Contents lists available at ScienceDirect

Construction and Building Materials

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

Highlights

Lightweight pixel-level semantic segmentation and analysis for sewer defects using deep learning

Construction and Building Materials xxx (xxxx) xxx

L. Minh Dang, Hanxiang Wang, Yanfen Li, Le Quan Nguyen, Tan N. Nguyen, Hyoung-Kyu Song, Hyeonjoon Moon*

- A manually collected sewer defect segmentation dataset that contains over 11,124 images.
- An efficient sewer defect segmentation framework based on DeepLabv3+.
- A frame reduction is introduced to reduce the computational complexity.
- Automatic report generation module to support real-life applications.
- Defect severity analysis based on the NASSCO PACP program.

Contents lists available at ScienceDirect

Construction and Building Materials

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

Lightweight pixel-level semantic segmentation and analysis for sewer defects using deep learning

L. Minh Dang [c,d,a], Hanxiang Wang [b], Yanfen Li [b], Le Quan Nguyen [b], Tan N. Nguyen [e], Hyoung-Kyu Song [a], Hyeonjoon Moon [b,*]

[a] Department of Information and Communication Engineering, and Convergence Engineering for Intelligent Drone, Sejong University, Seoul, Republic of Korea
[b] Department of Computer Science and Engineering, Sejong University, Seoul, Republic of Korea
[c] Institute of Research and Development, Duy Tan University, Da Nang 550000, Viet Nam
[d] Faculty of Information Technology, Duy Tan University, Da Nang 550000, Viet Nam
[e] Department of Architectural Engineering, Sejong University, 209 Neungdong-ro, Gwangjin-gu, 05006, Seoul, Republic of Korea

ARTICLE INFO

Keywords:
Deep learning
Semantic segmentation
Sewer network
Defect segmentation
Severity analysis

ABSTRACT

The underground sewer network is a vital public infrastructure in charge of large-scale wastewater collection and treatment. Complex defects can occur in sewer pipes due to various internal and external factors, which increase the demand for frequent inspection. Previous defect detection research mainly depended on manual inspection, which is tedious, costly, and error-prone. This study suggests an automatic pixel-level sewer defect segmentation framework based on DeepLabV3+, which can recognize the defect's type, location, geometric information and severity. The impacts of various backbones and pre-processing methods on the model's performance were carefully evaluated. In addition, four state-of-the-art segmentation models (U-Net, SegNet, PSPNet, and FCN) were compared with the presented model to demonstrate its superiority. The experimental results revealed that the DeepLabV3+ with the Resnet-152 backbone structure efficiently identified ten defect types under challenging conditions. The obtained mean pixel accuracy and mean intersection over union (IoU) were 0.97 and 0.68, respectively. In terms of severity analysis, it was revealed that the framework outputs were consistent with the NASSCO pipeline assessment certification program (PACP). In addition, during the testing process, the proposed frame reduction algorithm only required about 16% of the original time required to process an input video. Finally, with a generated detailed report for an inspection video, the suggested framework can offer a decision-making base for more precise and efficient defect inspection and maintenance.

1. Introduction

Sewerage systems are essential public infrastructure in metropolitan cities worldwide, which can significantly impact community assets and lives. For instance, the total length of the sewer network in Korea was reported to be 156,257 km in 2018 [1]. Even though underground sewer systems have a long lifespan by design, environmental factors or poor management can lead to the appearance of defects. Common defects, such as cracks, tree roots, joint dislocation, and blockages, are primary causes of pipe structural damage and may even lead to severe functional failure, which is time-consuming and requires high maintenance costs [2]. Some sample defects are described in Fig. 1(b). As a result, periodic sewer inspection is the most critical measure to prevent pipe deterioration, and there have been increasing demands for accurate and efficient sewer inspection systems [3].

1.1. CCTV-based sewer defect inspection

Multiple inspection approaches have been introduced for sewer investigation in the last decades. A recent review categorized them into vision-based, ultrasonic-based, acoustic-based, and electromagnetic-based technologies and discussed each technology in detail [4]. Among the technologies, vision-based closed-circuit television (CCTV) was considered a well-accepted and cost-effective technique [5,6]. As illustrated in Fig. 1(a), cameras and lighting equipment mounted on a carrier, such as a robot, can effectively be utilized to inspect sewer pipelines' internal conditions.

A huge number of CCTV videos were collected after the data collection process, requiring human inspectors to manually check each video for defects. This inspection process is time-intensive and error-prone due to the inspectors' conditions and experience [7]. Deep learning

* Corresponding author.
E-mail address: hmoon@sejong.ac.kr (H. Moon).

https://doi.org/10.1016/j.conbuildmat.2023.130792
Received 6 November 2022; Received in revised form 8 February 2023; Accepted 17 February 2023

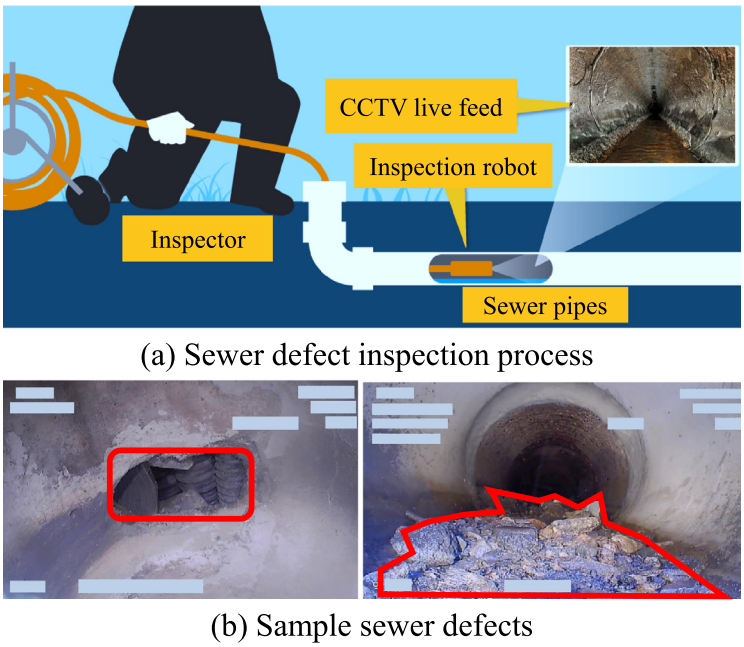


Fig. 1. Sewer defect inspection example. (a) introduces the sewer inspection process that uses a robot to record videos and (b) displays some defects in a sewer network.

(DL) models that automatically learn crucial features from data, have been extensively adopted for sewer inspection-related topics and have established state-of-the-art performances in recent decades. In general, DL has been implemented for three main sewer inspection topics with distinctive degrees of complexity, including defect classification, defect detection, and defect segmentation [3].

It is noticeable that there is also a lack of evaluation standards in current CCTV-based sewer inspection research [3]. In-depth training programs and uniform evaluation manuals, such as the NASSCO pipeline assessment certification program (PACP) [8] and the manhole assessment certification program (MACP), can partially solve these drawbacks.

1.1.1. Defect classification

Initially, convolutional neural network (CNN) models were adopted for sewer defect classification [7,9]. For instance, Hassan et al. demonstrated a defect classification system using the AlexNet structure, which supported video caption recognition [10]. Dang et al. proposed a deep learning-based sewer defect classification model that was robust against imbalanced data problems by implementing an ensemble approach [9]. Even though previous research has revealed the effectiveness and robustness of the various sewer defect classification models, an input image was classified into a specific defect type without revealing the exact location of the defects.

1.1.2. Defect detection

On the other hand, defect detection returns both the defect type and bounding boxes indicating the exact location of detected defects for an input image, which is more informative than defect classification [7,11]. Recent defect detection research has mainly applied state-of-the-art object detection structures, such as you only look once (YOLO), faster region-based CNN (Faster R-CNN), and the latest transformer-based approaches [12,13]. Notably, Oh et al. demonstrated that the latest YOLOv5 models achieved satisfying detection performance and could be implemented in real-time applications [13]. However, the defect detection approach still fails to deliver crucial geometrical data of the defects, such as area, width, length, and shape, which is essential for the defect severity assessment.

1.1.3. Defect segmentation

Although numerous studies have previously been proposed to identify sewer defects, most of them belonged to classification [2,9], or detection [12,13]. Semantic defect segmentation is the latest defect investigation trend, which segments various defects by classifying a defect class for each pixel of an input image. The pixels of different types of defects (e.g., obstacles, cracks, protruding) can be classified precisely. In addition, geometrical details of segmented defects can be used to perform in-depth defect analyses [3]. Mainstream deep segmentation structures, including fully convolutional network (FCN) [14], U-Net [15,16], SegNet [17], and DeepLab [18,19], have been increasingly applied to perform sewer defect investigation.

Previous studies mainly applied state-of-the-art segmentation networks to the defect segmentation systems, such as the SegNet model, which was fine-tuned by Han et al. [17] and U-Net's structure, which was improved by Pan et al. [16]. In a more recent study, Li et al. improved the state-of-the-art segmenting objects by locations (SOLOv2) model by introducing a novel backbone structure to perform segmentation for six different defect types [20]. However, the number of defect types was limited, and defect severity analysis was ignored. Moreover, defect severity analysis, an essential process for sewer inspection, was overlooked [21]. Finally, these studies did not compare the proposed models with other segmentation approaches when implemented in the same framework [3].

1.2. Main contributions

In order to fully solve these weaknesses, there is a desire for fully-automated, practical and robust methods for sewer defect detection and evaluation [9,18]. In consideration of the drawbacks of previous studies, this research introduces a novel sewer defect segmentation based on the DeepLabV3+ model that can accept both CCTV videos and images as input. The proposed framework also supports an in-depth analysis of the segmented defects, including defect severity. In summary, the main contributions are as follows.

- Previous defect segmentation research has been computationally intensive because they perform segmentation on every frame extracted from a CCTV video. This study introduces a novel frame reduction approach that recognizes the video's captions data to significantly lower the number of frames to be predicted.

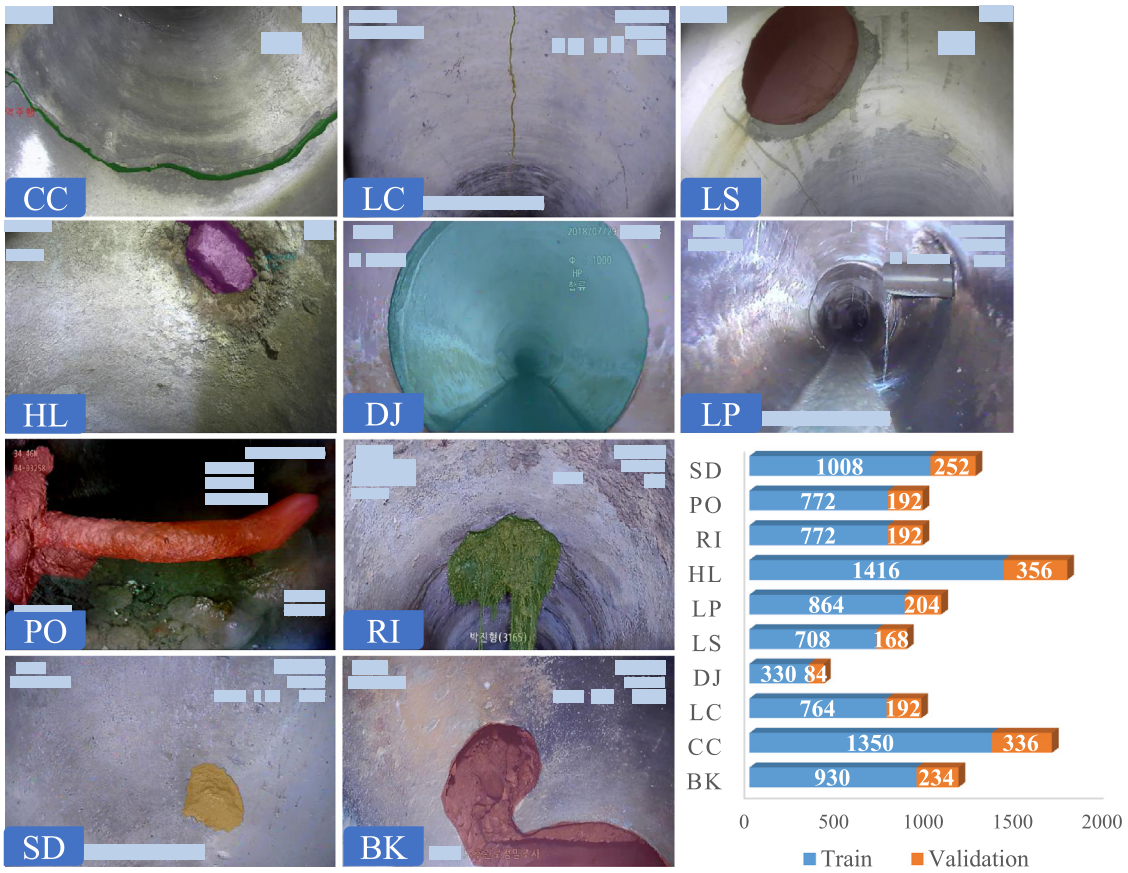


Fig. 2. Sample images, where the defects are highlighted, and a description of the number of training and validation images for ten defect types proposed in this study.

- To the best of our knowledge, the latest sewer defect segmentation framework can only detect a small number of defect classes, such as five [16] and six [20,22]. In this study, we introduce a high-quality sewer defect segmentation dataset carefully annotated by experts that include ten types of sewer defects.
- A comprehensive experiment is carried out in order to compare the proposed model with other well-known segmentation models, such as U-Net, SegNet, and FCN. In addition, various backbone structures and loss functions (9 combinations in total) are implemented to find out the best combination (accuracy and speed).
- The defect severity analysis, which was ignored in previous studies, is also added to the proposed system in order to enable a fully automatic defect evaluation process.

2. Dataset description

2.1. Defect segmentation dataset

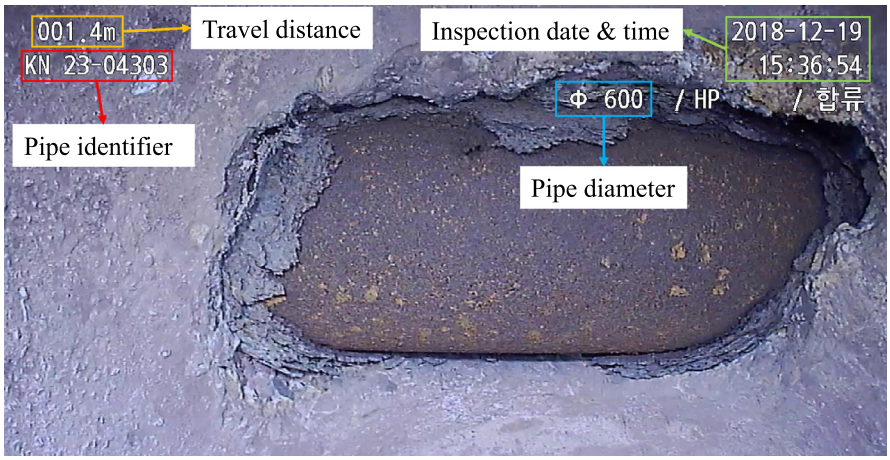
A total of 3699 frames containing defects were extracted from 60 closed-circuit television (CCTV) inspection videos recorded at various underground sewer manholes across South Korea. The time of the videos ranged from 3 to 10 min with a fixed frame rate of 30 frames per second (FPS) of various pipe lengths and inner conditions. The CCTV videos were collected by an advanced inspection robot carrying a high-resolution 1/3-inch SONY Exmor CMOS camera on its head that supports 360° rotation and up/down motions. In addition, the head of the robot was equipped with a powerful light-emitting diode (LED) bulb to enable it to record videos in the sewer pitch-dark environment.

As described in Fig. 2, ten different types of sewer defects are defined as follows.

- Broken pipe (BK): The inner pipe is partially or completely damaged. BK is a serious defect type requiring timely actions from experts.
- Longitudinal crack (LC): Concrete foundation settling effect that causes a vertical or diagonal crack emerges on the sewer's inner wall surface.
- Circumferential crack (CC): Pressure on the outside of the sewer's inner walls causes a crack to emerge parallel to the wall's horizontal axis. CC is considered a more serious defect compared to CL because it can cause permanent damage to the pipe foundation.
- Displaced joint (DJ): A slight dislocation of the pipe joints.
- Lateral protruding (LP): Sewerage connection protrudes to the primary sewer manhole.
- Lateral sealing (LS): Gap in sewerage connection but does not extend to the outer wall.
- Surface damage (SD): Minor damages on the sewer's inner wall surface, corrosion by environmental factors or abrasion.
- Root intrusion (RI): Sewer pipe diameter reduction due to the intrusion of roots.
- Hole (HL): Pipe diameter reduction due to the intrusion of roots into the sewer manhole.
- Permanent obstruction (PO): Obstacles, such as hanging objects, mortar, and welding byproducts, puncture through the sewer's surface or occupy part of the sewer manhole.

These defect categories were picked because they have been considered the most common type of sewer defect in Korea [3]. Most of these defect classes also appear in the PACP [23].

A group of 5 civil engineers from an inspection company participated in a two-month image annotation task, where each individual annotated about 20 images per day. The data annotation process adopted



Note: Crucial captions are travel distance, pipe identifier, inspection date, and pipe diameter.

Fig. 3. Essential captions on a single frame of an inspection video.

an open-source Labelme annotation tool programmed in Python.[1] No image cropping method was implemented to pick the relevant areas. At the end of the labeling process, a total of 3708 images were annotated. The images were resized to the input size requirement of the DeepLabV3+ network, which is 512 × 512 pixels. In order to increase the number of training images, three data augmentation methods (horizontal flip and random translation of ±20 pixels for the X-axis and Y-axis) were applied to the processed defect segmentation dataset. The number of images after this step increased to 11,124 images. Finally, the dataset was randomly split into two sets of 8914 images (80% of the data) used for training and validation and 2210 images (20% of the data) used for testing purposes. The chart in Fig. 2 shows the number of annotated images for each type.

2.2. Caption recognition dataset

As displayed in Fig. 3, captions added on every frame of a sewer inspection video provide essential information regarding the inspection and management of the sewers.

The explanation for some crucial captions is as follows.

- Pipe identifier: The unique identifier for a sewer pipe.
- Travel distance: The distance that the robot has moved inside a specific sewer pipe.
- Pipe diameter: The diameter of a sewer pipe.
- Inspection date & time: The inspection date (year/month/day) and current time (hour/minute/second).

A total of 1500 caption patches were extracted from the inspection videos. They were manually annotated for the text detection task. The dataset was split using the 8:1:1 ratio with 1200 images for training, 150 for validation and 150 for testing.

3. System overview

Fig. 4 describes the main components of the suggested sewer defect segmentation framework.

Comprehensive descriptions of each process are defined as follows.

- Preprocessing: Frames extracted from an input sewer video can be affected by external environments that lead to problems, such as hazy and low-light conditions. Therefore, various image preprocessing methods are applied to enhance the quality of the extracted frames and ensure the defect segmentation framework performance.
- Frame reduction: A novel approach to significantly reduce the computational intensive defect segmentation task by recognizing the caption information and selecting only crucial defect frames based on the nature of the CCTV inspection video.
- Defect segmentation: DeepLabV3+, a state-of-the-art semantic segmentation model, is tested with various backbone structures and loss functions and trained to learn abstract defect features under complex conditions to perform sewer defect segmentation.
- Defect severity analysis: Defect severity is useful and essential information for inspectors to determine whether a defect is minor or serious in order to commit timely action.
- Report generation: This module takes the output of defect segmentation and severity analysis agents to provide a comprehensive report showing the crucial information related to the segmented defects.

4. Methodology

4.1. Image pre-processing

Naturally, the environments inside a sewer pipe are always pitch dark and have high humidity due to continuous water vapour, which lead to various issues for the collected CCTV videos, such as low-light images, illumination, and foggy data. As a result, various algorithms were implemented to preprocess the dataset to enhance its visual quality.

First, contrast enhancement is an effective method to enhance image quality and reduce the low-light issue. Previous image contrast enhancement approaches have performed poorly on noisy images because the appearance of unwanted noise significantly affected the proper distribution of pixel intensities in the contrast-enhanced image. In order to overcome the issue, the contrast enhancement framework proposed by [24] was applied to process the extracted sewer frames. This approach first implemented the mean-shift algorithm to reduce irrelevant pixels and noise from the input images. After that, the Moth Swarm algorithm (MSA) was applied to redistribute the pixel intensities of

[1] https://github.com/wkentaro/labelme

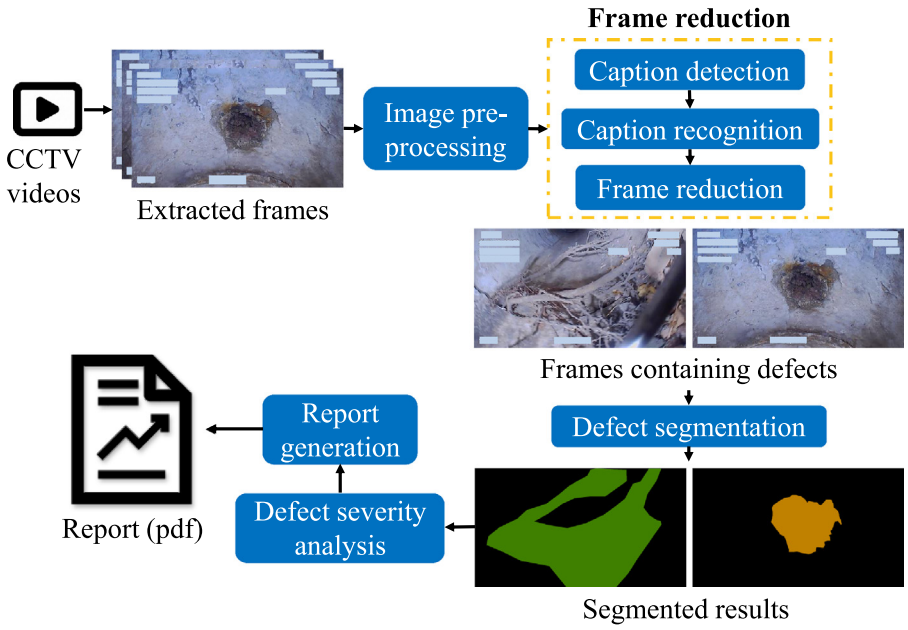


Fig. 4. Detailed description of the suggested sewer defect segmentation framework for an input CCTV video.

the preprocessed histogram to maximize the Kullback–Leibler entropy (KL-entropy) value.

After that, a pretrained lightweight DehazeNet model (LD-Net) [25] was fine-tuned to efficiently enhance recorded CCTV videos' quality at a low computational cost. All hyperparameters were set as suggested in the original article [25]. Different from previous methods, the LD-Net model simultaneously estimates the atmospheric light and transmission map based on an atmospheric scattering model to effectively lower the processing time and computational cost. Moreover, the color distortion issue is prevented in the output images using color visibility restoration module. LD-Net demonstrated high-quality denoising results without affecting the image's original brightness, and can be applied to real-time applications.

Fig. 5 displays the preprocessing outputs of 4 random input frames extracted from sewer inspection videos. The outputs of the two preprocessing methods, including LD-Net and MSA, show a significant enhancement of quality compared to the original inputs. For example, it is impossible to identify defects in the low-light scenario or captions in the illumination background, even with human eyes. However, a substantial improvement in the image quality after the preprocessing process enables a clear observation of defects or captions appearing in the preprocessed images. In addition, the preprocessing module does not introduce any artifacts to the output images.

4.2. Frame reduction

During the testing process, the users can feed either an image or CCTV video into the proposed framework. In the case of a video, after extracting every frame from the video, previous studies fed all the frames into the system, which is computationally complex and inefficient because most frames did not contain defects. Therefore, this study proposes a frame reduction module aiming to select only frames containing defects to be processed by the proposed defect segmentation framework based on the caption information. The two main processes are caption detection & recognition and frame reduction.

4.2.1. Caption detection & recognition

Transfer learning was applied to a pretrained YOLOv5 model to perform caption detection [26]. YOLOv5 includes three main components, the backbone, neck and head. Cross-Stage-Partial Darknet (CSPDarknet) was implemented as the backbone to obtain crucial features from the input images. After that, the neck part uses a path aggregation network (PANet) to create a feature pyramid network in order to perform feature aggregation and forward it to the head module for detecting the captions. The head performs the text detection by applying anchor boxes on extracted features and generates final output vectors, including class, class probabilities, bounding boxes, and size.

After the text detection step, candidate regions that possibly contain the captions are extracted. After that, a state-of-the-art four-stage scene text recognition framework, which includes transformation, feature extraction, sequence modeling, and prediction, was applied to recognize the detected captions [27]. Initially, a thin-plate spline (TPS) transformation is implemented in order to increase the diversity of aspect ratios of input text lines to produce normalized images. Next, in the feature extraction stage, a ResNet-152 model [28] is implemented to obtain abstract features from the normalized images in order to estimate the character from each receptive field. In the sequence modeling stage, the extracted feature maps are then converted into a sequence of features. Contextual information within a sequence of characters is then extracted by analyzing each frame of the sequence. Finally, in the prediction process, the attention-based sequence prediction (Attn) is implemented to estimate the flow of information within the identified features sequence to output the final character sequence.

4.2.2. Frame reduction

Important captions of the CCTV videos, which were detected and recognized in the previous section, are used in this section to select frames containing defects to be segmented. During the sewer pipe inspection, inspectors are required to follow predetermined steps when a defect is detected, as follows.

- The robot is driven near the defect position.
- The inspectors remotely control the mounted camera on the head of the robot to thoroughly capture the defect at various directions and angles.
- The process lasts for a few seconds, depending on the defect severity.

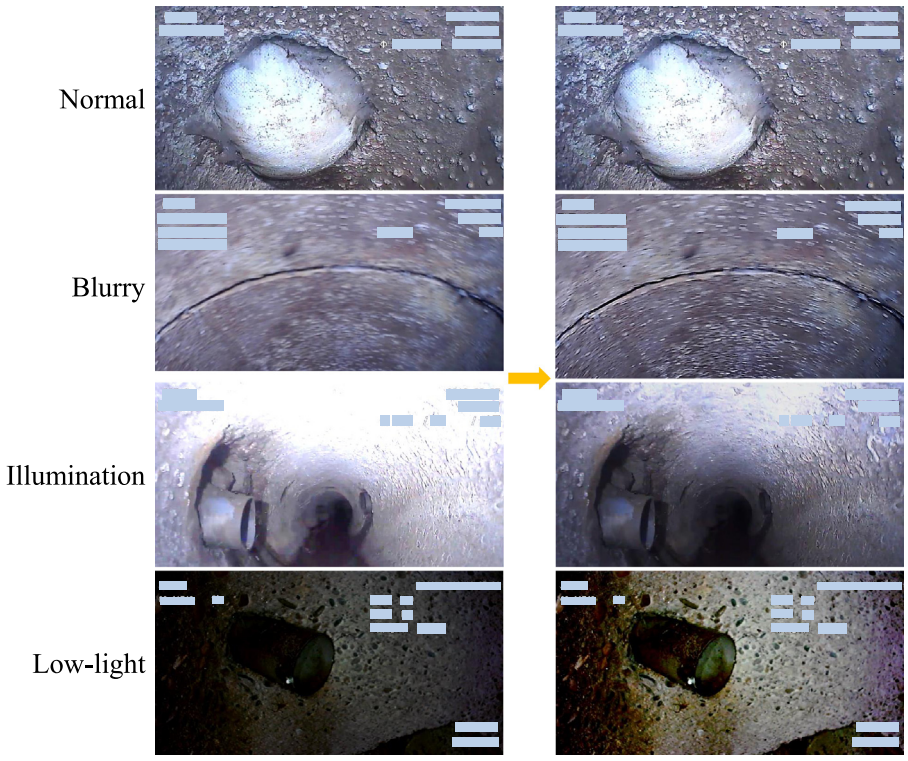
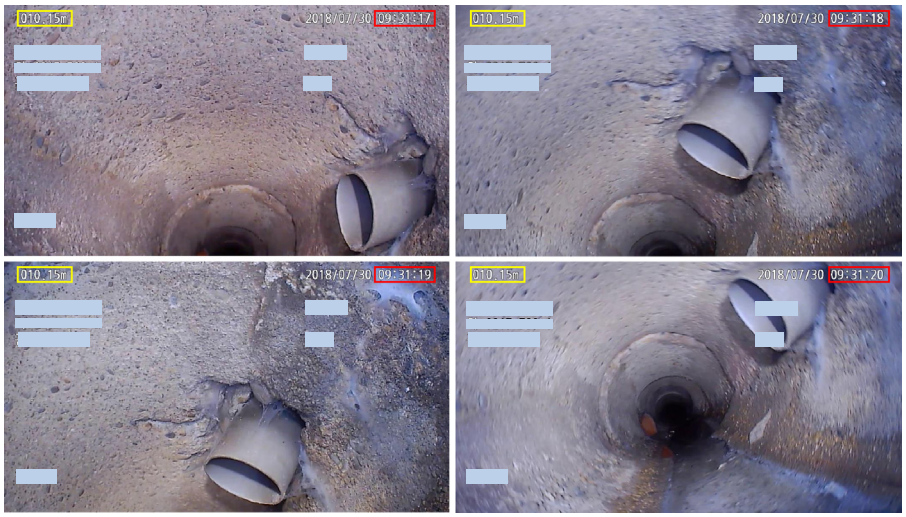


Fig. 5. Visualization of the raw sewer images and output images after implementing the pre-processing process.



Note: The yellow box indicates that the "travel distance" caption (010.15m) remains unchanged, whereas the "inspection time" highlighted in red increases.

Fig. 6. A sample of four lateral protruding defect samples that were captured when the robot stopped in order to observe the defect for 4s from 09:31:17 to 09:31:20. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

As shown in Fig. 6, the robot travel distance caption remains unchanged during this process, even when the inspection time changes.

Therefore, a frame reduction algorithm based on this characteristic is introduced in order to significantly reduce irrelevant frames being fed into the defect segmentation framework. Initially, all frames are obtained from an input CCTV video. After that, the suggested caption detection and recognition module from Section 4.2.1 is implemented to detect and recognize the travel distance and inspection time captions for each frame. These data are fed into the frame reduction algorithm. The difference between the beginning and ending inspection time for

the same "travel distance" caption is equal to or longer than 3000 ms, indicating that the robot is stopped to inspect a defect. Only the first and last frames are extracted and forwarded to the defect segmentation framework to improve computational efficiency during the whole period. A collection of defect frames will be extracted at the end of this process.

Algorithm 1 Frame reduction

Input: CCTV video V,
Output: Frame with defect F_d
Initialize: Frames $F \leftarrow null$, frame with defect $F_d \leftarrow null$, temporary data $TMP \leftarrow null$, temporary frames $F_{tmp} \leftarrow null$, moving distance $d \leftarrow -1$, current time $ct \leftarrow -1$, reset flag $r \leftarrow 1$
Extract all frames F from V
while F not null **do** ▷ Stop when all extracted frames were processed
    Get frame f
    Initialize: begin time $bt \leftarrow -1$, end time $et \leftarrow -1$
    $d \leftarrow ocr(f)$ ▷ Detect and recognize the moving distance caption
    $ct \leftarrow ocr(f)$ ▷ Detect and recognize the current time caption

    if $r = 1$ **then**
        $F_{tmp} \leftarrow f$ ▷ Add frame f to F_{tmp}
        $bt, et \leftarrow ct$
        $TMP \leftarrow \{d, bt, et\}$
        $r \leftarrow 0$
    else
        Select distance d_{TMP} from TMP
        if $d_{TMP} = d$ **then**
            $et \leftarrow ct$
            $TMP \leftarrow et$ ▷ Update TMP with a new et
        else
            $bt, et \leftarrow TMP$
            $totaltime \leftarrow et - bt$
            if $totaltime \geq 3000$ **then**
                $F_{tmp} \leftarrow f$ ▷ Extract the last frame containing the defect
                $F_d \leftarrow F_{tmp}$ ▷ Frames with defects are added to the final array
                $TMP, F_{tmp} \leftarrow null$; $r \leftarrow 1$;
Return F_d

4.3. Defect segmentation

A fine-tuned DeepLabV3+ [29] is implemented to learn discriminative features of ten primary defect types under challenging conditions in order to perform pixel-level defect segmentation. It is currently a well-known multi-class semantic segmentation model that has been increasingly utilized to perform defect segmentation in the latest structural engineering-related studies. DeepLabV3+ extends the previous DeepLabV3 by using an encoder–decoder structure. Fig. 7 shows the overall training process of the DeepLabV3+ in this study. Raw defect features are extracted using well-known DCNN backbone networks and fed into the DeepLabV3+. The encoder is in charge of reducing the spatial sizes of feature maps, extracting high-level semantic information, and eliminating some unnecessary artifacts. In DeepLabV3+, DeepLabV3 plays the role of the encoder. The decoder then boosts the learned features and gradually recovers the spatial information via the up-sampling operation.

For the encoder module, Atrous Spatial Pyramid Pooling (ASPP) is applied to downsize feature maps' dimensions and encode multi-scale contextual information. ASPP comprises 4 atrous convolution layers ($Conv_1$, $Conv_2$, $Conv_3$ and $Conv_4$) with different kernel sizes in order to enable it to obtain the filter's field of view. After that, all extracted multi-scale features are concatenated, go through one pooling layer, and are sent to the decoder.

The decoder module predicts segmentation outputs having the same size as the original images based on low-scale feature maps obtained by the encoder. The feature maps received from the encoder are first bilinearly upsampled by 4 using the transposed convolution operation in order to expand the feature maps' dimensions. These feature maps are fused with low-level features extracted directly from the backbone network via a special 1×1 convolution operation to enrich the extracted features. Finally, the initial spatial dimensions of the feature maps are slowly recovered through concatenation, convolution and up-sampling operations to produce the final segmentation outputs.

4.4. Defect severity analysis

Previous studies have ignored the severity analysis process of the defects after detecting them, which plays an essential role in indicating the sewer pipe's condition and provides inspectors with additional information to decide which defect needs to be maintained in time. This study performs severity analysis based on the PACP standard [8].

PACP assigns the structural defect grade based on further deterioration or failure risk, ranging from 1 to 5, with 1 indicating a minor defect (complete failure is unlikely to happen in the long term) and 5 indicating a severe defect (was ruined or will probably fail in the next five years). As displayed in Table 1, even though slight differences exist in naming some defect classes between this study and the PACP, most of the defect classes are interchangeable.

5. Experimental results

This section investigates the performance and robustness of the suggested framework thoroughly through several experiments with different settings. Section 5.1 shows the hardware and programming environment where the proposed framework was developed. After that, Section 5.2 explains the evaluation metrics computed to assess the framework. Moreover, the hyperparameters required for all models are also described. Finally, a series of experiments are carried out to investigate the framework's performance and robustness.

5.1. Implementation details

All experiments, including training and inference processes, were done on a Linux machine (Ubuntu 18.04) equipped with an Nvidia Tesla V100 32 GB. The framework was developed using Python programming language and PyTorch,[2] an open-source deep learning framework.

In the following subsections, five well-known segmentation models, including DeeplabV3+ [29], U-Net [30], SegNet [31], PSPNet [32], and FCN [33] were adapted to perform crack segmentation on the ten-class defect dataset introduced in this study. The dataset was labeled according to the scheme defined in the COCO benchmark dataset. In order to guarantee the fairness of the following experiments, all segmentation models used abstract features extracted from pre-trained standard backbone networks on the ImageNet dataset, including VGG19, Xception, ResNet152, and Densenet201. All models were trained for 80 epochs with the batch size fixed to 16. Moreover, the stochastic gradient descent optimization function with an initial learning rate of 0.01 and a momentum of 0.9 was applied to train the model.

[2] https://pytorch.org/

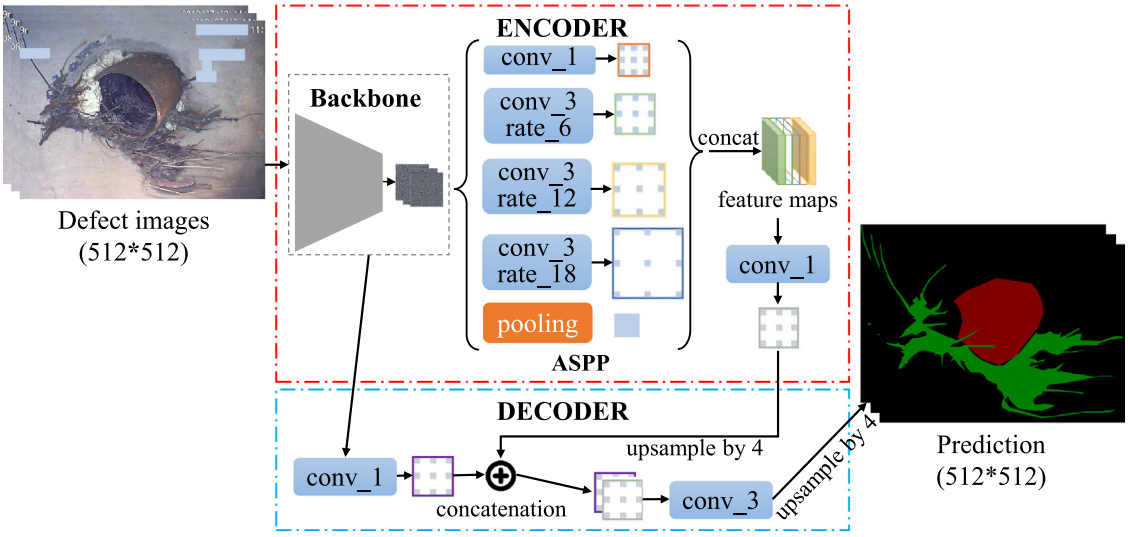


Fig. 7. Main processes of the sewer defect segmentation framework based on the DeepLabV3+ model.

Table 1
Defect grade comparison between this study and the PACP standard.

This study	PACP	Structural grade
Broken pipe	Pipe failures (broken)	5
Circumferential crack	Crack (circumferential)	1
Longitudinal crack	Crack (longitudinal)	2
Displaced joint	Joint (offset)	1
Surface damage	Surface damage	5
Protruding lateral	Obstacles/Obstructions (Pipe protruding through the wall)	3
Root intrusion	Roots (lateral)	3
Permanent obstruction	Obstacles/Obstruction (object)	2 (≤10%), 3 (≤20%), 4 (≤30%), 5 (>30%)
Hole	Pipe failures (hole)	5
Lateral sealing	External pipe in sewer	2 (≤10%), 3 (≤20%), 4 (≤30%), 5 (>30%)

5.2. Evaluation metrics

This study implements six standard segmentation evaluation metrics in order to evaluate the performance of the defect segmentation models. The short description and equation for each evaluation metric are described as follows.

- Mean intersection over union ($mIoU$): While IoU quantifies the percent overlap between the ground truth mask and predicted mask, m$mIoU$ indicates the class average IoU.

$$IoU = \frac{|A_p \cap A_g|}{|A_p \cup A_g|} \tag{1}$$

$$\text{mIoU} = \frac{\text{IoU}}{\text{K}} \tag{2}$$

where A_p and A_g is the area predicted and ground truth pixels/objects, correspondingly;

- Mean pixel accuracy (mPA) denotes the ratio of accurately segmented pixels to the total number of pixels in the image.

$$mPA = \frac{1}{K+1} \sum_{I=0}^{K} \frac{p_{ii}}{\sum_{j=0}^{K} p_{ij}} \tag{3}$$

where K is the number of class, p_{ii} corresponds to the total number of true positives for class i. p_{ij} is the total number of pixels labeled as class j.

- F-measure: reflects a model's ability to segment different defect types, computed based on the average between the recall and the precision.

$$\text{F-measure} = 2 \times \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

- Inference time: provides information regarding the model segmentation speed.

5.3. Image preprocessing

In order to demonstrate the necessity of the pre-processing approach, the comparison of the segmentation performances of DeeplabV3+ [29], U-Net [30], SegNet [31], PSPNet [32], and FCN [33] using the ResNet-152 backbone with and without the pre-processing method was carried out on the validation dataset.

Table 2 reveals the performances of the 5 models with and without the pre-processing process. The segmentation models trained with pre-processed images showed significantly better F-measure and mPA

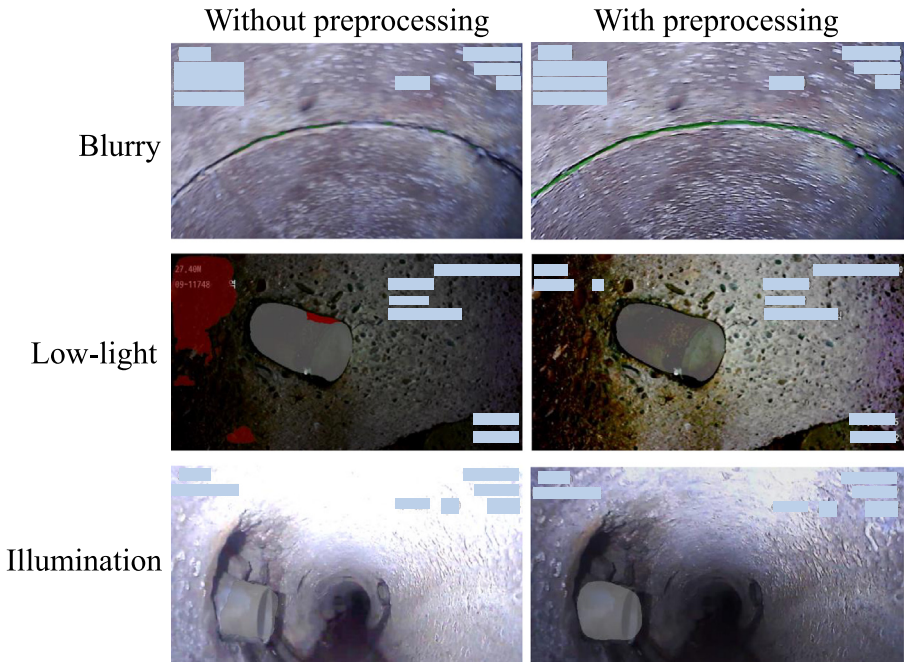


Fig. 8. Segmentation outputs of raw and pre-processed images for three typical CCTV challenging environments, including blurring, low-light, and illumination.

Table 2
Cross-model segmentation performance validation with and without implementing the pre-processing module.

Approach	Architecture	F-measure	mPA (%)
Without pre-processing	DeeplabV3+ (Our)	0.59	89.3
	U-Net [30]	0.07	85.2
	SegNet [31]	0.09	82.4
	PSPNet [32]	0.55	87.5
	FCN [33]	0.52	88.4
With pre-processing	DeeplabV3+ (Our)	**0.64**	**97.9**
	U-Net [30]	0.08	86.4
	SegNet [31]	0.09	88.1
	PSPNet [32]	0.61	96.8
	FCN [33]	0.57	95.5

scores than those trained with original images. Overall, the mPA scores showed a notable improvement between 1 to 9%. For the DeeplabV3+ model, in particular, the pre-processing approach allowed the mPA to be improved to 97.9% from the original 89.3%. Therefore, the pre-processing process is crucial for applications based on CCTV videos because it can improve sewer defect segmentation results remarkably.

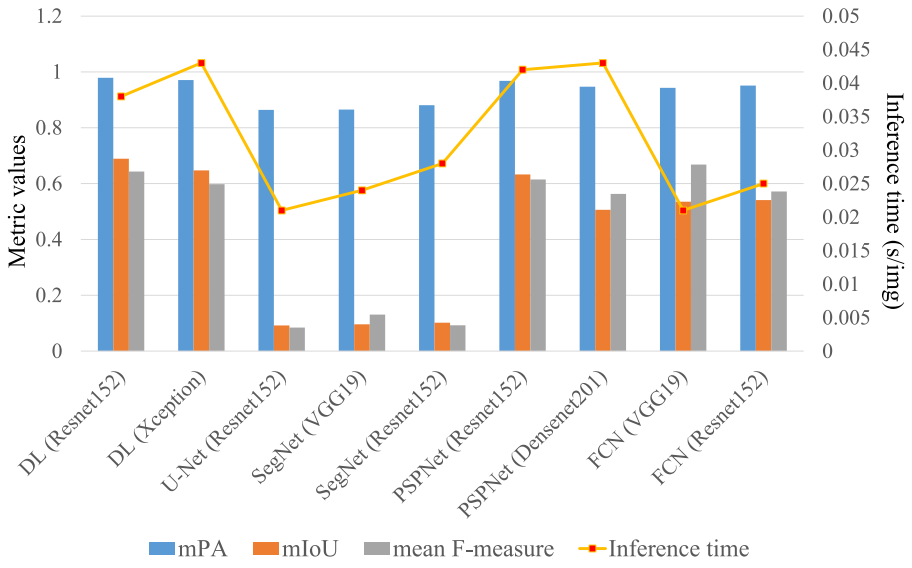
Fig. 8 demonstrates the effectiveness of the preprocessing procedure on the defect segmentation performance for three common scenarios appearing in CCTV videos, including blurring, low light, and illumination. Without the preprocessing step, the noise negatively affected the segmentation results. For example, the DeeplabV3+ model failed to recognize the CC defect in the blurring scenario or mistakenly recognized noise as a BK defect in the low-light scenario. When the environment was dark, and the contrast between the defects and the background was unclear, there was a high probability of false segmentations. However, the model segmented all defects correctly after they were preprocessed.

5.4. Defect segmentation performance analysis

In this section, various backbone networks were implemented to thoroughly validate the five segmentation models' performance on the proposed dataset after applying the pre-processing process. In summary, nine common model variations, including DeepLabV3+ (ResNet 152), DeepLabV3+ (Xception), U-Net (ResNet152), SegNet (ResNet 152), SegNet (VGG19), PSPNet (ResNet152), PSPNet (Densenet201), FCN (VGG19), and FCN (ResNet152), were examined. The segmentation results based on standard evaluation metrics (i.e. mPA, mloU, mean F-measure, and inference time) of all the network variations are described in Fig. 9.

Overall, the Resnet backbone showed better performance compared to other backbone networks. Two variations of the DeepLabV3+ model show the best segmentation performance for ten defect types. Except for the longer inference time, the DeepLabV3+ (ResNet152) outperformed other variations and obtained the best segmentation performances in all evaluation metrics. The mPA and mIoU of the DeepLabV3+ (ResNet152) were at 0.979 and 0.689, respectively, falling into the satisfying range for the segmentation topic [34]. In contrast, the mPA and mIoU of the SegNet and SegNet were low, which indicates that the network failed to learn representative features to perform the defect segmentation properly during the training phase. Therefore, this study does not recommend the U-Net and SegNet models for sewer defect segmentation.

The variations of PSPNet and FCN models showed comparable segmentation performances to the DeepLabV3+ model. The FCN model is recommended for applications focusing on speed over performance because its inference speed is almost two times faster than DeepLabV3+ and PSPNet. The DeepLabV3+(Resnet152) variation was considered the best model for sewer defect segmentation using the proposed dataset. Therefore, Fig. 10 displayed the accuracy and loss curves on the validation dataset to demonstrate the effectiveness of the training process and network convergence of the DeepLabV3+(Resnet152) model. Generally, the model achieved stable convergence during the training



Note: DL stands for DeepLabV3+ model and s/img indicates seconds per image.

Fig. 9. Reported validation results for nine segmentation model variations, which include mPA, mIoU, mean F-measure, and inference time.

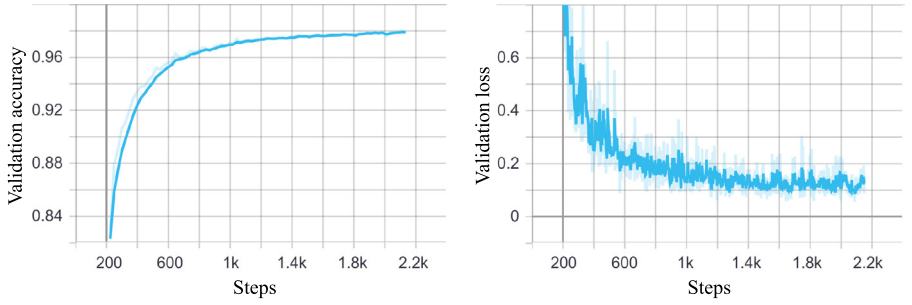


Fig. 10. Validation accuracy and loss curves of the DeepLabV3+ model using the ResNet152 backbone network.

process. The model's validation accuracy and loss curves change significantly before step 600 to approximately 94% and 0.2, respectively. After that, the curves change gradually and stop at about 97% accuracy and 0.15 loss. Therefore, the model showed good generalization ability and can deliver reliable statistics for the following experiments.

The proposed framework is also compared with models reported in previous studies in order to provide a general view of the progress of sewer defect segmentation research. Table 3 describes several features of previous sewer defect semantic segmentation research, including model, dataset, performance, and segmentation speed. Previous studies mainly relied on state-of-the-art semantic segmentation models, such as U-Net, PSPNet, and FCN, to perform defect segmentation. Even though mIoU and F-measure values were relatively lower compared to previous sewer defect segmentation studies [16,35], this study covered a significantly higher number of defect types (10 types) compared to previous studies (3–5 types) and contained the biggest number of images (11,097 images). Some defect types (such as permanent obstruction and broken pipes) were complicated because they could be mistakenly recognized as other defects (i.e. lateral sealing, root intrusion) and even overlap. Finally, the preprocessing module was implemented in the proposed framework in order to improve defect segmentation performance. The segmentation results and speed should only be used as a reference because the datasets used for training these models, the hardware, and the programming environment was different.

Table 3
Segmentation performance comparison between the proposed framework and latest defect segmentation studies.

Study	Dataset size	mIoU	Speed (FPS)
DilaSeg-CRF [35]	3 classes (1885 images)	0.84	9
PipeUNet [16]	4 classes (3654 images)	0.76	32
PSPNet [22]	5 classes (480 images)	0.53	20
DeepLabV3+ (Our)	10 classes (11097 images)	0.69	26

5.5. Qualitative evaluation

This section compares the best variation segmentation results quantitatively obtained from the previous section, which are DeepLabV3+ (Resnet152), PSPNet(Resnet152), U-Net(Resnet152), FCN(VGG19), and SegNet(VGG19). As shown in Fig. 11, generally, DeepLabV3+ showed the best segmentation results in terms of visual quality for ten defect types. For instance, the predicted output of PO using the DeepLabV3+ model was almost similar to the ground truth. Demonstratively, the

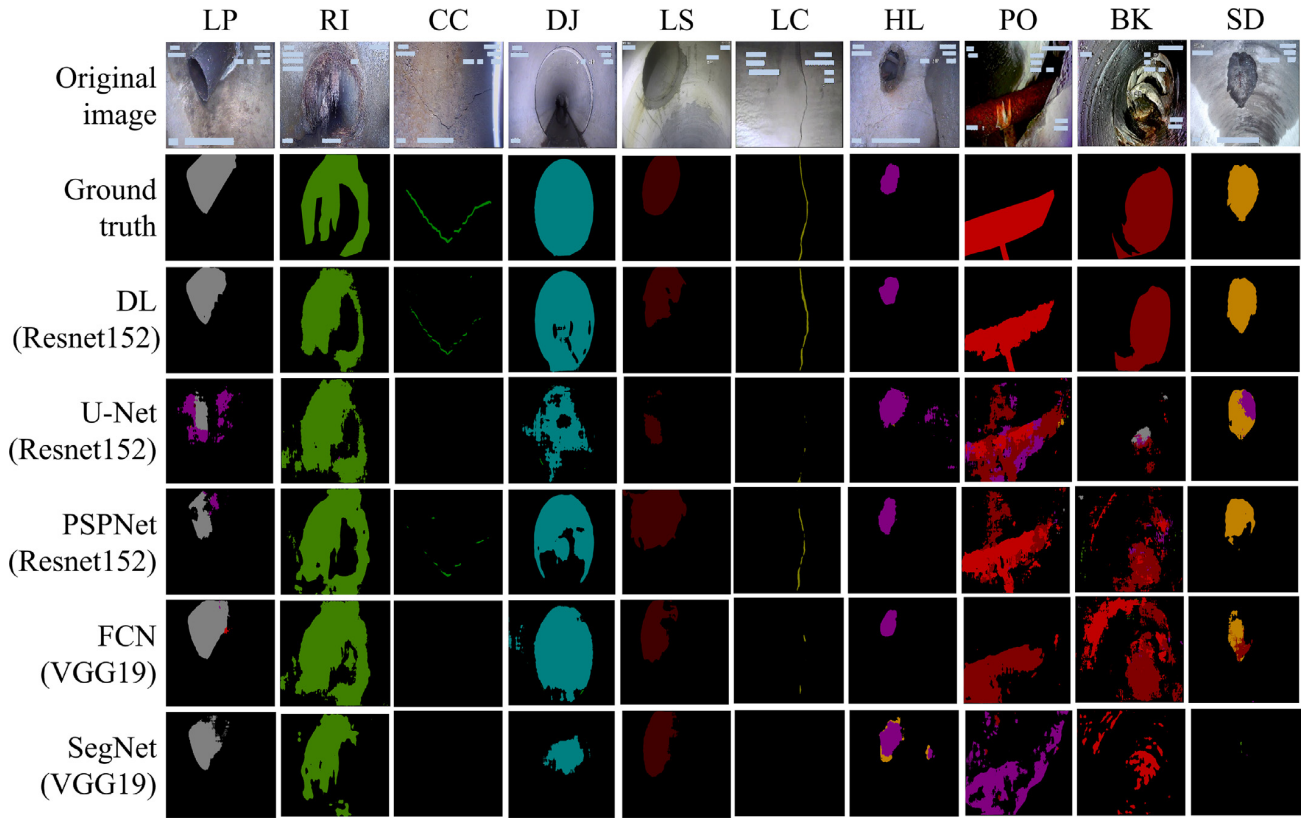


Fig. 11. Qualitative evaluation of the defect segmentation process for the ten defect types. The order is as follows: original images, ground truth images, and segmented masks produced by the best variations of each segmentation model.

model could recognize the small pipe on the bottom corner connecting to the main pipe.

Even though the segmented outputs of FCN and PSP models were scratchy, they can be considered acceptable for sewer defect segmentation. On the other hand, the SegNetand U-Net models failed to recognize complex defects, such as PO, BK, and CC. The two models could not find the defect boundaries and thus led to noisy segmented images. Too many defect regions were missed, making it challenging to perform defect severity analysis based on their segmentation results.

Despite the overall good segmentation results, poor/inaccurate segmentation results for challenging cases can still be witnessed during the testing process. Fig. 12 offers two cases where the model showed poor performance due to several potential causes. The first probable reason is the defect position, where a defect was influenced by other defects, such as an overlap between defects or defects near each other (Fig. 12 (Case 1)). Another reason is the similarity of geometric features of some defects, such as HL, BK, and PO, causing mixed segmentation predictions (Fig. 12 (Case 2)).

5.6. Caption detection & recognition and frame reduction analysis

The YOLOv5 was trained using the caption dataset explained previously in Section 4.2.1. The recorded validation mAP was 0.99, demonstrating that it correctly detected the captions with a good detection rate. Fig. 13 displays some text detection results using the YOLOv5 model. It can be seen that the model successfully detected all captions in the input images. After that, the detected captions were fed into a pretrained TPS-ResNet-BiLSTM-Attn model [27] in order to recognize the captions efficiently.

Table 4 demonstrates the effectiveness of the frame reduction algorithm on 10 randomly selected CCTV videos that range from 3 to 8 min. Without the frame reduction algorithm, all frames extracted from the input CCTV video need to be processed by the preprocessing and segmentation modules, which cost a significantly longer time.

It takes 0.33 s and 0.038 s for a single frame to be processed by the preprocessing module and the defect segmentation model, respectively. If the frame reduction algorithm is carried out, the time required for text detection is approximately 0.02 s and 0.032 s for text recognition. In summary, it takes approximately 0.368 s to process a frame without the frame reduction algorithm and costs 0.42 s if the frame reduction algorithm is implemented. Even though the algorithm requires more time and computational power, the number of frames was significantly reduced.

Take video ID number 1 as an example. The framework needs approximately 41 min to process 6701 frames ($= 6701 * 0.368$). On the other hand, it takes only 6 min to complete the defect investigation process using the frame reduction algorithm ($= 6701 * 0.052 + 22 * 0.368$). Table 4 reveals that the frame reduction algorithm saved, on average, about 86% of the computational time compared to when all frames were fed into the model.

5.7. Defect severity analysis and report generation

This section investigates the effectiveness of the report generation module, which is essential for real-life sewer inspection applications. The proposed framework automatically segments defects appearing in an input CCTV video, analyzes the segmented defect severity, and returns an inspection report. The reports generated by the proposed framework were then compared with ground truth inspection reports made manually by experts, who inspect the CCTV videos in order to identify defects and place them in the report.

Fig. 14 shows an inspection report sample produced by the suggested framework.

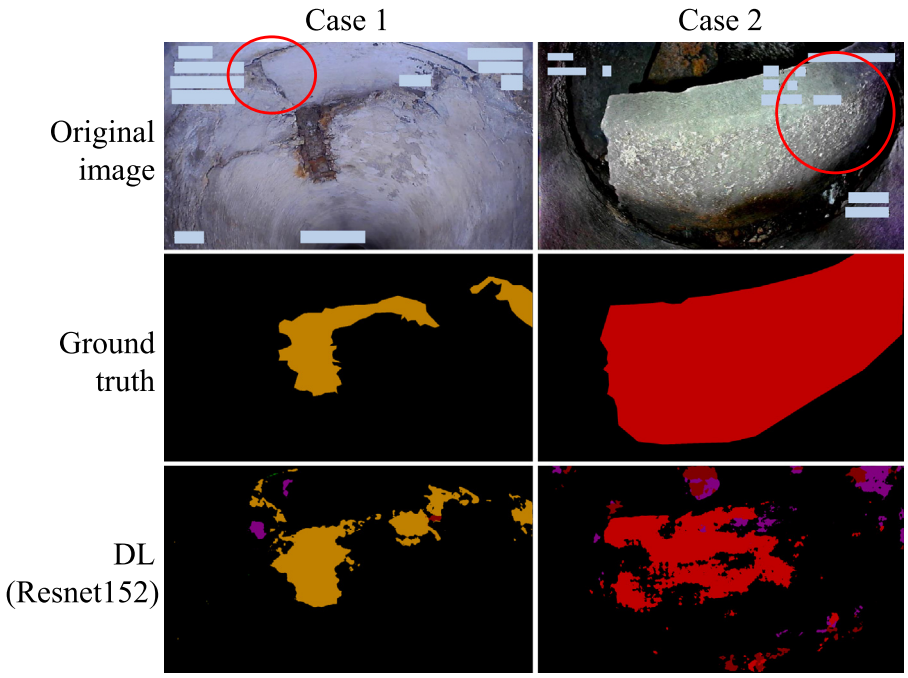


Fig. 12. Defect segmentation outputs of the DeepLabV3+ based model for challenging cases.

Fig. 13. Text detection outputs using the trained YOLOv5 model.

Important information includes visualized position of the robot in the sewer pipe, segmented defect images with the exact location highlighted, defect type, and defect severity score. As explained in Section 4.4, the defect severity score is the structural grade extracted from the PACP standard.

In order to demonstrate the effectiveness of the report generation module, five inspection videos were randomly selected to obtain automatically generated reports using the proposed framework. Fig. 15 compares the reports generated by the proposed framework and manual reports for five inspection videos.

In general, 84 defects were identified and reported in the automatically generated reports compared to 77 defects documented in the manual reports. The reasons that led to the differences between the generated and manual reports were a small number of additional defects were detected by the proposed framework, such as 05 – 06072 (1 additional defect), 05 – 06350 (2 additional defects), 05 – 06351 (5 additional defects), which was not reported in the manual reports. For example, Fig. 16 shows a BK defect (red bounding box) identified by the report generation module that did not appear in the manual report for the 05 – 06072.

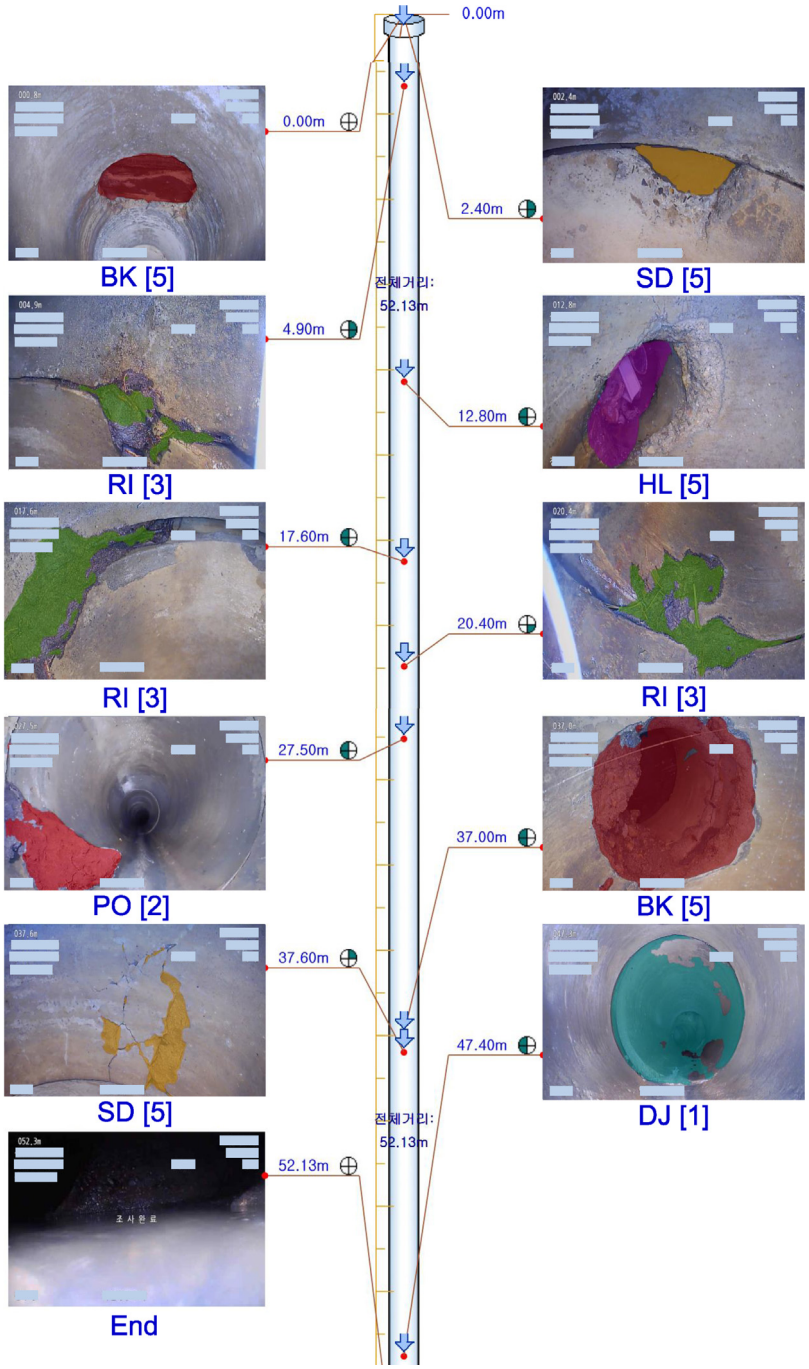


Fig. 14. A sample of the automatically generated inspection report for an input sewer inspection video using the proposed framework.

Assuming additional defects that the model detected as false positive, the overall accuracy of the generated reports was approximately 93.5%, where 72/77 defects were correctly detected.

6. Discussion

This study suggested a novel sewer defect segmentation framework based on the DeepLabV3+ that added various modules compared to previous techniques. The preprocessing module (Section 5.3), which includes MSA and LD-Net models, significantly improved the mPA of the DeepLabV3+ model by 8.6%. Other models also benefited from the preprocessing module and witnessed improvements in the segmentation performance. The main reason for the improvement was visualized and explained using Fig. 8, where noise from the raw images could cause the model to make false segmentations. Even though this module requires additional computation power and time, it can be easily disabled or enabled depending on the applications.

Five state-of-the-art segmentation models, including DeepLabV3+, U-Net, SegNet, PSPNet, and FCN, were trained using the proposed defect segmentation dataset with several combinations of backbone networks in order to select the best combination for the collected dataset. The DeepLabV3+ model (ResNet-152 backbone) achieved the highest mPA and IoU scores at 97.9% and 0.69, respectively. In addition, the DeepLabV3+ also surpassed previous state-of-the-art sewer defect segmentation research, including PipeUNet, PSPNet, and DilaSeg-CRF, demonstrating its superiority. Section 5.5 qualitatively evaluated the segmentation results of the five models for the ten defect types. Overall, the DeepLabV3+ showed satisfactory segmentation results that

Table 4
Processing time demanded by the framework with and without using the frame reduction algorithm for ten randomly selected CCTV videos.

Video ID	Video length	Travel distance	All frames		Frame reduction	
			Total frames	Processing time	Total frames	Processing time
1	224 s	0-52 m	6701	41 min	22	6 min
2	334 s	0-68 m	10002	61 min	38	9 min
3	336 s	0-66 m	10064	61 min	18	9 min
4	293 s	0-48 m	8775	53 min	30	7 min
5	301 s	0-49 m	9013	55 min	12	8 min
6	258 s	0-46 m	7734	47 min	16	6 min
7	485 s	0-73 m	14539	89 min	26	13 min
8	237 s	0-42 m	7098	43 min	18	6 min
9	375 s	0-42 m	11237	69 min	22	10 min
10	375 s	0-46 m	8244	50 min	12	7 min

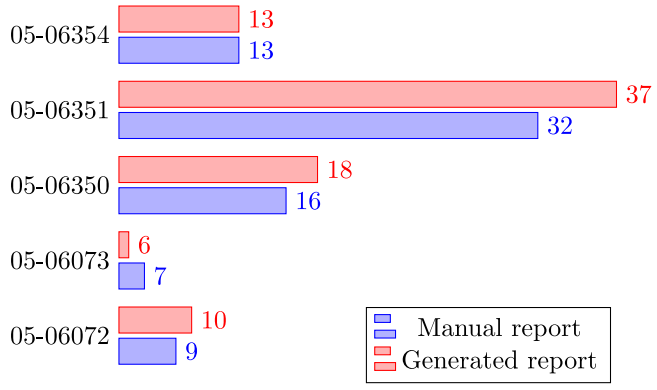


Fig. 15. A comparison between the number of defects identified by the proposed framework compared to that from the manual reports for five input videos.

resembled the ground truths. However, as displayed in Fig. 12, the DeepLabV3+ achieved poor segmentation performance for some challenging cases, such as overlap between defects or defects with similar geometric features, such as HL, BK, and PO.

Finally, the experiments in Section 5.6 on ten randomly selected CCTV videos revealed that the proposed frame reduction algorithm effectively reduced the processing time of an input CCTV video to 16% of the original processing time. Even though the frame reduction algorithm effectively saved computation time and power for the proposed defect detection framework without losing the system performance, it is only valid if the CCTV robot is stopped every time a defect appears, even for minor defects (such as cracks). This is the major drawback of the proposed algorithm compared to the conventional methods, which perform defect detection using all extracted frames. The proposed algorithm can be applied if the applications focus on reducing processing time with a limited reduction in defect detection rate and on detecting severe defects that need urgent attention.

7. Conclusions and future works

This research presents a novel deep learning-based defect segmentation framework for sewer CCTV videos. First, a total of 11,124 images for ten defect classes were manually created from CCTV inspection videos. The corresponding annotations were then manually annotated by professionals. Several variations of the five state-of-the-art defect segmentation models were trained on the proposed dataset to verify the defect segmentation performance. Unlike most previous defect segmentation research, which could only segment sewer defects, this study also determined the severity grade of detected defects. Finally, a frame reduction algorithm based on recognizing captions on the frames was introduced to reduce the computational complexity during the testing process.

The recorded results from various experiments demonstrated that the DeepLabV3+ model (Resnet-152 backbone) achieved the highest segmentation performance with mPA of 97% and mIoU of 0.68 for ten types of defects. Moreover, the mPA value of the model increased significantly from 89% to 97% if the preprocessing module was applied. The proposed framework also outperformed existing sewer defect segmentation models, including PipeUNet [16], PSPNet [22], and DilaSeg-CRF [35]. Finally, the suggested system demonstrated that it could determine the defect severity effectively based on the PACP manual and improve the computational efficiency due to the introduced frame reduction algorithm.

Even though the dataset presented in this study contains more defect types than previous sewer defect segmentation research, more defect types, such as lining cracks, water intrusion, and silty debris, can be supplemented for more accurate inspection. Moreover, the service life of any pipeline can be determined if additional sensor data are provided along with the inspection videos, which can greatly lower the processing time needed to process segmented defects to determine their severity. Finally, the suggested sewer defect segmentation system can be further improved in terms of performance and robustness.

CRediT authorship contribution statement

L. Minh Dang: Conceptualization, Methodology, Writing – review & editing. **Hanxiang Wang:** Methodology, Writing – review & editing. **Yanfen Li:** Visualization, Investigation. **Le Quan Nguyen:** Writing – review & editing. **Tan N. Nguyen:** Visualization, Investigation. **Hyoung-Kyu Song:** Funding acquisition. **Hyeonjoon Moon:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2020R1A6A1A03038540) and National Research Foundation of Korea (NRF) grant funded by the Korea government, Ministry of Science and ICT (MSIT) (2021R1F1A1046339) and by the Technology development Program (RS-2022-00156456) funded by the Ministry of SMEs and Startups (MSS, Korea).

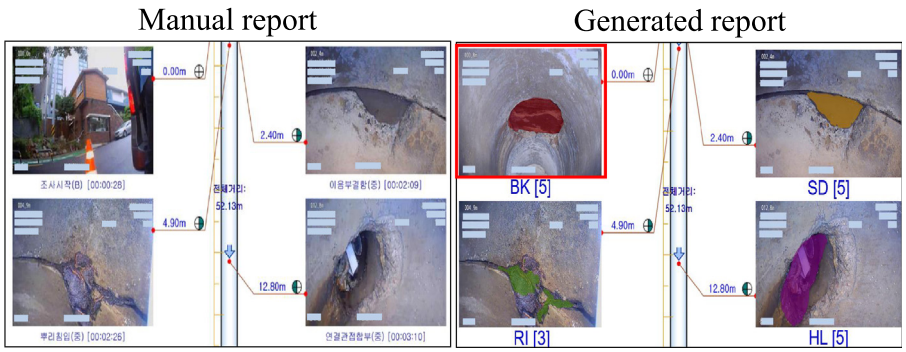


Fig. 16. A sample of an additional defect that was identified by the report generation module that does not appear in the manual report. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

References

[1] K.M. of Environments, Statics of Sewerage Works in Korea, 2022, https://www.hasudoinfo.or.kr/. (Accessed 21 September 2022).

[2] J.B. Haurum, T.B. Moeslund, Sewer-ml: A multi-label sewer defect classification dataset and benchmark, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 13456–13467.

[3] 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.

[4] F. Tscheikner-Gratl, N. Caradot, F. Cherqui, J.P. Leitão, M. Ahmadi, J.G. Langeveld, Y. Le Gat, L. Scholten, B. Roghani, J.P. Rodríguez, et al., Sewer asset management–state of the art and research needs, Urban Water J. 16 (9) (2019) 662–675.

[5] T.N. Nguyen, J. Lee, L. Dinh-Tien, L. Minh Dang, Deep learned one-iteration nonlinear solver for solid mechanics, Internat. J. Numer. Methods Engrg. 123 (8) (2022) 1841–1860.

[6] T.N. Nguyen, L.M. Dang, J. Lee, P.V. Nguyen, Load-carrying capacity of ultra-thin shells with and without cnts reinforcement, Mathematics 10 (9) (2022) 1481.

[7] 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.

[8] P. Assessment, Certification Program (pacp) Manual, Reference Manual, 2001.

[9] 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.

[10] 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.

[11] S. Moradi, T. Zayed, F. Golkhoo, Review on computer aided sewer pipeline defect detection and condition assessment, Infrastructures 4 (1) (2019) 10.

[12] 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.

[13] C. Oh, L.M. Dang, D. Han, H. Moon, Robust sewer defect detection with text analysis based on deep learning, IEEE Access 10 (2022) 46224–46237.

[14] Z. Tong, D. Yuan, J. Gao, Z. Wang, Pavement defect detection with fully convolutional network and an uncertainty framework, Comput.-Aided Civ. Infrastruct. Eng. 35 (8) (2020) 832–849.

[15] L.M. Dang, H. Wang, Y. Li, Y. Park, C. Oh, T.N. Nguyen, H. Moon, Automatic tunnel lining crack evaluation and measurement using deep learning, Tunnel. Undergr. Space Technol. 124 (2022) 104472.

[16] G. Pan, Y. Zheng, S. Guo, Y. Lv, Automatic sewer pipe defect semantic segmentation based on improved U-Net, Autom. Constr. 119 (2020) 103383.

[17] M. He, Q. Zhao, H. Gao, X. Zhang, Q. Zhao, Image segmentation of a sewer based on deep learning, Sustainability 14 (11) (2022) 6634.

[18] Y. Pan, L. Zhang, Dual attention deep learning network for automatic steel surface defect segmentation, Comput.-Aided Civ. Infrastruct. Eng. 37 (11) (2022) 1468–1487.

[19] H. Fu, B. Fu, P. Shi, An improved segmentation method for automatic mapping of cone karst from remote sensing data based on deeplab v3+ model, Remote Sens. 13 (3) (2021) 441.

[20] Y. Li, H. Wang, L.M. Dang, M.J. Piran, H. Moon, A robust instance segmentation framework for underground sewer defect detection, Measurement 190 (2022) 110727.

[21] D. Minh, H.X. Wang, Y.F. Li, T.N. Nguyen, Explainable artificial intelligence: A comprehensive review, Artif. Intell. Rev. 55 (5) (2022) 3503–3568.

[22] 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, Tunnel. Undergr. Space Technol. 123 (2022) 104403.

[23] H. Khaleghian, Y. Shan, P. Lewis, Development of a quality assurance process for sewer pipeline assessment and certification program (pacp) inspection data, in: Pipelines 2017, 2017, pp. 360–369.

[24] A. Luque-Chang, E. Cuevas, M. Pérez-Cisneros, F. Fausto, A. Valdivia-Gonzalez, R. Sarkar, Moth swarm algorithm for image contrast enhancement, Knowl.-Based Syst. 212 (2021) 106607.

[25] H. Ullah, K. Muhammad, M. Irfan, S. Anwar, M. Sajjad, A.S. Imran, V.H.C. de Albuquerque, Light-dehazenet: A novel lightweight cnn architecture for single image dehazing, IEEE Trans. Image Process. 30 (2021) 8968–8982.

[26] N. Gandhewar, S. Tandan, R. Miri, Deep learning based framework for text detection, in: 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks, ICICV, IEEE, 2021, pp. 1231–1236.

[27] J. Baek, G. Kim, J. Lee, S. Park, D. Han, S. Yun, S.J. Oh, H. Lee, What is wrong with scene text recognition model comparisons? dataset and model analysis, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 4715–4723.

[28] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[29] 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.

[30] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.

[31] T. Chen, Z. Cai, X. Zhao, C. Chen, X. Liang, T. Zou, P. Wang, Pavement crack detection and recognition using the architecture of segnet, J. Ind. Inf. Integr. 18 (2020) 100144.

[32] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2881–2890.

[33] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.

[34] S. Minaee, Y.Y. Boykov, F. Porikli, A.J. Plaza, N. Kehtarnavaz, D. Terzopoulos, Image segmentation using deep learning: A survey, IEEE Trans. Pattern Anal. Mach. Intell..

[35] M. Wang, J.C. 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.