3.06%	Missed some minor defect in severity assessment
[26]	2022	- Area ratio between defect and cross-section area - Number of pixels in area of defect	Not Specified	600 images with 5 defects	Best – Worst: 1 – 4	Accuracy = 70%	Not accurate for crack defects. Not quantify the locations and directions of cracks
[107]	2022	- Zone of influence (ZOI) based on mean attention weight feature map - Defect grade based on PACP	NASSCO – PACP	47,100 images with 10 defects	Best – Worst: 1 – 5	Not Specified	Unable to make maintenance decision.
[127]	2023	- Counting number of defects - Located defects along the pipeline	NASSCO – PACP	3699 images with 10 defects	Best – Worst: 1 – 5	Not Specified	Not examine the severity of defect. Counting number of defects and report its location in pipeline
[136]	2023	- Area ratio between defect and cross-section area	Mainland China – CJJ 181-2012	Dataset 1: 13 videos with 18 defects Dataset 2: 1,3633 images - 10 defects	Best – Worst: 1 – 4	Accuracy = 86.73% Deviation = 2.008%	Ability to replace manual assessment and get rid of human factors. Low accuracy for structural defects (crack, deformations, etc.)

[1] NASSCO - PACP: National Association of Sewer Service Companies – Pipeline Assessment Certificate Program.
[2] MAE: Mean Absolute Error.

and significant difference in image resolution and quality, causing barriers in making fair comparisons. These problems often stem from the harsh pipeline environments, characterised by high water level, debris and limited lighting. These conditions often result in low-quality images with significant noise and low contrast, which can affect the performance of inspection models. Such variability affects the performance of CV algorithms, reducing both their accuracy and generalisability.

Second, it is well known that choosing the right capture resolution is crucial for model performance. High-resolution images contain more details and textures, beneficial for inspecting small defects like roots or cracks, but require more storage and may need down-sampling to satisfy the model requirements. Conversely, low-resolution images may miss fine details. Future research should compare down-sampled high-resolution images with native low-resolution images to propose a standardised resolution for sewer defect inspection.

Next, most sewer defect inspection tasks rely on supervised learning, which are requires large amounts of labelled datasets for effective model training. This reliance poses a significant challenge in this domain, as many current datasets are unannotated or partially annotated, requiring manual annotation processes that are labour-intensive, time-consuming, and prone to human error. Furthermore, the manual annotation not only increases the likelihood of inconsistencies in the labels but also limits the scalability of data preparation, particularly for large-scale inspection projects. On the other hand, imbalanced datasets also pose a problem in which certain defect types may occur frequently and dominate the dataset. This problem can lead to biased models that perform well on the majority classes but fail to detect minor defect types.

Lastly, assessing defect severity based on area and pixel-to-

dimension ratio without considering defect geometry can be inaccurate. For example, inspectors cannot determine the type or impact of a crack based solely on area ratio. More comprehensive defect information, such as material properties and geometric context, is needed. High-quality capturing techniques like LiDAR, X-ray, or laser scanner can supplement datasets with 3D mapping and surface profiling.

8.2. Future directions

There are potential solutions and future directions to address the limitations and challenges in current sewer pipeline defect inspection:

- *Image Resolution and Quality*: Implementing a uniform image capturing technique, including specific resolutions for CCTV videos or images and lighting requirements, is essential. High-resolution cameras and scanners can provide clear, detailed textures and accurate colours. Image stabilisation can also reduce noise, increase contrast, and correct distortions.
- *Standardised Evaluation Metrics*: Unifying and standardising evaluation metrics will facilitate easier comparisons between different inspection tasks and models. More attention should be paid to the metrics for multiclass problems, as these are generally more complex than those for binary scenarios.
- *Detection Models*: Recent research on defect classification and detection, primarily from 2018 to 2023, has significantly advanced the field but often have difficulty to address newer, more complex defects found in sewer pipeline systems. To overcome this limitation, recent models, such as YOLOv7 and YOLOv8, with their enhanced feature extraction capabilities and improved detection accuracy, have been deployed in field applications. Besides, generative AI models like GANs and diffusion models can also address these problems of generating synthetic, high-quality images of diverse and underrepresented defect types, which can result in enhancing dataset diversity and robustness. The recent YOLO versions, from version 9 to 11, should also be examined in field for sewer pipeline conditions.
- *Segmentation Models*: While semantic segmentation models are commonly used in sewer pipeline inspection, they often lack granularity needed to differentiate and analyse individual defects. Future research should focus on advancing instance segmentation models, not only to identify but also to locate and outline specific instances of each defect class, enabling more detailed assessments and facilitating targeted maintenance strategies.
- *Learning Models*: Most current inspection models rely on supervised learning, necessitating the exploration and development of semi-supervised, self-supervised or unsupervised learning models to reduce the need for manual defect annotation. It allows model to learn useful feature representations from unlabelled data, thereby reducing the need for labour-intensive annotation and improving generalisability.
- *Edge Computing*: Edge computing enables real-time, on-site sewer inspection by deploying AI models on compact, high-performance devices like Nvidia Jetson or Google Coral [138]. This approach processes data locally, eliminating the need for large data transfer to centralised servers and reducing latency. To fully leverage the benefits of edge computing, the development of lightweight deep learning model is essential. These models are specifically optimised for low-power, resource-constrained environments while maintaining high detection performance. By enabling immediate inspection and analysis, the edge computing enhances efficiency, supports real-time decision-making and facilitates autonomous inspection workflows.
- *Defect Severity Assessment*: More precise measures of defects, such as size, depth, and shape, are needed for accurate assessment. High-resolution 3D imaging techniques can provide this critical spatial information. Furthermore, a standardised approach to quantify the severity of sewer defect should be developed and adopted nationally or internationally to ensure consistency, improve comparability, and support informed decision-making in sewer infrastructure management.

9. Conclusions

This paper presented a comprehensive review of the current state of sewer pipeline defect inspection utilising the CV technology. It focused on three primary categories of vision-based methods: detecting defects by image classification, locating defects by object detection, and characterising defects by image segmentation. In response to limitations identified in earlier reviews, this study provided a more comprehensive and technically detailed synthesis, including a critical analysis of inspection algorithms, benchmarking datasets, and condition assessment methodologies.

One of the key contributions of this study is the identification of several important findings in the application of CV into sewer pipeline defect inspection:

- A clear trend in sewer defect inspection methodologies was identified, transitioning from traditional image processing (before 2015) to ML (mainly between 2015 and 2018) and then DL (from 2018 onwards). DL methods currently dominate automated defect detection and characterisation while improving detection performance. There is a growing emphasis on deploying real-time detection systems on embedded systems and robots.
- The differences between various sewer defect datasets are identified, focusing on key factors such as resolution, dataset size, diversity of defects, capturing devices. Many image processing and augmentation algorithms to improve the quality of datasets were also described.
- Various inspection algorithms and evaluated their performance metrics for each level of sewer defect inspection in different categories of image classification, object detection and image segmentation.
- Current automated severity assessment methodologies and defect grading system were reviewed, as outlined in guidelines and standards. This process is expected to play an essential role in automated decision-making of repair and maintenance.

Furthermore, several research gaps and challenges in current CV-aided inspection systems have been identified, along with future directions to address these issues and enhance overall system performance. These include:

- There are issues with dataset availability and credibility, such as a lack of labelled datasets, inconsistencies in pixel accuracy, and insufficient diversity in sewer defect classes. Challenges also exist in defect severity assessment due to a lack of high-quality information, such as 3D data and surface profiling.
- There also exist inconsistencies in setting up hyperparameters and performance metrics, creating difficulties to conduct fair comparisons between studies. Reducing these inconsistencies can provide a more level playing field and higher research productivity for scholars and practitioners.
- Challenges in severity assessment require further investigations to enhance the algorithms and achieve better detection performance.

While significant progress has been made, considerable challenges and opportunities persist for future research aimed at developing more reliable and efficient automated inspection systems. This paper, through its systematic review of various CV models, datasets, and performance metrics, offers tailored insights and recommendations designed to guide researchers, especially newcomers, in addressing these opportunities and advancing the field of sewer inspection applications.

Ultimately, the insights garnered from this systematic review

underscore the transformative potential of CV in elevating the efficiency and objectivity of sewer defect inspection. Beyond academic advancements, the practical adoption of these technologies holds profound implications for urban infrastructure management, enabling proactive maintenance, mitigating costly failures, and significantly enhancing public health and environmental protection by preventing sewage leaks and pollution. There is a continued need for concerted efforts from researchers, industry practitioners, and policymakers to foster collaborative initiatives, standardise data collection and evaluation protocols, and invest in robust, scalable solutions that can seamlessly integrate into existing operational frameworks, thereby accelerating the transition towards smarter, more resilient sewer systems not just in developed countries but also around the globe.

CRediT authorship contribution statement

C. Long Nguyen: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Andy Nguyen:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Jason Brown:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Investigation. **L. Minh Dang:** Writing – review & editing, Visualization, Validation, Investigation.

Declaration of competing interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data availability

This review synthesises data from previously published studies. All data supporting the conclusions of this review are available in the cited references. No new data were generated or analysed during this study.

References

[1] M. Water, Sewer Relining and Maintenance Hole Rehabilitation Program, Melbourne Water, 2025. https://www.melbournewater.com.au/services/projects/sewer-relining-and-maintenance-hole-rehabilitation-program (accessed July 3 2024).

[2] F.K. Alqahtani, A. Alsharef, G.M. Hommadi, M.A. Alammari, Assessment framework for the maintainability of sewer pipeline systems, Appl. Sci. 13 (21) (2023) 11828, https://doi.org/10.3390/app132111828.

[3] S. Moradi, T. Zayed, F. Golkhoo, Review on computer aided sewer pipeline defect detection and condition assessment, Infrastructures 4 (1) (2019) 10, https://doi.org/10.3390/infrastructures4010010.

[4] W. Guo, L. Soibelman, J. Garrett Jr., Automated defect detection for sewer pipeline inspection and condition assessment, Autom. Constr. 18 (5) (2009) 587–596, https://doi.org/10.1016/j.autcon.2008.12.003.

[5] O. Duran, K. Althoefer, L.D. Seneviratne, State of the art in sensor technologies for sewer inspection, IEEE Sensors J. 2 (2) (2002) 73–81, https://doi.org/10.1109/JSEN.2002.1000245.

[6] R. Wirahadikusumah, D.M. Abraham, T. Iseley, R.K. Prasanth, Assessment technologies for sewer system rehabilitation, Autom. Constr. 7 (4) (1998) 259–270, https://doi.org/10.1016/S0926-5805(97)00071-X.

[7] B.F. Spencer, V. Hoskere, Y. Narazaki, Advances in computer vision-based civil infrastructure inspection and monitoring, Engineering 5 (2) (2019) 199–222, https://doi.org/10.1016/j.eng.2018.11.030.

[8] T. Czimmermann, G. Ciuti, M. Milazzo, M. Chiurazzi, S. Roccella, C.M. Oddo, P. Dario, Visual-based defect detection and classification approaches for industrial applications—a survey, Sensors 20 (5) (2020) 1459, https://doi.org/10.3390/s20051459.

[9] S. Kumar Srinath, M. Wang, M. Abraham Dulcy, R. Jahanshahi Mohammad, T. Iseley, C.P. Cheng Jack, Deep learning–based automated detection of sewer defects in CCTV videos, J. Comput. Civ. Eng. 34 (1) (2020) 04019047, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000866.

[10] R. Rayhana, Y. Jiao, A. Zaji, Z. Liu, Automated vision systems for condition assessment of sewer and water pipelines, IEEE Trans. Autom. Sci. Eng. 18 (4) (2020) 1861–1878, https://doi.org/10.1109/TASE.2020.3022402.

[11] Y. Li, H. Wang, L.M. Dang, H.-K. Song, H. Moon, Vision-based defect inspection and condition assessment for sewer pipes: a comprehensive survey, Sensors 22 (7) (2022) 2722, https://doi.org/10.3390/s22072722.

[12] L. Sun, J. Zhu, J. Tan, X. Li, R. Li, H. Deng, X. Zhang, B. Liu, X. Zhu, Deep learning-assisted automated sewage pipe defect detection for urban water environment management, Sci. Total Environ. 882 (2023) 163562, https://doi.org/10.1016/j.scitotenv.2023.163562.

[13] J.B. Haurum, T.B. Moeslund, A survey on image-based automation of CCTV and SSET sewer inspections, Autom. Constr. 111 (2020) 103061, https://doi.org/10.1016/j.autcon.2019.103061.

[14] K. Mostafa, T. Hegazy, Review of image-based analysis and applications in construction, Autom. Constr. 122 (2021) 103516, https://doi.org/10.1016/j.autcon.2020.103516.

[15] M.K. Scheuerman, A. Hanna, E. Denton, Do datasets have politics? Disciplinary values in computer vision dataset development, Proc. ACM. Hum. Comput. Interact. 5 (CSCW2) (2021) 1–37, https://doi.org/10.1145/3476058.

[16] S.I. Hassan, L.M. Dang, I. Mehmood, S. Im, C. Choi, J. Kang, Y.-S. Park, H. Moon, Underground sewer pipe condition assessment based on convolutional neural networks, Autom. Constr. 106 (2019) 102849, https://doi.org/10.1016/j.autcon.2019.102849.

[17] P. Huynh, R. Ross, A. Martchenko, J. Devlin, 3D anomaly inspection system for sewer pipes using stereo vision and novel image processing, in: 2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA), 2016, pp. 988–993, https://doi.org/10.1109/ICIEA.2016.7603726.

[18] S. Iyer, S.K. Sinha, A robust approach for automatic detection and segmentation of cracks in underground pipeline images, Image Vis. Comput. 23 (10) (2005) 921–933, https://doi.org/10.1016/j.imavis.2005.05.017.

[19] X. Fang, W. Guo, Q. Li, J. Zhu, Z. Chen, J. Yu, B. Zhou, H. Yang, Sewer pipeline fault identification using anomaly detection algorithms on video sequences, IEEE Access 8 (2020) 39574–39586, https://doi.org/10.1109/ACCESS.2020.2975887.

[20] N. Caradot, M. Riechel, M. Fesneau, N. Hernandez, A. Torres, H. Sonnenberg, E. Eckert, N. Lengemann, J. Waschnewski, P. Rouault, Practical benchmarking of statistical and machine learning models for predicting the condition of sewer pipes in Berlin, Germany, J. Hydroinf. 20 (5) (2018) 1131–1147, https://doi.org/10.2166/hydro.2018.217.

[21] L.V. Nguyen, D.T. Bui, R. Seidu, Comparison of machine learning techniques for condition assessment of sewer network, IEEE Access 10 (2022) 124238–124258, https://doi.org/10.1109/ACCESS.2022.3222823.

[22] A. Gedam, S. Mangulkar, B. Gandhi, Prediction of sewer pipe main condition using the linear regression approach, J. Geosci. Environ. Prot. 4 (5) (2016) 100–105, https://doi.org/10.4236/gep.2016.45010.

[23] J.B. Haurum, C.H. Bahnsen, M. Pedersen, T.B. Moeslund, Water level estimation in sewer pipes using deep convolutional neural networks, Water 12 (12) (2020) 3412, https://doi.org/10.3390/w12123412.

[24] X. Yin, Y. Chen, A. Bouferguene, H. Zaman, M. Al-Hussein, L. Kurach, A deep learning-based framework for an automated defect detection system for sewer pipes, Autom. Constr. 109 (2020) 102967, https://doi.org/10.1016/j.autcon.2019.102967.

[25] M. Wang, J. Cheng, Semantic segmentation of sewer pipe defects using deep dilated convolutional neural network, in: ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction 36, IAARC Publications, 2019, pp. 586–594, https://doi.org/10.22260/ISARC2019/0078.

[26] Q. Zhou, Z. Situ, S. Teng, H. Liu, W. Chen, G. Chen, Automatic sewer defect detection and severity quantification based on pixel-level semantic segmentation, Tunn. Undergr. Space Technol. 123 (2022) 104403, https://doi.org/10.1016/j.tust.2022.104403.

[27] C.H. Bahnsen, A.S. Johansen, M.P. Philipsen, J.W. Henriksen, K. Nasrollahi, T. B. Moeslund, 3d sensors for sewer inspection: a quantitative review and analysis, Sensors 21 (7) (2021) 2553, https://doi.org/10.3390/s21072553.

[28] Y. Tan, R. Cai, J. Li, P. Chen, M. Wang, Automatic detection of sewer defects based on improved you only look once algorithm, Autom. Constr. 131 (2021) 103912, https://doi.org/10.1016/j.autcon.2021.103912.

[29] J.B. Haurum, T.B. Moeslund, Sewer-ML: A multi-label sewer defect classification dataset and benchmark, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2021) 13456–13467, https://doi.org/10.1109/CVPR46437.2021.01325.

[30] D. Meijer, L. Scholten, F. Clemens, A. Knobbe, A defect classification methodology for sewer image sets with convolutional neural networks, Autom. Constr. 104 (2019) 281–298, https://doi.org/10.1016/j.autcon.2019.04.013.

[31] Y. Liu, X. Zhang, Y. Li, G. Liang, Y. Jiang, L. Qiu, H. Tang, F. Xie, W. Yao, Y. Dai, Y. Qiao, Y. Wang, VideoPipe 2022 challenge: real-world video understanding for urban pipe inspection, in: 2022 26th International Conference on Pattern Recognition (ICPR), 2022, pp. 4967–4973, https://doi.org/10.1109/ICPR56361.2022.9956055.

[32] Q. Xie, D. Li, J. Xu, Z. Yu, J. Wang, Automatic detection and classification of sewer defects via hierarchical deep learning, IEEE Trans. Autom. Sci. Eng. 16 (4) (2019) 1836–1847, https://doi.org/10.1109/TASE.2019.2900170.

[33] S.S. Kumar, D.M. Abraham, M.R. Jahanshahi, T. Iseley, J. Starr, Automated defect classification in sewer closed circuit television inspections using deep convolutional neural networks, Autom. Constr. 91 (2018) 273–283, https://doi.org/10.1016/j.autcon.2018.03.028.

[34] D. Li, A. Cong, S. Guo, Sewer damage detection from imbalanced CCTV inspection data using deep convolutional neural networks with hierarchical classification, Autom. Constr. 101 (2019) 199–208, https://doi.org/10.1016/j.autcon.2019.01.017.

[35] L.M. Dang, S. Kyeong, Y. Li, H. Wang, T.N. Nguyen, H. Moon, Deep learning-based sewer defect classification for highly imbalanced dataset, Comput. Ind. Eng. 161 (2021) 107630, https://doi.org/10.1016/j.cie.2021.107630.
[36] J.C.P. Cheng, M. Wang, Automated detection of sewer pipe defects in closed-circuit television images using deep learning techniques, Autom. Constr. 95 (2018) 155–171, https://doi.org/10.1016/j.autcon.2018.08.006.
[37] K. Chen, H. Hu, C. Chen, L. Chen, C. He, An Intelligent Sewer Defect Detection Method Based on Convolutional Neural Network, IEEE Int. Conf. Inf. Autom. 2018 (2018) 1301–1306, https://doi.org/10.1109/ICInfA.2018.8812445.
[38] M. Wang, J.C.P. Cheng, A unified convolutional neural network integrated with conditional random field for pipe defect segmentation, Comput.-Aided Civ. Infrastruct. Eng. 35 (2) (2020) 162–177, https://doi.org/10.1111/mice.12481.
[39] G.D. Finlayson, B. Schiele, J.L. Crowley, Comprehensive colour image normalization, in: Computer Vision—ECCV'98: 5th European Conference on Computer Vision Freiburg, Germany, June, 2–6, 1998 Proceedings I 5, Springer, 1998, pp. 475–490, https://doi.org/10.1007/BFb0055685.
[40] A. Badano, C. Revie, A. Casertano, W.-C. Cheng, P. Green, T. Kimpe, E. Krupinski, C. Sisson, S. Skrøvseth, D. Treanor, Consistency and standardization of color in medical imaging: a consensus report, J. Digit. Imaging 28 (2015) 41–52, https://doi.org/10.1007/s10278-014-9721-0.
[41] H. He, E.A. Garcia, Learning from imbalanced data, IEEE Trans. Knowl. Data Eng. 21 (9) (2009) 1263–1284, https://doi.org/10.1109/TKDE.2008.239.
[42] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, S. Savarese, Generalized intersection over union: A metric and a loss for bounding box regression, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 658–666, https://doi.org/10.1109/CVPR.2019.00075.
[43] A.P. Bradley, The use of the area under the ROC curve in the evaluation of machine learning algorithms, Pattern Recogn. 30 (7) (1997) 1145–1159, https://doi.org/10.1016/S0031-3203(96)00142-2.
[44] J.A. Hanley, B.J. McNeil, The meaning and use of the area under a receiver operating characteristic (ROC) curve, Radiology 143 (1) (1982) 29–36, https://doi.org/10.1148/radiology.143.1.7063747.
[45] T. Fawcett, An introduction to ROC analysis, Pattern Recogn. Lett. 27 (8) (2006) 861–874, https://doi.org/10.1016/j.patrec.2005.10.010.
[46] M. Sokolova, G. Lapalme, A systematic analysis of performance measures for classification tasks, Inf. Process. Manag. 45 (4) (2009) 427–437, https://doi.org/10.1016/j.ipm.2009.03.002.
[47] I. Ulku, E. Akagündüz, A survey on deep learning-based architectures for semantic segmentation on 2D images, Appl. Artif. Intell. 36 (1) (2022) 2032924, https://doi.org/10.1080/08839514.2022.2032924.
[48] T. Evgeniou, M. Pontil, Support Vector Machines: Theory and Applications, Advanced Course on Artificial Intelligence, Springer, 1999, pp. 249–257, https://doi.org/10.1007/3-540-44673-7_12.
[49] X. Ye, J.e. Zuo, R. Li, Y. Wang, L. Gan, Z. Yu, X. Hu, Diagnosis of sewer pipe defects on image recognition of multi-features and support vector machine in a southern Chinese city, Frontiers of, Environ. Sci. Eng. 13 (2019) 1–13, https://doi.org/10.1007/s11783-019-1102-y.
[50] M.-D. Yang, T.-C. Su, Automated diagnosis of sewer pipe defects based on machine learning approaches, Expert Syst. Appl. 35 (3) (2008) 1327–1337, https://doi.org/10.1016/j.eswa.2007.08.013.
[51] X. Zuo, B. Dai, Y. Shan, J. Shen, C. Hu, S. Huang, Classifying cracks at sub-class level in closed circuit television sewer inspection videos, Autom. Constr. 118 (2020) 103289, https://doi.org/10.1016/j.autcon.2020.103289.
[52] K. O'shea, R. Nash, An Introduction to Convolutional Neural Networks, arXiv preprint arXiv:1511.08458, 2015, https://doi.org/10.48550/arXiv.1511.08458.
[53] F.N. Iandola, S. Han, M.W. Moskewicz, K. Ashraf, W.J. Dally, K. Keutzer, SqueezeNet: AlexNet-level Accuracy With 50x Fewer Parameters and< 0.5 MB Model Size, arXiv preprint arXiv:1602.07360, 2016, https://doi.org/10.48550/arXiv.1602.07360.
[54] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2818–2826, https://doi.org/10.1109/CVPR.2016.308.
[55] D. Meijer, M. Kesteloo, A. Knobbe, Unsupervised anomaly detection in sewer images with a PCA-based framework, in: International Conference on Pattern Recognition and Artificial Intelligence (ICPRAI), 2018, pp. 354–359.
[56] J. Myrans, Z. Kapelan, R. Everson, Automatic identification of sewer fault types using CCTV footage, EPiC Ser. Eng. 3 (2018) 1478–1485, https://doi.org/10.29007/w41w.
[57] J. Myrans, Z. Kapelan, R. Everson, Using automatic anomaly detection to identify faults in sewers:(027), in: WDSA/CCWI Joint Conference Proceedings Vol.1, 2018.
[58] S.M. Khan, S.A. Haider, I. Unwala, A deep learning based classifier for crack detection with robots in underground pipes, in: 2020 IEEE 17th International Conference on Smart Communities: Improving Quality of Life Using ICT, IoT and AI (HONET), 2020, pp. 78–81, https://doi.org/10.1109/HONET50430.2020.9322665.
[59] Y. Chen, S.A. Sharifuzzaman, H. Wang, Y. Li, L.M. Dang, H.-K. Song, H. Moon, Deep learning based underground sewer defect classification using a modified RegNet, Comput. Mater. Continua 75 (3) (2023) 5455–5473, https://doi.org/10.32604/cmc.2023.033787.
[60] K.A. Joshi, D.G. Thakore, A survey on moving object detection and tracking in video surveillance system, Int. J. Soft Comput. Eng. 2 (3) (2012) 44–48.
[61] P.K. Mishra, G. Saroha, A study on video surveillance system for object detection and tracking, in: 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), IEEE, 2016, pp. 221–226.
[62] F. Pérez-Hernández, S. Tabik, A. Lamas, R. Olmos, H. Fujita, F. Herrera, Object detection binary classifiers methodology based on deep learning to identify small objects handled similarly: Application in video surveillance, Knowl.-Based Syst. 194 (2020) 105590, https://doi.org/10.1016/j.knosys.2020.105590.
[63] A. Gupta, A. Anpalagan, L. Guan, A.S. Khwaja, Deep learning for object detection and scene perception in self-driving cars: Survey, challenges, and open issues, Array 10 (2021) 100057, https://doi.org/10.1016/j.array.2021.100057.
[64] D. Fernandes, A. Silva, R. Névoa, C. Simões, D. Gonzalez, M. Guevara, P. Novais, J. Monteiro, P. Melo-Pinto, Point-cloud based 3D object detection and classification methods for self-driving applications: a survey and taxonomy, Inf. Fusion 68 (2021) 161–191, https://doi.org/10.1016/j.inffus.2020.11.002.
[65] A. Uçar, Y. Demir, C. Güzeliş, Object recognition and detection with deep learning for autonomous driving applications, Simulation 93 (9) (2017) 759–769, https://doi.org/10.1177/0037549717709932.
[66] E. Davies, The application of machine vision to food and agriculture: a review, Imaging Sci. J. 57 (4) (2009) 197–217, https://doi.org/10.1179/174313109X454756.
[67] A. Vibhute, S.K. Bodhe, Applications of image processing in agriculture: a survey, Int. J. Comput. Appl. 52 (2) (2012) 34–40, https://doi.org/10.5120/8176-1495.
[68] A.M. Roy, R. Bose, J. Bhaduri, A fast accurate fine-grain object detection model based on YOLOv4 deep neural network, Neural Comput. & Applic. 34 (5) (2022) 3895–3921, https://doi.org/10.1007/s00521-021-06651-x.
[69] Z. Li, M. Dong, S. Wen, X. Hu, P. Zhou, Z. Zeng, CLU-CNNs: Object detection for medical images, Neurocomputing 350 (2019) 53–59, https://doi.org/10.1016/j.neucom.2019.04.028.
[70] M. Baumgartner, P.F. Jäger, F. Isensee, K.H. Maier-Hein, nnDetection: a self-configuring method for medical object detection, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24, Springer, 2021, pp. 530–539, https://doi.org/10.1007/978-3-030-87240-3_51.
[71] R. Yang, Y. Yu, Artificial convolutional neural network in object detection and semantic segmentation for medical imaging analysis, Front. Oncol. 11 (2021) 638182, https://doi.org/10.3389/fonc.2021.638182.
[72] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A.C. Berg, Ssd: Single shot multibox detector, in: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14, Springer, 2016, pp. 21–37, https://doi.org/10.1007/978-3-319-46448-0_2.
[73] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: unified, real-time object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788, https://doi.org/10.1109/CVPR.2016.91.
[74] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2980–2988, https://doi.org/10.1109/ICCV.2017.324.
[75] H. Law, J. Deng, Cornernet: Detecting objects as paired keypoints, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 734–750, https://doi.org/10.1007/978-3-030-01264-9_45.
[76] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 580–587, https://doi.org/10.1109/CVPR.2014.81.
[77] R. Girshick, Fast r-cnn, Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1440–1448, https://doi.org/10.1109/ICCV.2015.169.
[78] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, Adv. Neural Inf. Proces. Syst. 28 (2015), https://doi.org/10.1109/TPAMI.2016.2577031.
[79] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications, arXiv preprint arXiv:1704.04861, 2017, https://doi.org/10.48550/arXiv.1704.04861.
[80] M. Wang, H. Luo, J.C. Cheng, Towards an automated condition assessment framework of underground sewer pipes based on closed-circuit television (CCTV) images, Tunn. Undergr. Space Technol. 110 (2021) 103840, https://doi.org/10.1016/j.tust.2021.103840.
[81] D. Shen, X. Liu, Y. Shang, X. Tang, Deep learning-based automatic defect detection method for sewer pipelines, Sustainability 15 (12) (2023) 9164, https://doi.org/10.3390/su15129164.
[82] C. Zhang, C.C. Chang, M. Jamshidi, Concrete bridge surface damage detection using a single-stage detector, Comput.-Aided Civ. Infrastruct. Eng. 35 (4) (2020) 389–409, https://doi.org/10.1111/mice.12500.
[83] J. Terven, D.-M. Córdova-Esparza, J.-A. Romero-González, A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas, Mach. Learn. Knowl. Extr. 5 (4) (2023) 1680–1716, https://doi.org/10.3390/make5040083.
[84] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, Yolov10: real-time end-to-end object detection, Adv. Neural Inf. Proces. Syst. 37 (2024) 107984–108011, https://doi.org/10.48550/arXiv.2405.14458.
[85] R. Huang, J. Pedoeem, C. Chen, YOLO-LITE: a real-time object detection algorithm optimized for non-GPU computers, in: 2018 IEEE International

Conference on Big Data (Big Data), IEEE, 2018, pp. 2503–2510, https://doi.org/10.1109/BigData.2018.8621865.
[86] K.-J. Kim, P.-K. Kim, Y.-S. Chung, D.-H. Choi, Performance enhancement of YOLOv3 by adding prediction layers with spatial pyramid pooling for vehicle detection, in: 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), IEEE, 2018, pp. 1–6, https://doi.org/10.1109/AVSS.2018.8639438.
[87] J. Redmon, A. Farhadi, YOLO9000: better, faster, stronger, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 7263–7271, https://doi.org/10.1109/CVPR.2017.690.
[88] Q. Zhou, Z. Situ, S. Teng, W. Chen, G. Chen, J. Su, Comparison of classic object-detection techniques for automated sewer defect detection, J. Hydroinf. 24 (2) (2022) 406–419, https://doi.org/10.2166/hydro.2022.132.
[89] Z. Situ, S. Teng, W. Feng, Q. Zhong, G. Chen, J. Su, Q. Zhou, A transfer learning-based YOLO network for sewer defect detection in comparison to classic object detection methods, Dev. Built Environ. 15 (2023) 100191, https://doi.org/10.1016/j.dibe.2023.100191.
[90] J. Redmon, A. Farhadi, Yolov3: An Incremental Improvement, arXiv preprint arXiv:1804.02767, 2018, https://doi.org/10.48550/arXiv.1804.02767.
[91] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2117–2125, https://doi.org/10.1109/CVPR.2017.106.
[92] S.S. Kumar, D.M. Abraham, A deep learning based automated structural defect detection system for sewer pipelines, in: ASCE International Conference on Computing in Civil Engineering 2019, American Society of Civil Engineers, Reston, VA, 2019, pp. 226–233, https://doi.org/10.1061/9780784482445.029.
[93] C.-Y. Wang, H.-Y.M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, I.-H. Yeh, CSPNet: A new backbone that can enhance learning capability of CNN, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 390–391, https://doi.org/10.1109/CVPRW50498.2020.00203.
[94] S. Liu, L. Qi, H. Qin, J. Shi, J. Jia, Path aggregation network for instance segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 8759–8768, https://doi.org/10.1109/CVPR.2018.00913.
[95] K. He, X. Zhang, S. Ren, J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, IEEE Trans. Pattern Anal. Mach. Intell. 37 (9) (2015) 1904–1916, https://doi.org/10.1109/TPAMI.2015.2389824.
[96] Z. Zekuan, H. Chunlin, Research on defect detection method of drainage pipe network based on deep learning, in: 2022 19th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), 2022, pp. 1–6, https://doi.org/10.1109/ICCWAMTIP56608.2022.10016589.
[97] M. Tan, R. Pang, Q.V. Le, Efficientdet: scalable and efficient object detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 10781–10790, https://doi.org/10.1109/CVPR42600.2020.01079.
[98] Z. Situ, S. Teng, X. Liao, G. Chen, Q. Zhou, Real-time sewer defect detection based on YOLO network, transfer learning, and channel pruning algorithm, J. Civ. Struct. Heal. Monit. 14 (1) (2024) 41–57, https://doi.org/10.1007/s13349-023-00681-w.
[99] C.-Y. Wang, A. Bochkovskiy, H.-Y.M. Liao, YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 7464–7475, https://doi.org/10.48550/arXiv.2207.02696.
[100] R. Liu, Z. Shao, Q. Sun, Z. Yu, Defect detection and 3D Reconstruction of complex urban underground pipeline scenes for sewer robots, Sensors 24 (23) (2024) 7557, https://doi.org/10.3390/s24237557.
[101] Z. Lv, S. Dong, J. He, B. Hu, Q. Liu, H. Wang, Lightweight sewer pipe crack detection method based on amphibious robot and improved YOLOv8n, Sensors 24 (18) (2024) 6112, https://doi.org/10.3390/s24186112.
[102] J. Dong, M. Liao, Defect Detection of Urban Drainage Pipeline Based on Improved YOLO-V8, in: 2024 IEEE 7th International Conference on Information Systems and Computer Aided Education (ICISCAE), IEEE, 2024, pp. 284–289, https://doi.org/10.1109/ICISCAE62304.2024.10761785.
[103] C.-Y. Wang, I.-H. Yeh, H.-Y. Mark Liao, Yolov9: learning what you want to learn using programmable gradient information, in: European Conference on Computer Vision, Springer, 2024, pp. 1–21, https://doi.org/10.48550/arXiv.2402.13616.
[104] M.R. Halfawy, J. Hengmeechai, Automated defect detection in sewer closed circuit television images using histograms of oriented gradients and support vector machine, Autom. Constr. 38 (2014) 1–13, https://doi.org/10.1016/j.autcon.2013.10.012.
[105] G. Heo, J. Jeon, B. Son, Crack automatic detection of CCTV video of sewer inspection with low resolution, KSCE J. Civ. Eng. 23 (3) (2019) 1219–1227, https://doi.org/10.1007/s12205-019-0980-7.
[106] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, S. Zagoruyko, End-to-end object detection with transformers, in: European Conference on Computer Vision – ECCV 2020, Springer, 2020, pp. 213–229, https://doi.org/10.1007/978-3-030-58452-8_13.
[107] L.M. Dang, H. Wang, Y. Li, T.N. Nguyen, H. Moon, DefectTR: End-to-end defect detection for sewage networks using a transformer, Constr. Build. Mater. 325 (2022) 126584, https://doi.org/10.1016/j.conbuildmat.2022.126584.
[108] A. Hawari, M. Alamin, F. Alkadour, M. Elmasry, T. Zayed, Automated defect detection tool for closed circuit television (cctv) inspected sewer pipelines, Autom. Constr. 89 (2018) 99–109, https://doi.org/10.1016/j.autcon.2018.01.004.
[109] D. Li, Q. Xie, Z. Yu, Q. Wu, J. Zhou, J. Wang, Sewer pipe defect detection via deep learning with local and global feature fusion, Autom. Constr. 129 (2021) 103823, https://doi.org/10.1016/j.autcon.2021.103823.
[110] Q. Yuan, Y. Shi, M. Li, A review of computer vision-based crack detection methods in civil infrastructure: progress and challenges, Remote Sens 16 (16) (2024) 2910, https://doi.org/10.3390/rs16162910.
[111] A. Bochkovskiy, C.-Y. Wang, H.-Y.M. Liao, Yolov4: Optimal Speed and Accuracy of Object Detection, arXiv preprint arXiv:2004.10934, 2020, https://doi.org/10.48550/arXiv.2004.10934.
[112] T.-C. Su, M.-D. Yang, T.-C. Wu, J.-Y. Lin, Morphological segmentation based on edge detection for sewer pipe defects on CCTV images, Expert Syst. Appl. 38 (10) (2011) 13094–13114, https://doi.org/10.1016/j.eswa.2011.04.116.
[113] G. Pan, Y. Zheng, S. Guo, Y. Lv, Automatic sewer pipe defect semantic segmentation based on improved U-Net, Autom. Constr. 119 (2020) 103383, https://doi.org/10.1016/j.autcon.2020.103383.
[114] Y. Li, H. Wang, L.M. Dang, M. Jalil Piran, H. Moon, A robust instance segmentation framework for underground sewer defect detection, Measurement 190 (2022) 110727, https://doi.org/10.1016/j.measurement.2022.110727.
[115] L.J. Sartor, A.R. Weeks, Morphological operations on color images, J. Electron. Imaging 10 (2) (2001) 548–559, https://doi.org/10.1117/1.1353199.
[116] M. Zeng, J. Li, Z. Peng, The design of top-hat morphological filter and application to infrared target detection, Infrared Phys. Technol. 48 (1) (2006) 67–76, https://doi.org/10.1016/j.infrared.2005.04.006.
[117] T.-C. Su, M.-D. Yang, Application of morphological segmentation to leaking defect detection in sewer pipelines, Sensors 14 (5) (2014) 8686–8704, https://doi.org/10.3390/s140508686.
[118] M.-D. Yang, T.-C. Su, Segmenting ideal morphologies of sewer pipe defects on CCTV images for automated diagnosis, Expert Syst. Appl. 36 (2, Part 2) (2009) 3562–3573, https://doi.org/10.1016/j.eswa.2008.02.006.
[119] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3431–3440, https://doi.org/10.1109/TPAMI.2016.2572683.
[120] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, Springer, 2015, pp. 234–241, https://doi.org/10.1007/978-3-319-24574-4_28.
[121] M. Li, M. Li, Q. Ren, H. Li, L. Xiao, X. Fang, PipeTransUNet: CNN and Transformer fusion network for semantic segmentation and severity quantification of multiple sewer pipe defects, Appl. Soft Comput. 159 (2024) 111673, https://doi.org/10.1016/j.asoc.2024.111673.
[122] S. Woo, J. Park, J.-Y. Lee, I.S. Kweon, Cbam: Convolutional block attention module, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 3–19, https://doi.org/10.1007/978-3-030-01234-2_1.
[123] V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: a deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (12) (2017) 2481–2495, https://doi.org/10.1109/TPAMI.2016.2644615.
[124] M. He, Q. Zhao, H. Gao, X. Zhang, Q. Zhao, Image segmentation of a sewer based on deep learning, Sustainability 14 (11) (2022), https://doi.org/10.3390/su14116634.
[125] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs, IEEE Trans. Pattern Anal. Mach. Intell. 40 (4) (2018) 834–848, https://doi.org/10.1109/TPAMI.2017.2699184.
[126] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 801–818, https://doi.org/10.1007/978-3-030-01234-2_49.
[127] L.M. Dang, H. Wang, Y. Li, L.Q. Nguyen, T.N. Nguyen, H.-K. Song, H. Moon, Lightweight pixel-level semantic segmentation and analysis for sewer defects using deep learning, Constr. Build. Mater. 371 (2023) 130792, https://doi.org/10.1016/j.conbuildmat.2023.130792.
[128] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask r-cnn, Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2961–2969, https://doi.org/10.1109/ICCV.2017.322.
[129] D. Bolya, C. Zhou, F. Xiao, Y.J. Lee, Yolact: Real-time instance segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 9157–9166, https://doi.org/10.1109/ICCV.2019.00925.
[130] D. Ma, H. Fang, N. Wang, H. Zheng, J. Dong, H. Hu, Automatic defogging, deblurring, and real-time segmentation system for sewer pipeline defects, Autom. Constr. 144 (2022) 104595, https://doi.org/10.1016/j.autcon.2022.104595.
[131] X. Wang, R. Zhang, T. Kong, L. Li, C. Shen, Solov2: Dynamic and fast instance segmentation, Adv. Neural Inf. Proces. Syst. 33 (2020) 17721–17732.
[132] R. Alshawi, M.M. Ferdaus, M. Abdelguerfi, K. Niles, K. Pathak, S. Sloan, Imbalance-aware culvert-sewer defect segmentation using an enhanced feature pyramid network, IEEE Trans. Syst. Man Cybern.: Syst. (2024) 1–16, https://doi.org/10.1109/TSMC.2025.3579706.
[133] Y. Li, Y. Yang, Y. Liu, F. Zhong, H. Zheng, S. Wang, Z. Wang, Z. Huang, A novel method for semantic segmentation of sewer defects based on StyleGAN3 and improved Deeplabv3+, J. Civ. Struct. Heal. Monit. (2025) 1–18, https://doi.org/10.1007/s13349-025-00919-9.
[134] NASSCO'S Pipeline Assessment Certification Program. NASCCO: NASCCO, https://nassco.org/resource/nasscos-pipeline-assessment-certification-program/?scLang=en#:~:text=An%20introduction%20to%20NASSCO's%20Pipeline,

assessment%20coding%20of%20underground%20infrastructure, 2023 (accessed 2024).
[135] S. Daher, T. Zayed, A. Hawari, Defect-based condition assessment model for sewer pipelines using fuzzy hierarchical evidential reasoning, J. Perform. Constr. Facil. 35 (1) (2021) 04020142, https://doi.org/10.1061/(ASCE)CF.1943-5509.0001554.
[136] P. Jia, Y. Liao, Q. Zhao, M. He, M. Guo, Defect severity assessment model for sewer pipeline based on automated pipe calibration, J. Pipeline Syst. Eng. Pract. 14 (3) (2023) 04023025, https://doi.org/10.1061/JPSEA2.PSENG-1454.
[137] M. Elmasry, A. Hawari, T. Zayed, Defect based risk assessment model for prioritizing inspection of sewer pipelines, in: Pipelines 2018, American Society of Civil Engineers Reston, VA, 2018, pp. 1–9, https://doi.org/10.1061/9780784481653.001.
[138] C.L. Nguyen, A. Nguyen, J. Brown, T. Byrne, B.T. Ngo, C.X. Luong, Optimising concrete crack detection: a study of transfer learning with application on nvidia jetson nano, Sensors (Basel, Switzerland) 24 (23) (2024) 7818, https://doi.org/10.3390/s24237818.