CSCE Annual Conference
Growing with youth – Croître avec les jeunes

Laval (Greater Montreal)
June 12 - 15, 2019

A NEURAL NETWORK-BASED APPLICATION FOR AUTOMATED DEFECT DETECTION FOR SEWER PIPES

Xianfei Yin[1,8], Yuan Chen[2], Qin Zhang[3], Ahmed Bouferguene[4], Hamid Zaman[5], Mohamed Al-Hussein[4], Randy Russell[6], Luke Kurach[7]
[1] PhD Student, Department of Civil and Environmental Engineering, University of Alberta, Canada
[2] PhD Candidate, Department of Civil and Environmental Engineering, University of Alberta, Canada
[3] Undergraduate Student, Department of Computer Science, University of Alberta, Canada
[4] Professor, Department of Civil and Environmental Engineering, University of Alberta, Canada
[5] General Supervisor, Waste Management, City of Edmonton, Canada
[6] Engineering Technologist, EPCOR Drainage Services, Edmonton, Canada
[7] Senior Infrastructure Engineer, EPCOR Drainage Services, Edmonton, Canada
[8] xianfei@ualberta.ca

Abstract: Manual defect identification and classification for sewer pipes using footage from closed-circuit television (CCTV) monitoring is generally time-consuming and can have varying degrees of accuracy depending on the expertise of the technologist conducting the analysis. In order to address this issue, automation is proposed as an alternative to human visual inspection and consists of extracting still frames from collected videos, examining whether these frames include defects, and finally classifying these defects into different types (e.g., cracks or fractures). A classifier based on a new convolutional neural network, called You only look once (YOLO), is proposed in this paper which consists of four parts: (1) extracting the colour frames including defects from the video; (2) transferring information in the selected frames in order to highlight the part of the image containing the defect; (3) using the training images with the corresponding information as inputs to generate a classifier by means of the YOLO network; and (4) testing performance of the automatic classifier based on the images for validation. The proposed framework is then applied to a case study of the City of Edmonton in order to automatically detect the number and location of defects in sewer pipes to facilitate improved productivity for human visual defect identification and human resource allocation. The results show that the YOLO-based classifier performs accurately: up to 96% accuracy in automated defect detection, although some mistakes occurred such as mix-up other type of defect.

1 INTRODUCTION

The deterioration of urban infrastructure is commonly regarded as a critical problem that all countries worldwide are currently facing, a challenge which is compounded by the reality of limited budgets and finite resources. Sewer pipes represent one of the most important components of the municipal infrastructure, essential to public health and quality of life in urban environments. To curb the potentially dire consequences of an aging infrastructure, substantial budgets and resources are needed to carry out preventive maintenance activities for sewer pipes. For instance, it is estimated that $47 billion will be spent on restoring the condition of sewers in Canada to an acceptable level (Sinha and Knight 2004). In North America, current procedures often use closed circuit television (CCTV) monitoring to inspect the structural integrity of sewer pipes due to the wide commercial availability and practical advantages of

CCTV monitoring (Iseley et al. 1997, Madryas and Przybyla 1998, Makar 1999). After recording the video in the field, the footage is examined by technologists (i.e., trained individuals) in the analysis facility. The defects in a video are detected, classified into different types and labelled according to a standardized nomenclature. For instance, four primary categories of pipe defects are listed based on pipeline assessment and certification program (PACP) (NASSCO 2015): structural defects, operation and maintenance defects, construction defects, and miscellaneous defects. In addition, each type of defect can have several subclasses. The traditional approach to conducting the video-based assessment is to employ human visual identification, but this process is labour-intensive, time-consuming, and is likely to be error-prone since the accuracy of the analysis is impacted by the experience of the technologist. Furthermore, in many instances, such as in recently developed neighborhoods, large portions of the video footage will not show any defects and thus the time spent on watching these portions is essentially wasted. To improve the productivity of analysing video footage, some innovative approaches and other artificial intelligence techniques have been developed for CCTV data. Moselhi and Shehab-Eldeen (1999) employed a neural network-based approach to realize image processing, segmentation, and feature extraction. Moselhi and Shehab-Eldeen (2000) used a three-layer neural network with a back-propagation algorithm to classify four types of defects in sewer pipes (i.e., cracks, joints, spalling, and cross-sectional areas). Sinha and Fieguth (2006) integrated neural networks with fuzzy logic to build a classifier to classify defects in segmented buried pipe images and compared the performance of the proposed approach with traditional classification methods such as the k-nearest neighbours (K-NN) and conventional back-propagation network. Yang and Su (2008) applied image processing techniques to describe pipe textures and used machine learning methods including back-propagation neural network, radial basis network, and support vector machine to classify defects in pipes and compare the performance of each method. Yang and Su (2009) developed a diagnostic system based on radial basis network for automated detection of pipe defects on CCTV images; the critical step of which is to define ideal morphologies of pipe defects. Based on the current literature review, four relevant points emerged and are summarized as follows: (1) different types of defects are detected and classified in the existing research, some of which focus on only one special type while others are able to identify multiple types at the same time; (2) the studied pipes are diverse in their production materials, such as concrete, clay, and PVC (i.e., poly vinyl chloride); (3) there exists diversity in the methods that are used for automated defect identification, while neural network is the more widely-used approach; and (4) image pre-processing is the essential first step before defect detection. Most of the current methods depend on black and white videos or the processed images created using grey scale conversion. With respect to the aforementioned findings in the literature review, this paper introduces a neural network-based method to detect multiple types of defects automatically, which is performed using the colour images of defects in sewer pipes captured using CCTV monitoring. The proposed method in the present study could prove to be beneficial in replacing the traditional method for defect identification, improving detection productivity, and helping manage labour resources. The following sections present a more detailed explanation of the methodology and its application to a real case study.

2 METHODLOGY

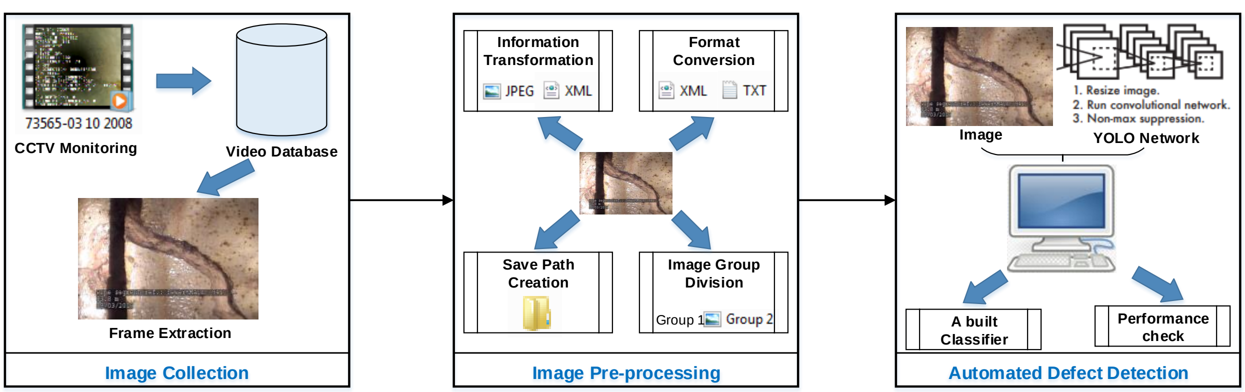

Figure 1: The framework of methodology

The proposed method for automatic defect detection is composed of three parts: image collection, image pre-processing, and automated defect detection. An overview of the framework is provided in Figure 1, where Python and a supercomputer are used as the primary tools to achieve the research objective.

2. 1 Image Collection

The original files used for analysis are videos recorded by CCTV monitoring. Therefore, the first step is to collect all video frames that contain defects in the observed sewer pipes. In general, 30 frames per second can be extracted from the video, by means of the Pinnacle system, which are saved as images in jpeg format. Considering several consecutive frames where defects may be identical, an assumption is made for selecting these frames to make sure that the selected frames for training and validation are individually distinct. For instance, if the time threshold is set to 1 second, several frames with the same defect(s) during that time period are regarded as one frame. The task of image collection is implemented by programs encoded in Python.

2. 2 Image Pre-processing

Defects in the selected images can be visually identified by technologists, but the information should be transferred into a file format that the computer is able to read. Therefore, image pre-processing is an imperative in this step. We label each image and save its corresponding file in xml format, which describes the location of a specific type of defect in this image. Meanwhile, the format conversion and save path creation for the needed files are completed. To test the automatic classifier in the next step, all the selected images are divided into two groups for training and validation, respectively. The above tasks are implemented by programs encoded in Python.

2. 3 Automated Defect Detection

Similar to previous studies, the main method selected for this study is neural networks. According to the function principle, different neural network approaches are generated for pattern recognition and classification, such as the back-propagation neural network (Buscema 1998, Nolan 2002) and radial basis network (Yang and Su 2008, 2009). This study selects a comparably new neural network approach called You only look once (YOLO) to build a defect classifier for sewer pipes in order to achieve automated detection. This approach currently has been applied to visual object tracking (Ning et al. 2017), soil hydraulic function predicting (Minasny et al. 2004), autonomous driving (Wu et al. 2017), and text localization in natural images (Gupta et al. 2016). The principle underlying YOLO is that it uses a single convolutional neural network to predict multiple bounding boxes and their class probabilities at the same time; meanwhile, full images act as the train objects in YOLO and this approach can realize optimization of detection performance (Redmon et al. 2016). Since YOLO does not depend on image pixels, it then succeeds in transferring object detection into a single regression problem. As such, YOLO is superior compared to traditional methods in the area of object detection, which is evidenced by the fact it requires much less computation time, it encodes contextual information and appearance of each class, and it learns generalizable representations of objects (Redmon et al. 2016, Redmon and Farhadi 2017, 2018). A brief introduction to the theory behind YOLO is presented below.

First, the input (i.e., full images) is divided into an $S \times S$ grid and each grid cell can predict whether a bounding box includes an object with a binary variable defined as $\Pr(Object)$. If there exists an object (i.e., $\Pr(Object) = 1$), then this system judges whether the predicted box is accurate as the ground truth with the variable (i.e., intersection over union (IOU)) defined as IOU_{pred}^{truth}. Based on this, the confidence score for each bounding box is equal to $\Pr(Object) \times IOU_{pred}^{truth}$. In addition to the confidence score, each bounding box also has other four predictions: the box center coordinate (x,y), width (w) and height (h). In addition, another variable, conditional class probability, is also predicted in each grid cell, which is defined as $\Pr(Class_i \mid Object)$. Therefore, for each bounding box, its class-specific confidence score is equal to $\Pr(Class_i \mid Object) \times \Pr(Object) \times IOU_{pred}^{truth}$, which can be expressed as $\Pr(Class_i) \times IOU_{pred}^{truth}$.

Before executing YOLO in practice, some input information is needed, such as number and name of classes to detect, parameters in the configuration section, the convolutional section, and the YOLO section. The above information can be modified in order to seek the optimal classification performance. Considering that a general computer would consume a lot of time and require large storage space to achieve automated defect detection, the whole task is accomplished with the support of a remote supercomputer.

3 CASE STUDY

3. 1 Data Source and Preparation

The data used in this study was collected from 63 CCTV monitoring videos for clay sewer pipes, which were recorded by EPCOR Drainage Services in the City of Edmonton, Canada. The results of the pipe condition assessment can be retrieved from two databases of media inspection and condition; the database of media inspection provides the information for all videos recorded by CCTV monitoring in the field, while the conditions database describes the type of defects and their locations in the pipe examined by each technologist. Based on the aforementioned procedure, 1,451 images are generated to display the location and shape of each defect in the pipes. The image selection is based on the consistency principle within 1 second.

The next step is image pre-processing. There are seven classes of defects included in this case: (a) structural defects: cracks, fractures, broken, and hole; (b) operation and maintenance defects: deposit and root; (c) construction defects: tap. Based on this, we label each image, save its defect description file in xml format (see example in Figure 2), and complete the format conversion and save the path creation for all needed files.

941-20180390928_725_HSV.xml


```xml
<annotation>
    <folder>941-20180390928</folder>
    <filename>725_HSV.jpg</filename>
    <path>C:/Users/zqqld/Downloads/941-20180390928/725_HSV.jpg</path>
    <source>
        <database>Unknown</database>
    </source>
    <size>
        <width>720</width>
        <height>480</height>
        <depth>3</depth>
    </size>
    <segmented>0</segmented>
    <object>
        <name>HSV</name>
        <pose>Unspecified</pose>
        <truncated>0</truncated>
        <difficult>0</difficult>
        <bndbox>
            <xmin>309</xmin>
            <ymin>3</ymin>
            <xmax>502</xmax>
            <ymax>51</ymax>
        </bndbox>
    </object>
</annotation>
```


Figure 2: An example for defect description in xml format

3. 2 The Automatic Classifier

We divide the selected 1,451 images into the training group (1,331 images) and the testing group (120 images). In other words, the number of images for training accounts for 91.8% of the total number. Before executing YOLO, some input information must be specified in advance. For instance, the number of classes in this case is seven and their names are listed in section 3.1. Parameters in the convolutional section and the YOLO section are shown in Figure 3, where the number of filters is equal to 3*(number of classes+5).

```
[convolutional]
size=1
stride=1
pad=1
filters=36
activation=linear

[yolo]
mask = 6,7,8
anchors = 10,13,  16,30,  33,23,  30,61,  62,45,  59,119,  116,90,  156,198,  373,326
classes=7
num=9
jitter=.3
ignore_thresh = .5
truth_thresh = 1
random=1

[route]
layers = -4
```

Figure 3: Part of parameter setting in YOLO network

3. 3 Performance Check

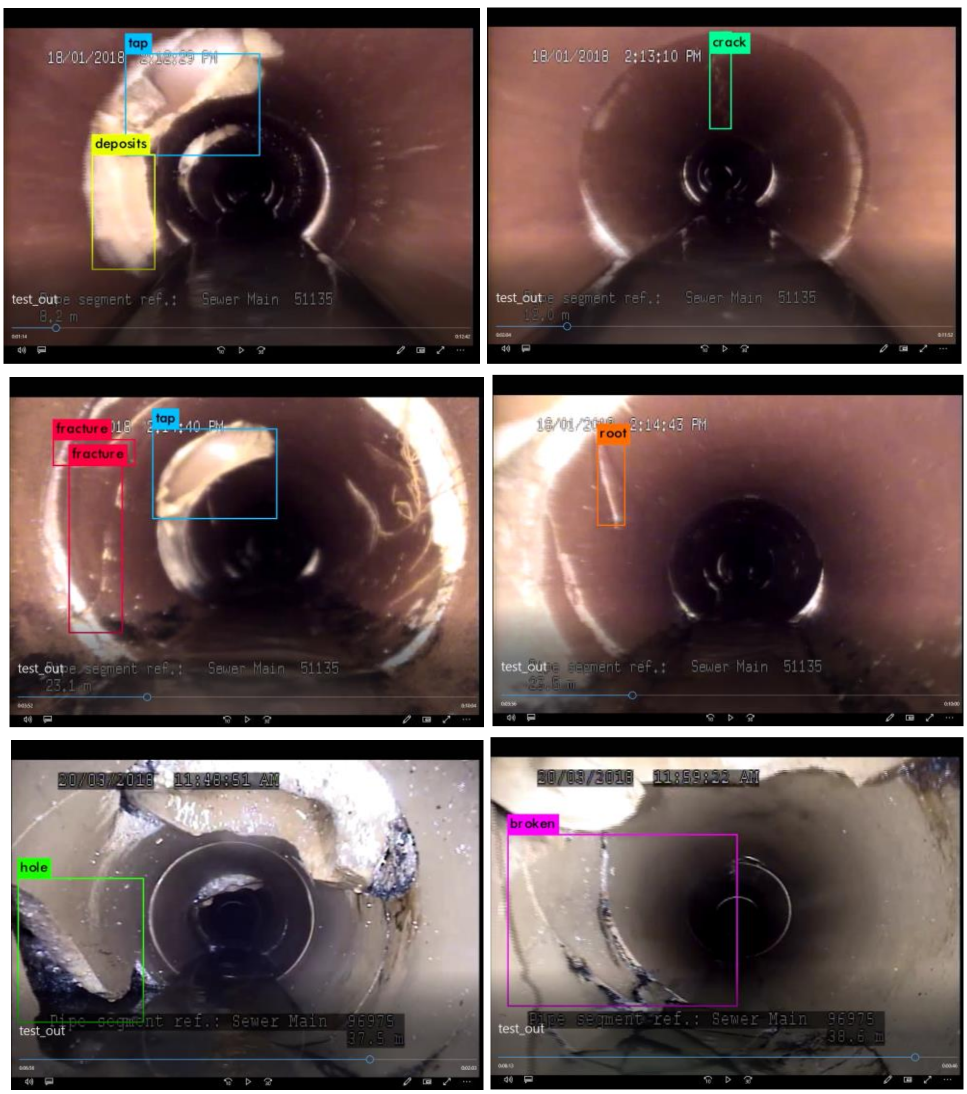

Figure 4: Example of YOLO labelled images

Using the automatic classifier derived from the training set of images, the output of the testing results is a labelled video. As the video plays, the defect will be marked by a bounding box pointing out the corresponding class. Figure 4 shows the example of YOLO labelled images in the video. In addition, a confusion matrix is developed to compare the accuracy of the automatic classifier with the human visual defect identification (see Table 1).

Table 1: Confusion matrix for defect detection

Type	Correct/Total	Mix-up Tap	Mix-up Root	Mix-up Fracture	Mix-up Crack	Mix-up Deposit	Mix-up Hole	Mix-up Broken
Broken	10/10	1	2	2		1	3	
Hole	10/10		1			1		2
Deposit	20/20	2	1		1			1
Crack	20/20	1	1	1		1		
Fracture	19/20	1	1		6		2	2
Root	18/20			4	5	3	1	1
Tap	19/20		1		1	2		

From Table 1, it can be seen that the automatic classifier based on YOLO has high accuracy (i.e., 116/120 = 96%) when identifying the correct type of defects. However, this classifier also tends to mix up other types of defects, such as root, fracture, and crack. For instance, in one of the pipes in the testing dataset, while YOLO did not miss any defects, it did mistakenly detect other defects such as deposits inside a broken tap (see Figure 5). This is, however, not a significant problem since the inspector will go into the filed to check the condition of the tap manually. Additionally, using YOLO also created some false alarms, due to its sensitive detection (see examples in Figure 6). Basically, IOU has a greater influence that results from a small error in a small box (rather than in a large box), and these errors in YOLO are derived from incorrect localization (Redmon et al. 2016).

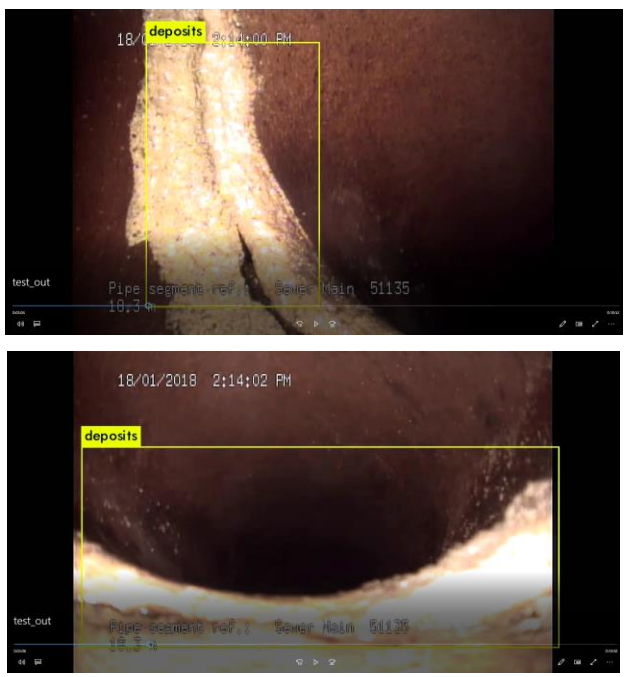

Figure 5: Example of faulty defect detection in broken tap

(a) Too sensitive to detect deposits

(b) Labeling of the same defect twice with two types of classification

Figure 6: Example of false detection by YOLO

4 CONCLUSION

This paper develops a neural network-based method to automatically detect defects for sewer pipes. Three steps are involved in the process: image collection, image pre-processing, and automated defect detection. This approach offers several advantages. First, colour images of defects derived from CCTV monitoring videos are used as inputs directly. In other words, there is no need to convert these images to grey scale as done in previous studies. Second, this approach enables the detection of multiple types of defects simultaneously. Third, the defect classifier is built by means of a novel neural network approach, YOLO, which is superior over traditional methods in the area of object detection. A case study of the sewer network in Edmonton is presented to demonstrate the applicability of the framework to automated defect detection for sewer pipes. The results show that 96% of images are partitioned into the correct classes using the automatic classifier, in spite of some errors such as the mix-up other types of defects. The proposed method will be beneficial when it comes to replacing the traditional method used for defect identification, improving detection productivity, and helping manage labour resources. Future study will focus on improving the classification accuracy and avoiding other interferences by optimizing the YOLO network, making it more suitable to defect detection for sewer pipes. In fact, for a specific class such as crack or fracture, the class can be divided into sub-classes according to defect severity. The automatic classifier is currently only able to identify the large classes of defects, and it does not subdivide these defects into specific classes. Thus, future study will consider the development of a classifier that can better identify each class and subclass of defect.

ACKNOWLEDGEMENTS

We gratefully acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada (CRDPJ 503647-16).

REFERENCES

Buscema, M. 1998. Back propagation neural networks. *Substance use & misuse*, **33**(2): 233-270.

Gupta, A., Vedaldi, A., and Zisserman, A. 2016. Synthetic data for text localisation in natural images. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2315-2324.

Iseley, T., Abraham, D.M., and Gokhale, S. 1997. Intelligent sewer condition evaluation technologies. *Proceedings of the North American NO-DIG Conference*, North American Society for Trenchless Technology, Seattle, WA, USA, 254-265.

Madryas, C. and Przybyla, B. 1998. Inspection of pipes as an element of operating municipal sewerage networks. *Tunneling and Underground Space Technology*, **13**(1): 57-64.

Makar, J.M. 1999. Diagnostic techniques for sewer systems. *Journal of Infrastructure Systems*, **5**(2): 69-78.

Minasny, B., Hopmans, J.W., Harter, T., Eching, S.O., Tuli, A., and Denton, M.A. 2004. Neural networks prediction of soil hydraulic functions for alluvial soils using multistep outflow data. *Soil Science Society of America Journal*, **68**(2): 417-429.

Moselhi, O. and Shehab-Eldeen, T. 1999. Automated detection of surface defects in water and sewer pipes. *Automation in Construction*, **8**(5): 581-588.

Moselhi, O. and Shehab-Eldeen, T. 2000. Classification of defects in sewer pipes using neural networks. *Journal of infrastructure systems*, **6**(3): 97-104.

NASSCO. (2015). *Pipeline Assessment and Certification Program (PACP)*. 7th ed., NASSCO, Marriottsville, MD, USA.

Ning, G., Zhang, Z., Huang, C., Ren, X., Wang, H., Cai, C., and He, Z. 2017. Spatially supervised recurrent convolutional neural networks for visual object tracking. *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS),* Baltimore, MD, USA, 1-4.

Nolan, J. R. 2002. Computer systems that learn: An empirical study of the effect of noise on the performance of three classification methods. *Expert Systems with Applications*, **23**(1): 39-47.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. 2016. You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Las Vegas, NV, USA, 779-788.

Redmon J. and Farhadi, A. 2017. YOLO9000: Better, faster, stronger. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, Hawaii, USA, 7263-7271.

Redmon, J. and Farhadi, A. 2018. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

Sinha, S.K. and Knight, M.A. 2004. Intelligent system for condition monitoring of underground pipelines. *Computer‐Aided Civil and Infrastructure Engineering*, **19**(1): 42-53.

Sinha, S.K. and Fieguth, P.W. 2006. Neuro-fuzzy network for the classification of buried pipe defects. *Automation in Construction*, **15**(1): 73-83.

Wu, B., Iandola, F.N., Jin, P.H., and Keutzer, K. 2017. SqueezeDet: Unified, small, low power fully convolutional neural networks for real-time object detection for autonomous driving. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, Hawaii, USA, 446-454.

Yang, M.D. and Su, T.C. 2008. Automated diagnosis of sewer pipe defects based on machine learning approaches. *Expert Systems with Applications*, **35**(3): 1327-1337.

Yang, M.D. and Su, T.C. 2009. Segmenting ideal morphologies of sewer pipe defects on CCTV images for automated diagnosis. *Expert Systems with Applications*, **36**(2): 3562-3573.