Deep Learning Based Underground Sewer Defect Classification Using a Modified RegNet

Yu Chen[1], Sagar A. S. M. Sharifuzzaman[2], Hangxiang Wang[1], Yanfen Li[1], L. Minh Dang[3], Hyoung-Kyu Song[3] and Hyeonjoon Moon[1,*]

[1]Department of Computer Science and Engineering, Sejong University, Seoul, 05006, Korea
[2]Department of Intelligent Mechatronics Engineering, Sejong University, Seoul, 05006, Korea
[3]Department of Information of Communication Engineering, and Convergence Engineering for Intelligent Drone, Sejong University, Seoul, 05006, Korea
*Corresponding Author: Hyeonjoon Moon. Email: hmoon@sejong.ac.kr
Received: 27 June 2022; Accepted: 10 October 2022

Abstract: The sewer system plays an important role in protecting rainfall and treating urban wastewater. Due to the harsh internal environment and complex structure of the sewer, it is difficult to monitor the sewer system. Researchers are developing different methods, such as the Internet of Things and Artificial Intelligence, to monitor and detect the faults in the sewer system. Deep learning is a promising artificial intelligence technology that can effectively identify and classify different sewer system defects. However, the existing deep learning based solution does not provide high accuracy prediction and the defect class considered for classification is very small, which can affect the robustness of the model in the constraint environment. As a result, this paper proposes a sewer condition monitoring framework based on deep learning, which can effectively detect and evaluate defects in sewer pipelines with high accuracy. We also introduce a large dataset of sewer defects with 20 different defect classes found in the sewer pipeline. This study modified the original RegNet model by modifying the squeeze excitation (SE) block and adding the dropout layer and Leaky Rectified Linear Units (LeakyReLU) activation function in the Block structure of RegNet model. This study explored different deep learning methods such as RegNet, ResNet50, very deep convolutional networks (VGG), and GoogleNet to train on the sewer defect dataset. The experimental results indicate that the proposed system framework based on the modified-RegNet (RegNet+) model achieves the highest accuracy of 99.5 compared with the commonly used deep learning models. The proposed model provides a robust deep learning model that can effectively classify 20 different sewer defects and be utilized in real-world sewer condition monitoring applications.

Keywords: Deep learning; defect classification; underground sewer; computer vision; convolutional neural network; RegNet

1 Introduction

Sewer monitoring is essential to support urban subterranean infrastructure. There are approximately more than 800,000 miles of municipal sewage pipelines and 500,000 miles of residential sewage pipelines in the U. S. alone [1]. While governments have extended sewage lines to accommodate expansion and upgrade water systems, sewer monitoring has received much less attention [2]. The Clean Watersheds Needs Survey points out that there isn't enough money to monitor wastewater infrastructure [3,4]. Moreover, towns throughout the United States are dealing with aging sewage infrastructure that has to be maintained, repaired, or replaced. Additionally, these sewage lines must be examined regularly to minimize pipe breakage or reduced sewer efficiency [5].

Sewer condition monitoring is presently carried out on-site by a qualified investigator who operates a remote-controlled robot with a camera through the sewage line at the same time. The investigators must watch a video stream for an extended period of time, which is difficult and exhausting labor. This may lead to inaccurate assessments, contributing to sewage structural damage in the worst-case scenario. In order to minimize maintenance expenses and enhance the performance of a computerized assessment, robots and scanning equipment have been widely employed in recent years to monitor and manage different structures [6–8]. Consequently, there is fierce competition between many leading industrial robot companies to develop a more advanced inspection robot. One example is ETRI's utility hole condition monitoring robot, which uses a 360-degree field of vision to accurately examine utility holes up to a distance of 50 feet in sewerage systems. Furthermore, the robot's high-quality sensor allows it to accurately record the internal characteristics of the medium and big sewer lines. Furthermore, closed-circuit video taken by robots is a budget and appropriate method of monitoring the condition of the sewer in environmentally fragile or challenging monitoring scenarios.

Researchers recently utilized deep learning techniques to automatically extract sewer defect information from a given image. Although deep learning based solutions provide promising results to classify the sewer defects, it needs a large dataset to achieve state-of-the-art accuracy. The accuracy of existing deep learning based sewer classification models are not as high as other classification problems. Moreover, the overfitting issue of the current available solution affects the model's robustness when applied in real-world sewer condition monitoring applications. The lack of a large dataset with extensive classes of defects also affects the performance of the existing solutions. Therefore, a large dataset with more defect classes and a robust deep learning model must be introduced, which can be implemented in the constraint environment.

This paper utilized a dataset that was generated using the Closed Circuit Television (CCTV) video frames provided by the Korea Institute of Civil Engineering and Building Technology. The datasets for the sewer condition monitoring comprise many defect classes, each of which has multiple photos collected from the footage and thoroughly checked by the inspection officer. Due to a large number of CCTV recordings, an automated sewer condition monitoring system must be developed to automatically identify defects and obtain relevant information regarding the sewer condition. An automated sewer condition monitoring system has several advantages, including (1) reducing investigation mistakes caused by tiredness, perceptions, and varying skill levels of investigators, (2) detecting defects that go undetected by the naked eye, and (3) allowing rapid assessments and process monitoring of CCTV footage to maintain monitoring system robustness [9]. The contribution of our proposed sewer defect classification system are as follows,

1. A large manually collected sewer defect dataset with 20 different sewer defect classes can be used to train deep learning models to provide a robust sewer defect classification solution.

2. The original RegNet deep learning model is modified by introducing the dropout layer to the sequence excitation block to reduce overfitting issues and the activation function of the sequence excitation block also changed to LeakyReLU from ReLU to increase the performance of the model.

3. The extensive experiments show that our proposed model outperforms other state-of-the-art deep learning in all evaluation metrics, which can be implemented in the constraint sewer condition monitoring application to classify sewer defects effectively.

This study proposes a deep learning based sewer condition monitoring framework that supports automated defect monitoring in sewer frames extracted from CCTV inspection videos. We have implemented different deep learning-based image classification techniques such as RegNet, ResNet50, VGG16, VGG19, and GoogleNet to evaluate the performance of detecting defects in the sewer [10–13]. We have found that the proposed RegNet+ model outperforms other deep learning-based methods in detecting sewer defects.

2 Related Works

Due to technological advancements, notably in computer vision there is an expanding number of deep learning-based sewage monitoring methods available in the literature [14]. Myrans et al. provided an automated system for recognizing different kinds of defects in CCTV recordings [15]. First, they calculate each frame's feature representation before analyzing with other methods. The information from each frame was then analyzed using two machine learning methods. The Hidden Markov Model and the filter approach were used to gather data from a sequence of images to enhance the model's performance. Their model has achieved an accuracy of more than 80%. However, the dataset employed in this study was insufficient, with just 1000 photos, and more than half of the sample fell into the no defects class. Ye et al. proposed feature extraction and machine learning methods for detecting sewer defects [16]. They used a support vector machine (SVM) model to classify seven classes of sewer pipe issues. The performance of the model was measured at 84.1% after being applied to 28,760 m of sewage lines. Fang et al. presented a system for defect detection that employed an unsupervised machine learning-based defect detection method on CCTV data [17]. The achieved accuracy was over 90% for the proposed model. Although machine learning approaches have been widely utilized for sewer monitoring, they depend significantly on pre-processing techniques and proper feature extraction in particular scenarios.

Deep learning, a subset of machine learning (ML) that uses multi-layered artificial neural networks, has delivered state-of-the-art results in detection, classification, and other domains. Hassan et al. presented a convolutional neural network-based model for identifying sewage cracks using pictures taken from CCTV recordings [18]. The presented method used in their study is able to achieve 96.33% accuracy in classifying six basic forms of defects. However, the inequitable distribution of data throughout the class impacted the model's performance. Cheng et al. proposed an automated detection of sewage cracks using the Faster R-CNN technique [19]. Several tests were conducted to assess the model's performance, including accuracy and computing costs. The high detection accuracy was attributed to changes in variables such as stride parameters and filter sizes, resulting in an 83% mean average precision (mAP). The presented model was used for still photos; hence it is necessary to study video analysis. Xie et al. used a deep convolutional neural network (CNN) model to automatically extract sewer defect features [20]. Several trials demonstrated that the framework generalized the new data effectively, yielding a classification accuracy of more than 94% on state-of-the-art datasets. However, the system had high processing costs and was unable to identify common

features such as defect deformation and deflections. Meijer et al. presented a collection of sewer crack photos taken from CCTV video that included more than 21,000 photos [21]. The authors then used the acquired dataset to conduct a deep learning method for crack detection. They also introduced a leave-two-inspections-out cross-validation method, substantially avoiding data leakage bias. The proposed classifier didn't meet state-of-the-art crack classification criteria. Moreover, their paper lacks authority by using the recall value as an experimental indicator because the accuracy rate is the most important basis for judging the model's performance in practical use. Oh et al. proposed a novel automated framework for identifying sewage pipe faults using the enhanced you only look once v5 (YOLOv5) architecture for CCTV footage [22]. Their proposed model achieved a mAP of 75.9% for real-time sewer defect detection outperforming other traditional methods. Although the proposed method achieved very good accuracy, the model is not lightweight, affecting the performance of the resource-constrained system used to monitor the underground sewer pipeline, such as CCTV camera devices. Li et al. introduced pipe segmenting objects by locations (SOLO), a new automated instance defect segmentation model that can segment six different classes of sewage pipe defects. Their proposed model outperformed other traditional methods for segmenting sewer defects, achieving a mAP of 59.3%. However, the proposed segmentation can only be implemented on the still images, which is not suitable for analyzing the real-time CCTV video analysis. Dang et al. introduced an effective and sustainable deep learning-based system for automatically detecting and evaluating sewer defects [23]. In addition, an ensemble-based methodology and a cost-sensitive learning-based approach were presented to address the unbalanced data issue. They were able to identify seven different types of sewage defects with an overall accuracy of 97.6%. However, the proposed system cannot detect more than one defect in a single image, which is not suitable for a real-time CCTV sewer defect classification system where one image can contain multiple defect classes. Zhou et al. proposed a CNN based automated sewer defect classification model to classify six common sewer defects [24]. They achieved an average accuracy of 90% during the training and the prediction accuracy was over 95% during prediction of the different sewer defects. However, the feature difference of different objects in an image such as obstacles and walls may deteriorate the proposed model's performance. Moreover, the classification accuracy of the proposed model is comparatively low compared to the other classification model discussed earlier. Ma et al. proposed Style generative adversarial network (GAN) and sharpness discrimination model to generate the sewer defect images to classify multiple sewer defects [25]. They used Fusion CNN model to classify the sewer defects and achieved a mean accuracy of 95.64%. They only considered four types of sewer pipeline defects in their research, Moreover, there is a certain difference between the generated images and the real images, which can affect the performance of the classification model.

3 Materials and Methods

Fig. 1 shows the detailed architecture of the proposed sewer condition monitoring system using deep learning models. The system framework consists of four parts: Image collection, data augmentation, model training, and defect classification output. The sewer defect images are obtained by examining the images which were retrieved from Closed Circuit videos. The images are then classified into twenty defect classes by manual inspection. The training dataset is fine-tuned using data augmentation approaches before training the proposed model. The dataset is then utilized for training several deep learning models to assess their performance, then the best model is chosen to classify sewer pipeline defects.
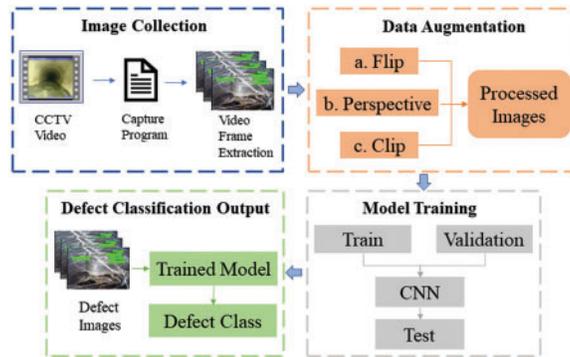
Figure 1: The comprehensive structure of the automated sewer condition monitoring system includes four major parts. (1) image collection from videos; (2) data augmentation on the acquired images; (3) CNN classification is used to train a different class of defects; (4) defect classification; and different models are trained to evaluate the performance of the proposed model with state of the models

3.1 Data Preparation

The robot-based solution was used to collect the data since the inspectors may check the sewage system remotely using a remote camera. This study used 7733 CCTV videos of sewers, which ranged in duration from 30 s to 15 min and had a pixel of 1280×720. The dataset was initially collected by the Korea Institute of Civil Engineering and Building Technology, and we have used their dataset in our study.

The robot was equipped with a 1.3-megapixel Exmor complementary metal-oxide-semiconductor (CMOS) camera and could rotate fully 360 degrees, tilt up or down, and observe 240 degrees of side views. In addition, six 35 W high-power led bulbs were attached to the robot, which enabled the robot to take videos in different scenarios.

Sewer defect datasets were developed by manually analyzing CCTV footage from the original collection with a total of twenty classes. Then different data augmentation methods are used to pre-process data before training the sewer monitoring model. The distribution of images per class extracted from the CCTV footage is given in Table 1.

Table 1: The description of the proposed underground sewer defect classification dataset

Defect name	Code	Before augmentation	After augmentation
Buckling	BC	200	510
Broken pipe	BK	400	510
Crack, circumferential	CC	400	510
Crack, longitudinal	CL	400	510

(Continued)

Table 1: Continued

Defect name	Code	Before augmentation	After augmentation
Crack, multiple	CM	453	510
Deformation	DF	170	510
Debris, greasy	DG	85	510
Deposits, silty	DS	400	510
Hole	HL	391	510
Joint, displaced	JD	400	510
Joint, faulty	JF	400	510
Joint, separated	JS	400	510
Lining defect	LD	93	510
Lateral, protruding	LP	400	510
Lateral, sealing defective	LS	236	510
Permanent obstruction	PO	122	510
Root intrusion	RI	158	510
Surface damage	SD	400	510
Sags	SG	150	510
Temporary obstruction	TO	75	510

3.2 Data Augmentation

Data augmentation refers to approaches for enhancing the quantity of data in a dataset based on various alterations in order to increase the number of instances in the entire dataset. Data augmentation not only contributes to the expansion of the dataset but also enhances its variety. Data augmentation serves as a regularizer and prevents overfitting in training deep learning-based classification models. The data augmentation method for the classification model involves crop, clip, flip, perspective, rescale, and rotation; among them, clip, flip, and perspective methods are used in this study.

Fig. 2 shows the data augmentation performed on our sewer datasets, enhancing the datasets' variety and robustness to overfitting methods. We have selected three augmentation techniques such as clip, flip and perspective, where the flip and perspective are done randomly. The processed images are then utilized for training the sewer monitoring classification model to classify different defects.

3.3 Underground Sewer Defect Classification Model

Neural architecture search (NAS) [26] has recently gained much popularity, despite its high computational cost. Some major drawback of the conventional NAS technique is its inflexibility, generalizability, and interpretability issues. The conventional NAS approach relies on sampling individual networks. So, researchers have presented a global estimate of network design space. Naturally, we can determine the optimal network size if we can establish the functional correlations between different network features, such as network dimensions. This is because we will better understand how these elements operate together. EfficientNet's design concept is used to extend the network dimension using a simple and efficient composite factor. This approach does not randomly modify the network's dimensions as the conventional method does. NAS uses neural structure search technology

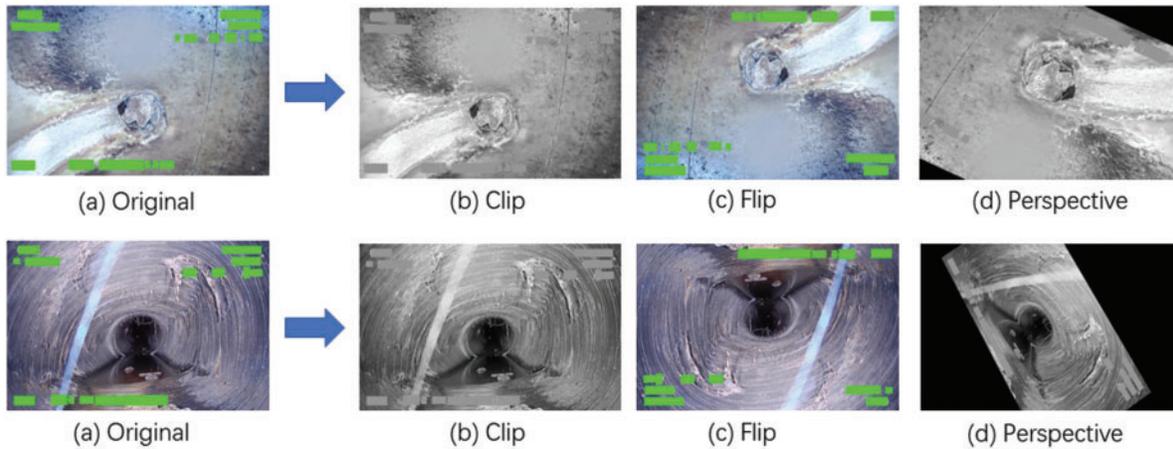to seek out the optimal model within a certain computing expense and find the optimal combination of parameters.

Figure 2: The data augmentation method used in our proposed dataset to enhance the quality of the dataset, where three common augmentation technique such as clip, flip and perspective

RegNet incorporates NAS technology as well, although not in the same way as other prior NAS systems (such as MobileNetV3 and EfficientNet). Conventional NAS employed search techniques to determine the optimal set of parameters within a specified search frame. RegNet, on the other hand, investigates the construction of regions and the particular network design principles instead of just combining parameters. In contrast to EfficientNets, RegNet does not concentrate on a particular network or collection of networks. RegNet outperforms the present available EfficientNet in terms of accuracy while also being five times quicker on the Graphics processing unit.

The RegNet network is divided into three components, as shown in Fig. 3a: stem, body, and head. In this network, the body's architecture takes priority over the stem and head of the network.
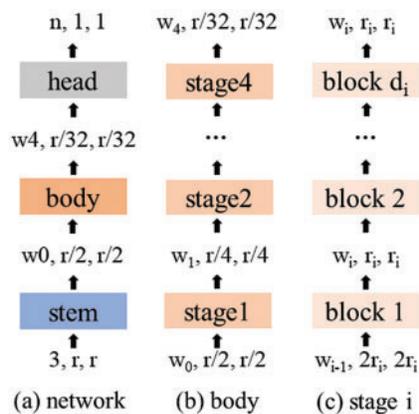

Figure 3: RegNet general network structures for deep learning models which consists of stem, body, and head part

A regular convolution layer serves as the stem (including batch normalization and Relu). Fig. 3b depicts the body's structure, which is categorized into four stages, similar to a stack. There will be a 50% reduction in the input feature matrix's height and width after each step, as depicted in Fig. 3c. A stage is comprised of a number of block stacks. Every stage's initial block consists of two shortcut layers and two main convolution layers. The classification network's head classifier is comprised of a global average pooling layer and a fully connected layer.

Fig. 4 illustrates the architecture of the block, with Fig. 4a illustrating the instance of step stride, which is equal to 1, and Fig. 4b illustrating the instance of step stride, which is equal to 2. We can see in Fig. 4 that the RegNet block is quite similar to the ResNet block. It consists of two 1×1 convolution layer and a 3×3 convolutional layer, including batch normalization and ReLU. The model does not process further when the stride is equal to 1 on the shortcut connection. A 1×1 convolution layer is used for downsampling when the stride is 2. The input and output resolution (r) stay constant if $s = 1$; if $s = 2$, the output resolution is reduced input to 50%.
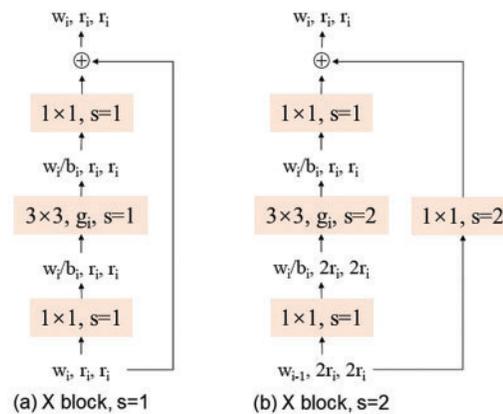

Figure 4: RegNet block structure based on the residual bottleneck block, where w represents the characteristic matrix's channel, g represents each group's group width, and b represents the bottleneck ratio

Figs. 5 and 6 show the proposed RegNet+ structure and block diagram used in this study to classify different sewer pipeline defects. The proposed structure consists of three parts: stem, body, and head. The stem part is consisting of a 3×3 convolution layer, batch normalization, and LeakyReLU activation function. 12 convolution layers and 4 2D convolution blocks are used in the structure's body part. The block section of the original RegNet is modified to reduce overfitting and improve the performance of the model. The block consists of two 1×1 convolutional layer, one 3×3 convolutional layer, a squeeze excitation (SE) block, and a dropout layer. The adaptive average pooling is removed from the SE block and we have changed squeeze channel calculation method by multiplying the input channel with squeeze ratio whereas the original the SE block divide the input channel by the squeeze ratio. The squeeze ratio is set to 0.25 to provide a constant ratio to the model. The shortcut connection is used to solve the gradient divergence problem. The LeakyReLU activation function is used as it performs better than other methods such as the sigmoid function.
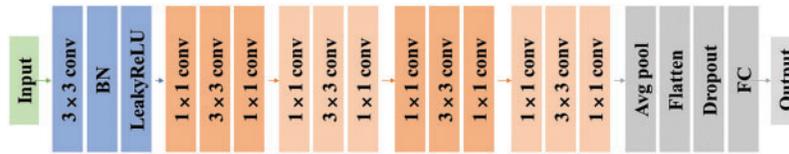
Figure 5: The proposed model is based on the RegNet structure, where FC represents fully connected layer, BN represents batch normalization
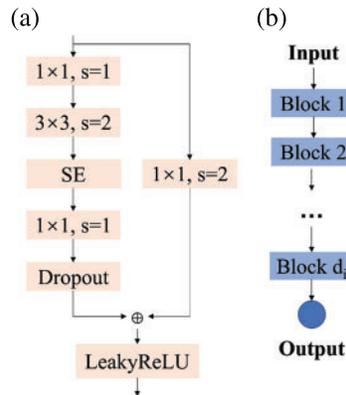

Figure 6: The block diagram of the proposed RegNet+ structure where s represents stride, SE represents sequence excitation block

The overall prediction procedure of the sewer defect classification can be seen in algorithm 1. The system first obtains CCTV videos from the remote robot. Then the system automatically extracts every frame from the sewer defect videos. The extracted images are then fed into the CNN classifier model one by one to predict the defect for each frame. However, the system shows error message if it does not find any defect image dataset.

4 Results and Discussion

An extensive experiment was performed on the sewer defect classification dataset to evaluate the performance of the proposed model. Different available deep learning models were used to train on the proposed dataset, and then they were evaluated using the standard multiclass evaluation metrics such as accuracy precision, F1-score, and recall. The best model is then trained again by optimizing the hyperparameters to achieve highest performance from the model. The hyperparameter optimized model is then used on the different noisy sewer images to predict the sewer defects and evaluate the performance of the model. Lastly, the proposed model is compared with the available sewer defect classification solutions found in the literature.

4.1 Classification Evaluation Metrics Experiments

The data augmentation method was implemented on the proposed dataset to increase the images of the sewer defect. We randomly selected 12000 images from the augmented dataset, where every class was distributed to 600 images. The proposed dataset is then split into train and validation sets using the 85:15 method. The training set contained 10200 images, and the validation set contained 1800 images. The deep learning models were trained in the Pytorch library, which is one of the most popular libraries for training deep learning models. We have implemented five mostly used deep learning models to train on our datasets and evaluate these models' performance. A stochastic gradient descent optimizer was utilized for training the model. The initial learning rate for all deep learning models used in this experiment was set to 0.001. The batch size and epochs for all the models were set to 64 and 200, respectively. The image sizes for both training and validation were set to 224×224. Fig. 7 shows the training and validation accuracy of all the models used in this experiment. It can be seen from the Fig. 7 that most of the model's accuracy becomes stable around 40 epochs. However, the proposed RegNet+ model's training accuracy became stable around 25 epochs and achieved the highest accuracy among other deep learning models for both training and testing. Table 2 illustrates the accuracy in terms of training and testing for all deep learning models. It can be seen that the accuracy of our proposed model for both training and test was 99.1 and 98.79, respectively.
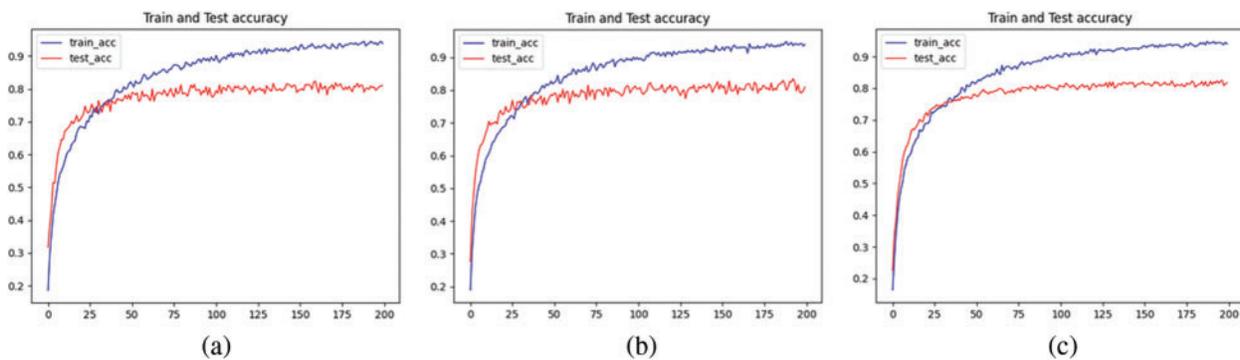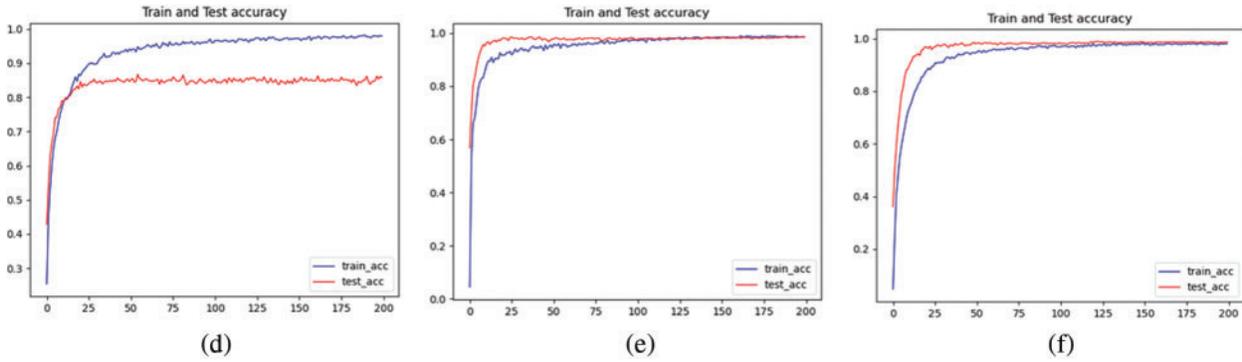

Figure 7: (Continued)

Figure 7: Training and testing results of the implemented deep learning models, (a) VGG16, (b) VGG19, (c) GoogleNet, (d) ResNet50, (e) Original RegNet, (f) RegNet+

Table 2: Training and testing accuracy of the implemented deep learning models

Model name	Training accuracy	Testing accuracy
VGG16	94.5	82.3
VGG19	94.8	83.5
GoogleNet	94.8	82.5
ResNet50	98.1	86.6
RegNet	98.9	98.68
RegNet+	**99.1**	**98.79**

The confusion matrix is also calculated to evaluate the performance of our proposed model. 1800 images with 20 classes were utilized to calculate the confusion matrix of the proposed model. The detailed confusion matrix for two best deep learning models are given in the Fig. 8. It can be seen that the classwise performance of the proposed model is very high.

We utilized standard multiclass classification evaluation metrics. In each classification test, we computed true positive (TP), true negative (TN), false negative (FN), and false-positive (FP). Using the given formula, we obtained the average classification accuracy (A), average recall (R), average precision (P), and F-1 score (F-1).

$$A = \frac{TP + TN}{TP + TN + FP + FN}$$
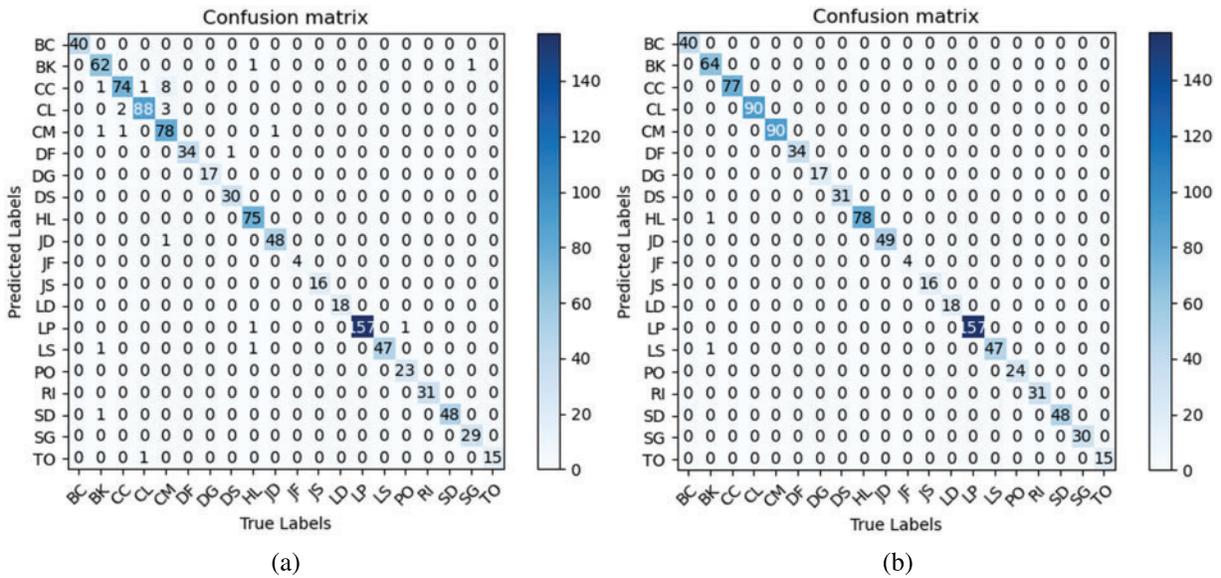
$$R = \frac{TP}{TP + FN}$$

Figure 8: The confusion matrix of the implemented RegNet and RegNet+ deep learning model. (a) represents the RegNet model, (b) represents the RegNet+ model

$$P = \frac{TP}{TP + FP}$$

$$F = 2 * \frac{R * P}{R + P}$$

Table 3 shows the metric evaluation results used in this study, where average precision, recall, F1-score, and average were calculated for the five deep learning models. It can be seen that GoogleNet performs worse than other methods which were 69.97, 51.67, 59.44, 57.38 in terms of precision, recall, f1-score and accuracy, respectively. On the other hand, our proposed model outperforms other models for all the performance metrics. The proposed model's average precision, recall, F1-score, and accuracy were 98.83, 98.85, 98.83, and 98.79, respectively.

Table 3: The average performance metrics of the multi class classification used in this study

Model name	Precision	Recall	F1-score	Accuracy
VGG16	65.94	61.09	51.30	64.40
VGG19	67.57	69.06	68.30	69.85
GoogleNet	69.97	51.67	59.44	57.38
ResNet50	86.99	84.90	85.93	86.69
RegNet	98.82	98.83	98.82	98.68
RegNet+	**98.83**	**98.83**	**98.83**	**98.83**

4.2 Hyperparameters Experiment

The hyperparameters play an important role in achieving a particular model's highest accuracy. The hyperparameters for the classification model include the learning rate and optimization algorithms. When a model's learning rate is not optimized, the loss varies and convergence speed also becomes slow [27,28]. The two most common optimizer methods, Adam and SGD, are also used to determine the optimized hyperparameters for our proposed model.

Table 4 shows the effect of different hyperparameters on the performance of the model. We have trained different deep learning models along with our proposed model using Adam and SGD optimizers. We selected three learning such as 0.01, 0.005 and 0.001 for every optimizer. Then we calculate the testing accuracy for all deep learning models to compare the results to choose best model for further evaluation. It can be seen that the proposed RegNet+ model achieved the highest accuracy using the SGD optimizer, with a 0.005 learning rate and 0.9 momentum. Therefore, we have selected the learning rate of 0.005 and SGD optimizer for our proposed model.

Table 4: The effect of model performance using different hyperparameters for defect classification system. Two optimizer such as adaptive moment estimation (Adam) and stochastic gradient descent (SGD) with three learning rate such as 0.01, 0.005, 0.001 were used to train the model on our proposed dataset

Optimizer	Learning rate	VGG16	VGG19	GoogleNet	ResNet50	RegNet	RegNet+
Adam	0.01	16.3	16.3	71.5	76.7	98.5	80.8
	0.005	16.3	16.3	77.9	78.3	98.4	85.1
	0.001	60.4	60.1	83.5	83.4	96.6	80.2
SGD	0.01	82.6	83.5	83.8	83.5	98.6	99.2
	0.005	82.7	83.3	83.1	86.1	98.5	**99.5**
	0.001	82.3	83.5	82.5	86.6	98.6	98.7

4.3 Model Evaluation for Constraint Environment

The acquired images from different sewers with different equipment can contains images with noise. Therefore, the robustness of the defect classification toward different noises should be evaluated prior to implementation in the real time sewer condition monitoring application. In this paper, we considered four common noises that can be found in the sewer images to evaluate the performance of our proposed model. These four common noises are crop, rotate, partial overlapping and block noise which can occur due to the constraint environment and monitoring equipment as well.

Fig. 9 shows the prediction score of the proposed model on the random cropped sewer images. The original image can be seen in Fig. 9a, which was then randomly cropped and fed into classifier to predict the defect. It can be seen that the proposed model accurately predicts the defect class with high prediction score. The average prediction score for crop image was over 90% with one exception which can be seen in Fig. 9d. This is due to the lack of the defect surface distribution to the different images.
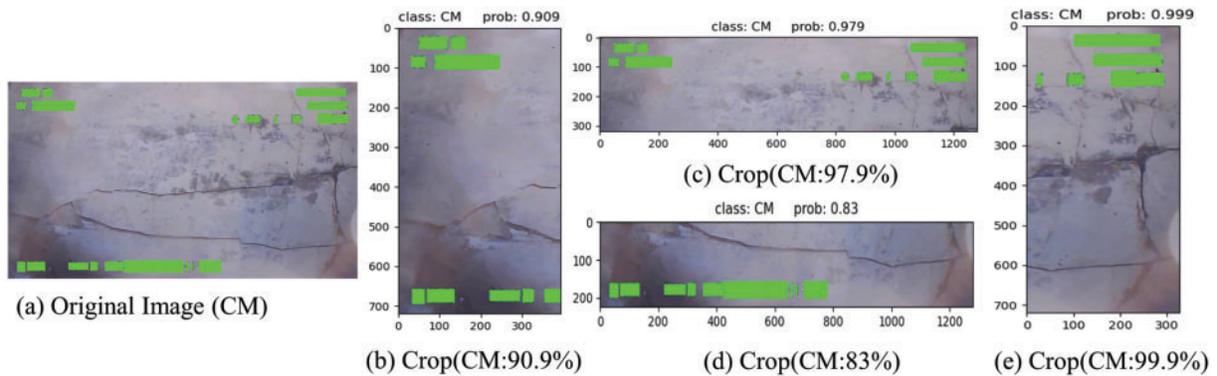
Figure 9: The prediction results of our proposed model on the cropped images. (a) represents the original image from the crack multiple class. (b–e) are the predicted results using our proposed model on randomly cropped images

Fig. 10 shows the prediction score of the surface damage sewer defect with random rotation. It can be seen that the predicted accuracy of our proposed classification model was more than 95 for rotated images. The proposed model achieved high prediction score for rotated images because the distribution of defect did not change when the image was rotated. Nonetheless, our proposed model could classify the defects for constrained underground sewer defect images, demonstrating that the proposed model can be implemented in the different constraint environments to detect and classify the underground sewer defect.
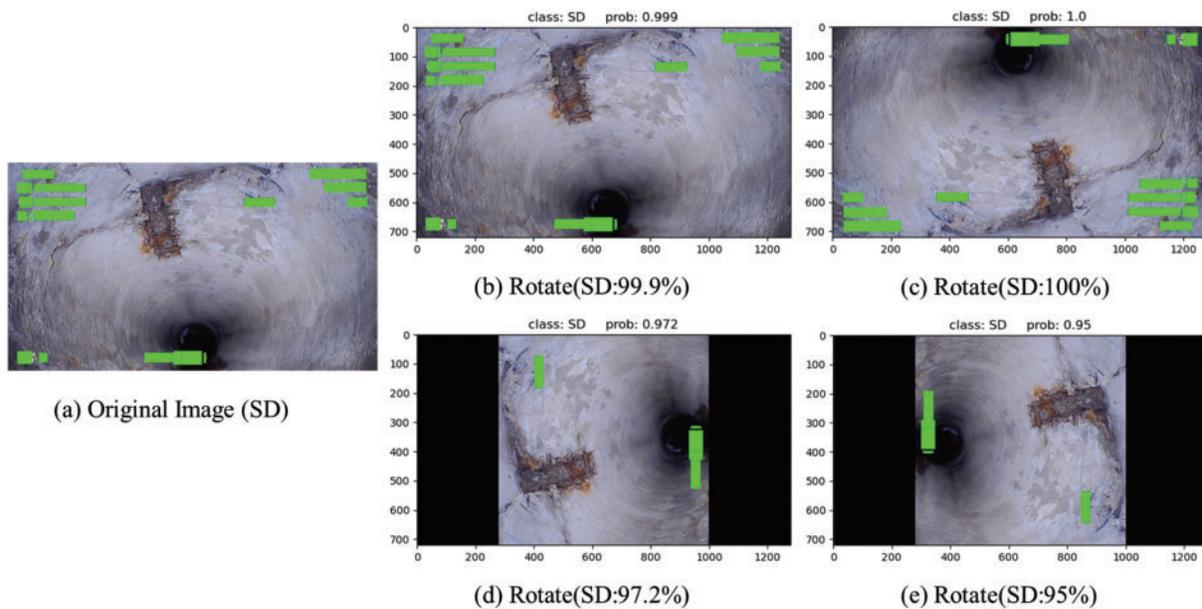

Figure 10: The prediction results of our proposed model on the rotated images. (a) represents the original image from the surface damage class. (b–e) are the predicted results using our proposed model on the rotated images

We have also evaluated our proposed sewer defect classification model against various noises which can be found in the real-world surveillance camera footage. We have utilized images from the lateral protruding class and put the images on other images to make the image partially overlap to check the prediction's performance. Fig. 11 shows the proposed model's predicted results on the partially overlapping images. Random overlapping was done for the same image, and the proposed model predicted all the images correctly. The lowest accuracy was achieved in Fig. 11b, which was 83.4%, whereas the highest accuracy was achieved in Fig. 11d, which was 97.5%.
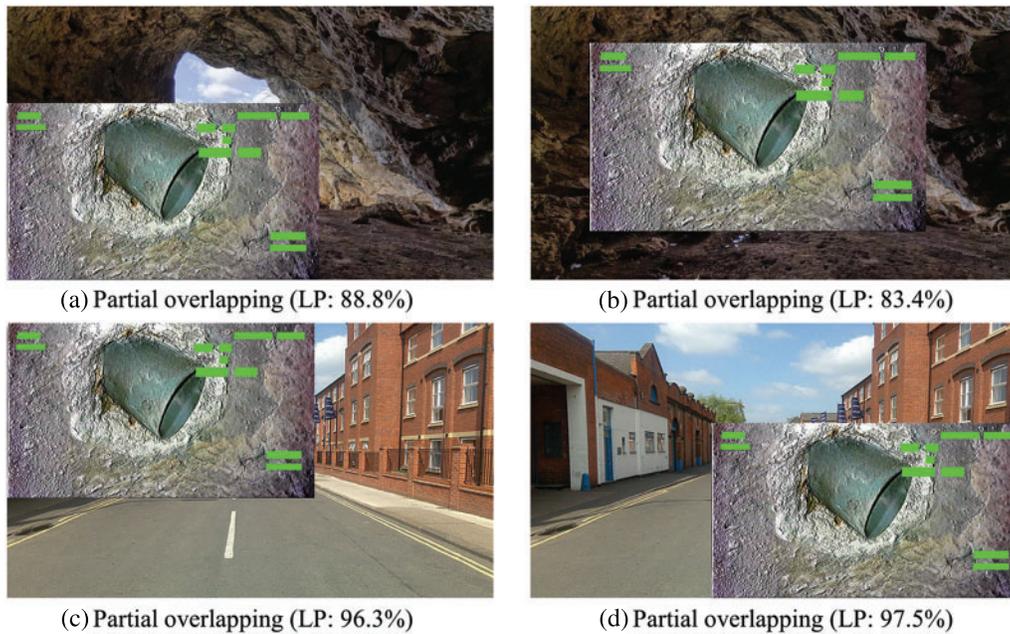

(a) Partial overlapping (LP: 88.8%)   (b) Partial overlapping (LP: 83.4%)

(c) Partial overlapping (LP: 96.3%)   (d) Partial overlapping (LP: 97.5%)

Figure 11: The prediction results of our proposed model on the partially overlapping images

The noisy images were also fed into the proposed model to evaluate the prediction performance of the proposed model. Fig. 12 shows the predicted result on different block noise images with good accuracy. Different block noises were added to Figs. 12b–12d prior to being fed into the classifier. The proposed model predicted all the images correctly for lateral protruding class, where the lowest accuracy among them was achieved for Fig. 12c, which was 91.3%. Figs. 12e and 12f represent the partial defect block images, where some or major sections of the defect were blocked with a black screen and sent to the proposed model for prediction. We have observed that the proposed model can accurately classify the defect images, but the accuracy is not as high as other images. We observed the lowest accuracy of 46.1% for Fig. 12e, where a major part of the defect was blocked. After analyzing the result, it is evident that the proposed model is robust to the different noises and constrained environments that can be found in the real world.
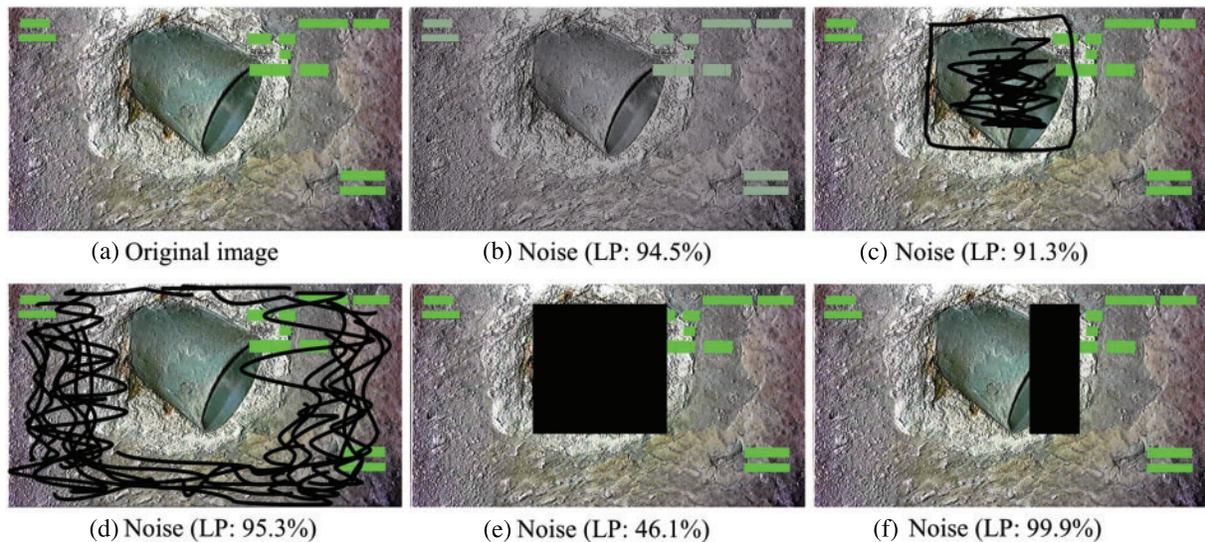
(a) Original image   (b) Noise (LP: 94.5%)   (c) Noise (LP: 91.3%)

(d) Noise (LP: 95.3%)   (e) Noise (LP: 46.1%)   (f) Noise (LP: 99.9%)

Figure 12: The prediction results of our proposed model on the noisy images

4.4 Comparison with Other Works

This section's primary objective is to show that our modified RegNet+ model outperforms the previous model for sewer defect classification. This section provides a comprehensive analysis and comparison of the recent sewer condition monitoring approaches. Dang et al. used a fine-tuned VGG19 deep learning model to classify 12 different defects [23]. They achieved the highest accuracy of 97.6% for sewer defect classification. Ma et al. proposed a StyleGAN-SDM-based method to pre-process the small dataset and then introduced a multi-defect classification model based on the multi-defect classification model (MDCM) to classify sewer defects [25]. They achieved an accuracy of 95.6% using 14451 images with 4 different defect classes. Li et al. introduced modified ResNet18 on the imbalanced sewer classification dataset [29]. They used images of 7 classes to train and validate the model and achieved an accuracy of 64.8%. Kumar et al. proposed a CNN-based deep learning model on a dataset having 8 common defect classes [30]. They achieved the highest classification accuracy of 86.2% for sewage defects. Situ et al. also utilized StyleGAN to generate defect images to train the synthetic images with different classifier model [31]. They used a variant of StyleGAN cascaded with adaptive discriminator augmentation (ADA) to prepare the datasets. They found out that Inception_v3 deep learning model performs well in classifying different sewer defect, where the model achieved an accuracy of 94% for four defect classes. On the other hand, the proposed method used a small dataset with 20 different defect classes to classify sewer defects in a sewage system. Despite having a small dataset, the proposed model outperforms state-of-the-art sewer classification models in terms of accuracy. The proposed model achieved an accuracy of 99.5%, which demonstrates that the proposed model has the highest performance in sewer defect classification compared to previous methods. The comprehensive comparisons with different method available in the literature are shown in Table 5.

Table 5: Comparisons between the proposed method and the recent sewer defect classification found in the literature

Classification approach	Defect class number	Accuracy
Fine-tuned VGG19 [23]	12	97.6
StyleGAN-SDM+MDCM [25]	4	95.6
Modified ResNet18 [29]	7	64.8
CNN [30]	8	86.2
StyleGAN-ADA+CNN [31]	4	94.0
Our method	20	**99.5**

5 Conclusion

This study proposed a defect classification framework using a CNN-based deep learning model on the collected CCTV underground sewer dataset. A large dataset of sewer defects dataset contains 20 different classes is proposed to classify various defects found in the sewer. The dataset was prepared by manually extracting the CCTV video frame from a remote robot. The data quality is also enhanced by applying data augmentation, which further improves the classification model's performance. Finally, we have modified the original RegNet by adding the dropout layers in the RegNet structure block to reduce the model's overfitting. Moreover, the LeakyReLU activation function was used instead of ReLU to achieve the highest accuracy by optimizing the model network.

An extensive experiment was performed to evaluate the performance of our proposed model. The experimental results demonstrated that the proposed model outperforms other models in terms of precision, recall, F1-score, and accuracy, which are 98.83, 98.85, 98.83, 98.79, respectively. The hyperparameter optimization is performed to further improve the proposed model's accuracy. We have achieved a testing accuracy of 99.5% by optimizing the hyperparameters such as learning rate-0.005 with SGD optimizer. Moreover, we have added different noises to the testing images which can occur in the real-world scenarios to evaluate the prediction score of the proposed model. The result shows that the proposed model can effectively predict the correct class with an overall prediction score of 90%, whereas the lowest prediction score of the correct class is 46.1%. Therefore, the proposed model can be used for underground sewer condition monitoring and defect classification tasks.

Although our proposed model performs well in the single defect class-based images, the proposed model performance deteriorates when predicting multiclass images. Different methods such as meta heuristic learner, object detection method can be implemented to detect multi defects in a single image. The dataset introduced in this paper also has limitations, such as the distribution of the images for all classes are not the same. More images can be extracted to balance dataset for every defect classes. In the future, more challenging experiments can be done to evaluate the performance of the proposed model. The proposed dataset should be extended further to acquire more images from the CCTV videos to accommodate more defect classes. The proposed system can be extended to the real-time classification of CCTV images, which can help investigators detect multiple defects in less time.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] American Society of Civil Engineers (ASCE) Report card for America's infrastructure Retrieved from: https://www.infrastructurereportcard.org/wpcontent/uploads/2017/01/astewater-Final.pdf, (accessed on 20 June 2022), 2017.

[2] N. Caradot, H. Sonnenberg, I. Kropp, A. Ringe, S. Denhez *et al.,* "The relevance of sewer deterioration modelling to support asset management strategies," *Urban Water Journal*, vol. 14, no. 10, pp. 1007–1015, 2017.

[3] United States Environmental Protection Agency (EPA) Clean water sheds needs survey (CWNS) Retrieved from: https://www.epa.gov/cwns, (accessed on 20 June 2022), 2012.

[4] Y. Li, H. Wang, L. Dang, H. Song and H. Moon, "Vision-based defect inspection and condition assessment for sewer pipes: S comprehensive survey," *Sensors*, vol. 22, no. 7, pp. 2722, 2022.

[5] Y. Li, H. Wang, L. Dang, M. Piran and H. Moon, "A robust instance segmentation framework for underground sewer defect detection," *Measurement*, vol. 190, pp. 110727, 2022.

[6] J. Haurum, J. Bruslund and T. Moeslund, "A survey on image-based automation of CCTV and SSET sewer inspections," *Automation in Construction*, vol. 111, pp. 103061, 2020.

[7] L. Dang, H. Wang, Y. Li, T. Nguyen and H. Moon, "DefectTR: End-to-end defect detection for sewage networks using a transformer," *Construction and Building Materials*, vol. 325, pp. 126584, 2022.

[8] H. Wang, Y. Li, L. Dang, S. Lee and H. Moon, "Pixel-level tunnel crack segmentation using a weakly supervised annotation approach," *Computers in Industry*, vol. 133, pp. 103545, 2021.

[9] M. Halfawy and J. Hengmeechai, "Integrated vision-based system for automated defect detection in sewer closed circuit television inspection videos," *Journal of Computing in Civil Engineering*, vol. 29, no. 1, pp. 04014024, 2015.

[10] I. Radosavovic, R. Kosaraju, R. Girshick, K. He and P. Dollár, "Designing network design spaces," in *Proc. 2020 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, pp. 10428–10436, 2020.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 770–778, 2016.

[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[13] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed *et al.,* "Going deeper with convolutions," in *Proc., 2015 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, pp. 1–9, 2015.

[14] W. Guo, L. Soibelman and J. Garrett Jr, "Visual pattern recognition supporting defect reporting and condition assessment of wastewater collection systems," *Journal of Computing in Civil Engineering*, vol. 23, no. 3, pp. 160–169, 2009.

[15] J. Myrans, R. Everson and Z. Kapelan, "Automated detection of faults in sewers using CCTV image sequences," *Automation in Construction*, vol. 95, pp. 67–71, 2018.

[16] X. Ye, J. Zuo, R. Li, Y. Wang, L. Gan *et al.,* "Diagnosis of sewer pipe defects on image recognition of multi-features and support vector machine in a southern Chinese city," *Frontiers of Environmental Science & Engineering*, vol. 13, no. 2, pp. 1–13, 2019.

[17] X. Fang, W. Guo, Q. Li, J. Zhu, Z. Chen *et al.,* "Sewer pipeline fault identification using anomaly detection algorithms on video sequences," *IEEE Access*, vol. 8, pp. 39574–39586, 2020.

[18] S. Hassan, L. Dang, I. Mehmood, S. Im, C. Choi *et al.,* "Underground sewer pipe condition assessment based on convolutional neural networks," *Automation in Construction*, vol. 106, pp. 102849, 2019.

[19] J. Cheng and M. Wang, "Automated detection of sewer pipe defects in closed-circuit television images using deep learning techniques," *Automation in Construction*, vol. 95, pp. 155–171, 2018.

[20] Q. Xie, D. Li, J. Xu, Z. Yu and J. Wang, "Automatic detection and classification of sewer defects via hierarchical deep learning," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 4, pp. 1836–1847, 2019.

[21] D. Meijer, L. Scholten, F. Clemens and A. Knobbe, "A defect classification methodology for sewer image sets with convolutional neural networks," *Automation in Construction*, vol. 105, pp. 281–298, 2019.

[22] C. Oh, L. Dang, D. Han and H. Moon, "Robust sewer defect detection with text analysis based on deep learning," *IEEE Access*, vol. 10, pp. 46224–46237, 2022.

[23] L. Dang, S. Kyeong, Y. Li, H. Wang, T. Nguyen *et al.,* "Deep learning-based sewer defect classification for highly imbalanced dataset," *Computers & Industrial Engineering*, vol. 161, pp. 107630, 2021.

[24] Q. Zhou, Z. Situ, S. Teng and G. Chen, "Convolutional neural networks–based model for automated sewer defects detection and classification," *Journal of Water Resources Planning and Management*, vol. 147, no. 7, pp. 04021036, 2021.

[25] D. Ma, J. Liu, H. Fang, N. Wang, C. Zhang *et al.,* "A multi-defect detection system for sewer pipelines based on StyleGAN-SDM and fusion CNN," *Construction and Building Materials*, vol. 312, pp. 125385, 2021.

[26] C. Han, L. Zhu and S. Han, "Proxylessnas: Direct neural architecture search on target task and hardware," arXiv preprint arXiv:1812.00332, 2018.

[27] J. Yang and G. Yang, "Modified convolutional neural network based on dropout and the stochastic gradient descent optimizer," *Algorithms*, vol. 11, no. 3, pp. 28, 2018.

[28] H. Wang, Y. Li, L. Dang, J. Ko, D. Han *et al.,* "Smartphone-based bulky waste classification using convolutional neural networks," *Multimedia Tools and Applications*, vol. 79, no. 39, pp. 29411–29431, 2020.

[29] D. Li, A. Cong and S. Guo, "Sewer damage detection from imbalanced CCTV inspection data using deep convolutional neural networks with hierarchical classification," *Automation in Construction*, vol. 101, pp. 199–208, 2019.

[30] S. Kumar, D. Abraham, M. Jahanshahi, T. Iseley and J. Starr, "Automated defect classification in sewer closed circuit television inspections using deep convolutional neural networks," *Automation in Construction*, vol. 91, pp. 273–283, 2018.

[31] Z. Situ, S. Teng, H. Liu, J. Luo and Q. Zhou, "Automated sewer defects detection using style-based generative adversarial networks and fine-tuned well-known CNN classifier," *IEEE Access*, vol. 9, pp. 59498–59507, 2021.