Visual Inspection of Storm-Water Pipe Systems using Deep Convolutional Neural Networks

Ruwan Tennakoon, Reza Hoseinnezhad, Huu Tran and Alireza Bab-Hadiashar
School of Engineering, RMIT University, Melbourne, Australia

Keywords: Storm-Water Pipe Inspection, Automated Infrastructure Inspection, Deep Convolutional Neural Networks.

Abstract: Condition monitoring of storm-water pipe systems are carried-out regularly using semi-automated processors. Semi-automated inspection is time consuming, expensive and produces varying and relatively unreliable results due to operators fatigue and novicity. This paper propose an innovative method to automate the storm-water pipe inspection and condition assessment process which employs a computer vision algorithm based on deep-neural network architecture to classify the defect types automatically. With the proposed method, the operator only needs to guide the robot through each pipe and no longer needs to be an expert. The results obtained on a CCTV video dataset of storm-water pipes shows that the deep neural network architectures trained with data augmentation and transfer learning is capable of achieving high accuracies in identifying the defect types.

1 INTRODUCTION

Condition monitoring of storm-water pipe systems are often carried-out to provide an understanding of the current status of the storm-water system, which enables the prediction of future deterioration of the pipes and facilitate investment planning. These information can also be used in allocating maintenance and repair resources efficiently.

An on-site inspection with closed-circuit television (CCTV) is currently the most common and commercially available method for condition assessment of storm-water pipes. The typical inspection process can be described as follows. A certified technician guides a CCTV camera mounted on a robot that travels inside a pipe segment. The technician must visually detect the defects in the pipe segment by observing the video feed. Once a defect is detected the technician manually rotates and zoom the camera to gain a better understanding of the defect and adds the information relating to that defect (i.e. defect type, defect parameters) to the video together with additional information such as pipe diameter, location, inspection date. The recorded video is then used for further analysis including discrete condition rating, deterioration modelling and planning (Tran et al., 2010).

The above described CCTV inspection is considered semi-automated and is time consuming, expensive and produces varying and relatively unreliable results in some cases due to operators fatigue and novicity. In addition, training a professional technician to be able to classify all the defect types, estimate defect parameters and conduct inspection is costly. Due to the above limitations of the manual inspection process, only around ten percent of the storm-water pipe system in Melbourne, Australia can be inspected given limited budget. Increasing the portion of the inspected pipes would increase the reliability of the network as well as improve the resource allocation and planning processes.

In this paper, we propose an innovative method to automate the defect detection and condition assessment within the pipe inspection process. With the proposed method, the operator only needs to guide the robot through each pipe and no longer needs to be an expert in piping. A computer vision algorithm based on deep-neural network architecture is designed to classify the defect types automatically. The block-diagram of the overall process is shown in Figure 1. In the proposed system, the technician still needs to drive the robot through the pipe and record a clear video of all the internal conditions of the pipe. The video is then fed to the model and the model will go through the video frame by frame to detects the underlying defects in each frame. After successfully detecting a defect, the system extracts those frames with defects and classify the defect type and extract de-

Tennakoon, R., Hoseinnezhad, R., Tran, H. and Bab-Hadiashar, A.
Visual Inspection of Storm-Water Pipe Systems using Deep Convolutional Neural Networks.
DOI: 10.5220/0006851001350140
In *Proceedings of the 15th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2018) - Volume 1*, pages 135-140
ISBN: 978-989-758-321-6

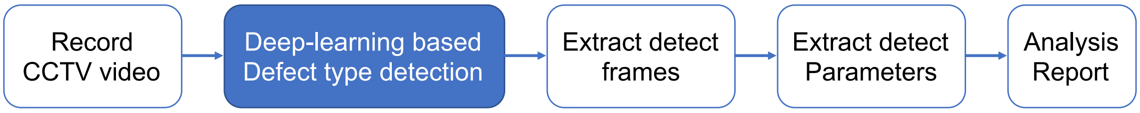


Figure 1: The overall block-diagram of the automated storm-water pipe inspection process.

fect parameters required for condition assessment and further analysis. By applying automated visual inspection, the reliability of the inspection process can be improved. In addition, this automatic system reduces the cost and time in comparison with the manual visual inspection process.

The remainder of the paper is organized as follows: Section 2 provides a review of existing pipe inspection methods and deep neural networks. Section 3 provides a description of the overall method and Section 4 show the results of of our experiments. Finally Section 5 concludes the paper.

2 BACKGROUND

2.1 Automated Inspection of Storm-water/Sewage Pipes

Numerous attempts have been made to automate the pipe inspection process using computer vision and machine learning techniques. Xu et al. (Xu et al., 1998) proposed an automated method for pipe deformation analysis and crack detection that uses image processing techniques such as edge detection and binary image thresholding combine with boundary segment analysis.

Shehab and Moselhi (Shehab and Moselhi, 2005) propose a machine learning based method for infiltration detection in pipes. They first extracted 17 features from images of pipes using a sequence of image processing operations including: dilation, background subtraction, thresholding, and segmentation. These features were then used in a neural network to predict the presence of infiltration which was trained using back propagation. Yang and Su (Yang and Su, 2008) also proposed a machine learning based automatic pipe inspection framework. They extracted texture based features from the image using techniques including wavelet transform and computation of co-occurrence matrices. These features are used with, three machine learning approaches: back-propagation neural network (BPN), radial basis network (RBN), and support vector machine (SVM) to classify pipe defect patterns to following categories: broken pipe, crack, fracture, and open joint. By analysing the above mentioned classifiers they concluded that SVM and RBN are better than BPN.

Yang and Su (Yang and Su, 2009) proposed a pipe defect detection method that utilized both supervised and un-supervised techniques. In their method images from CCTV camera were first converted to a set of features using morphology based segmentation technique. The most important features were then identified using principle component analysis and used in a Radial basis network (RBN) to classify them into one of the following defect types: broken pipe, crack, fracture, and open joint. Su and Yang (Su and Yang, 2014) also proposed a morphological segmentation based method for detecting defects in CCTV video of sewer pipelines. This method was only designed to identify cracks and open joints in pipelines.

Halfawy and Hengmeechai (Halfawy and Hengmeechai, 2014) proposed a method that first extract image region of interest using image segmentation techniques. Next histogram of gradient features were extracted from those regions and used in a SVM classifier to predict weather the region is defective or not.

None of the above mentioned methods are reliable enough to completely replace the current manual inspection due to the limitation of data size, data collection techniques, image processing and pattern classification approaches (Guo et al., 2009). Also, most of them only cover few of the defect types.

2.2 Deep Convolutional Neural Networks

Since winning the ImageNet competition in 2012 (Russakovsky et al., 2015), deep-learning method has gained significant attention in computer vision community with many applications in image classification and segmentation.

Deep convolutional neural networks (CNN) used in image classification comprises of multiple layers of convolution operations coupled with non-linear operations. The output of the convolutional stack is fed through a classification neural network that output the probability of the input image belonging to each of the predifined categories (Krizhevsky et al., 2012). The parameters of the overall network is learned end-to-end using back propagation algorithm on labelled training data. Many CNN architectures has been proposed so far for image classification tasks and, the stateof-the art method include (Simonyan and Zisserman, 2014) , (He et al., 2016) and (Szegedy et al., 2015).

Unlike traditional machine learning that require the features to be hand-crafted, CNNs learn the relevant features from data. However, end-to-end training of a deep neural network requires large amount of labelled data which might not be available for many applications. Several approachers have been proposed to solve this problem including: Unsupervised pre-training of feature layers, data-augmentation and transfer learning (use off-the shelf pre-trained models and fine-tune the final classification layer).

3 PROPOSED METHOD

In this section we describe our proposed method for storm water pipe inspection. The focus of the paper is the novel defect type detection module which is designed to detect five main types of defects found in storm water pipes i.e. 1) Breaking - complete separation of a pipe segment due to a radial crack 2) Cracks - either radial or longitudinal 3) Deposition - sediment build-up on the floor of the pipe 4) Root intrusion - intrusion of tree roots through a gap in the pipes at a crack or at the place where two pipes segments join and 5) Holes. Examples of each defect type is shown in Figure 2. Unlike the existing methods that use hand-crafted features combined with a learned classifier, in this paper we intend to use a deep neural network that learns end-to-end using data alone.

3.1 CNN Architecture and Cost Functions

Given a set of labeled video frames, $\mathcal{X} = [x_i, y_i]_{i=1}^{N}$, where x_i is a video frame, y_i is the corresponding class and N is the total number of training instances, the intention here is to learn a parametrised function $f(x_i;\theta)$ that maps an unseen image to a corresponding class. In this paper we test two network architectures to model this function. The first network was a shallow network with only six layers, which includes three convolution layers (Conv), a Global average pooling layer (GAP) and two fully connected layers (FC). The above model has only a few parameters (670,981 trainiable parameters) compared to typical deep networks, and the architecture is shown in Tabel 1. The next model is based on the well known ResNet-50 architecture (He et al., 2016). This network consists of 50 residual blocks and it is selected as it is a deep architecture that provides an appropriate balance between complexity and accuracy for image classification task. However this architecture has millions of parameters (23,597,957 trainable parameters) and training of the network needs large amount of data.

Table 1: Architecture of the shallow network. Relu stands for Rectified linear units.

	Layer Type	Activation	Shape	Filters
1	Conv	Relu	11x11	128
2	Conv	Relu	5x5	256
3	Conv	Relu	3x3	512
4	GAP			
5	FC	Relu		128
6	FC	Softmax		5

Training of the modes require a loss function that quantifies the errors made by comparing the model output with the supervision signal (ground truth labels for each image). Here we used a categorical cross-entropy as the loss function. The categorical cross-entropy loss function can be written as:

$$\mathcal{L} = \sum_{i=1}^{N}\sum_{j=1}^{C} y_{ij} \ln \hat{y}_{ij} \quad (1)$$

where y_{ij} is the ground-truth indicating whether image i belongs to category j, $\hat{y}_{ij}$ is the predicted probability of image i belonging to category j, N is the number of instances in the dataset and C is the number of detection classes which is 5 in our application.

3.2 Transfer Learning

The way the model parameters are initialized would have a significant effect on the final result. One way to initialize the parameters is to set them to randomly chosen values. This method of model training is called training from scratch and this does not involve any prior information. As a neural network contained millions of parameters training from scratch effectively requires a large dataset. It is not economically feasible to generate such a large dataset in our application.

Another well known method to train networks with limited data is to start from a set of parameters that is trained on a different domain and fine tune the parameter with the limited dataset collected for the task of interest. As we have only a limited dataset we adopted this approach and used the parameters of the ResNet-50 model that was trained on natural image classification task in ImageNet competition (He et al., 2016). The ImageNet challenge involved classifying natural images into 1000 different classes. Because our application involves only five classes and does not map into any class that is in ImageNet competition, we removed the last classification layer of the network and added a new layer which was initialized to random values.

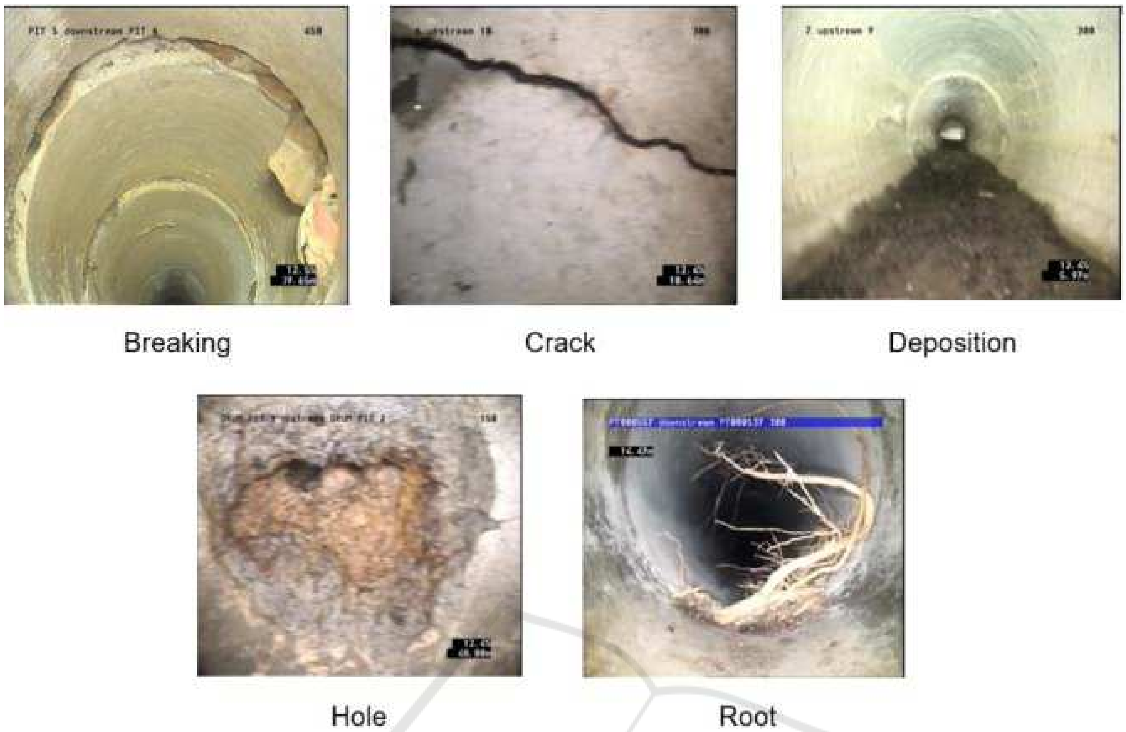


Figure 2: Examples of defect types.

3.3 Model Training

3.3.1 Dataset Preparation

We obtained 90 videos of storm water pipe inspections and most of the video range from 15 to 25 minutes in duration recorded at 25 frames per second. The videos were first divided into training and validation splits of 80% and 20% respectively. These videos were then broken down into images. After converting these videos to images, those images containing defects were taken out and moved to different folders with respect to its type of defects. Each image has a resolution of 720 X 576, which matches the resolution of the video. Since images are generated from video frame by frame, it is inevitable to get many duplicated photos, which would cause over-fitting of the model. A software called “Duplicate Photo Fixer” was used to remove all the duplicated images at 83% similarity. We achieved 13 classes of images at the very beginning, which includes cracking, breaking, hole, sparling, fracture, intrusion, root, steel reinforcement explosion, deposition, water accumulation, and angular, longitudinal and radial joint displacement. But due to image data insufficiency for some of the defect types, we only kept the breaking, crack, deposition, hole and root defect instances.

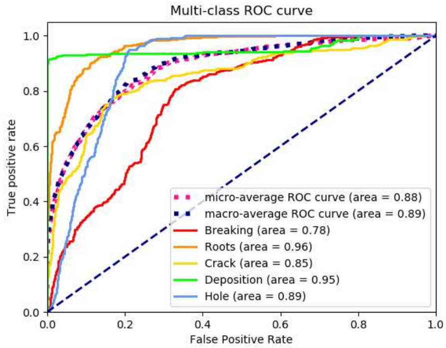
3.3.2 Class Balancing

The resulting dataset was not balanced as it had large number of instances from some classes and few instances of some other classes. Training a model with such class imbalance would result in the model learning to predict only the dominant classes in the dataset. To overcome this issue, we balanced the classes by oversampling the instances in less frequent classes.

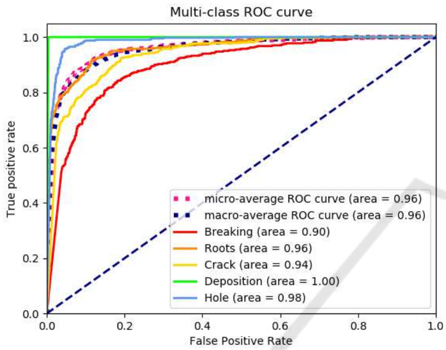
3.3.3 Data Augmentation

Due to the limited number of images and variation of the classes, over-fitting would be a major concern. Consequently, it is a challenge to achieve high classification accuracy by the limited number of data we have. To reduce this issue, we applied images augmentation on the dataset. New images were created by randomly zooming, shearing and horizontally flipping the originals, so that a relatively larger dataset exist to train the model and reduce the likelihood of over-fitting.

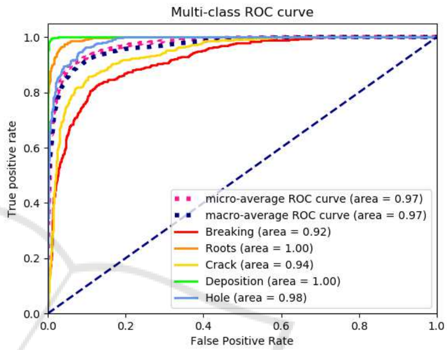
Once the dataset was prepared, we trained the model using ADAM optimization. The network was trained on a Nvidia Titan X GPU with 12 GB of RAM for 150 epoch. The hyper parameters for training are: Batch size: 32, learning rate 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$.



(a) S-net



(b) Resnet-RND



(c) Resnet-TL

Figure 3: Area under ROC curve plots for each tested network types.

Table 2: The confusion matrix on the validation set for the method Resnet-TL. Dep stands for the class deposition.

	Break	Roots	Crack	Dep	Hole
Break	290	33	48	9	20
Roots	7	383	1	3	6
Crack	88	6	300	0	6
Dep	1	0	2	397	0
Hole	33	12	30	0	325

Table 3: Per-Class Precision and Recall of each model.

Class	**S-net**		**Resnet-RND**		**Resnet-TL**	
	Precision	Recall	Precision	Recall	Precision	Recall
Breaking	0.45	0.38	0.65	0.70	0.69	0.72
Roots	0.80	0.70	0.89	0.74	0.88	0.96
Cracks	0.78	0.42	0.73	0.76	0.79	0.75
Deposition	0.71	0.93	0.95	1.00	0.97	0.99
Hole	0.50	0.72	0.88	0.88	0.91	0.81
Average	0.65	0.63	0.82	0.82	0.85	0.85

4 RESULTS

We tested the trained models on a held out validation set created from 20% of the original inspection videos. The validation set consists of 400 images per each category. The evaluations were done using area under the receiver operating characteristics (ROC) curve. We also report the precision and recall for each category.

The results for the shallow network (S-net), Resnet-50 without random initialized weights (Resnet-RND) and Resnet-50 with transfer learning (Resnet-TL) are shown in Figure 3 and Table 3.

The results show that S-net, even with fewer parameters compared to Resnet, has only been able to achieve an overall ROC value of 0.88. However both Resnet with and without transfer learning has been able to obtain high ROC values of 0.97 and 0.96 respectively. The Resnet with transfer learning shows slightly better ROC values in classifying breaking and roots whereas the ROC values across other categories are similar to that without transfer learning. The results indicate that data augmentation has enabled accurate learning of a deep network with limited data in storm-water pipe inspection.

The confusion matrix on the validation set for the method Resnet-TL is shown in Table 2. The confusion matrix shows that there is some misclassification between the classes cracks and breaking. This behaviour is understandable given that the two defect types mentioned above share similar physical characteristics.

5 CONCLUSIONS

The paper presents a new method for automated visual inspection of the storm water pipes. The main novelty of our method is to use a deep convolutional neural network in identifying the defect types. The results obtained on a held out validation set shows that proposed deep neural network architectures trained with data augmentation and transfer learning are capable of achieving high accuracies in identifying the defect types.

In these experiments we have only used five defect types due to the limited availability of data from other categories and we intend to increase this in future work. Defect parameters such as the crack width are also important in decision making and we intend to extend our work towards automated prediction of defect parameters.

ACKNOWLEDGEMENTS

The Authors would like to thank Mr. Juncheng Li for his help with the implementation part of this work.

REFERENCES

Guo, W., Soibelman, L., and Garrett, J. (2009). Automated defect detection for sewer pipeline inspection and condition assessment. *Automation in Construction*, 18(5):587 – 596.

Halfawy, M. R. and Hengmeechai, J. (2014). Automated defect detection in sewer closed circuit television images using histograms of oriented gradients and support vector machine. *Automation in Construction*, 38:1 – 13.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252.

Shehab, T. and Moselhi, O. (2005). Automated detection and classification of infiltration in sewer pipes. *Journal of Infrastructure Systems*, 11(3):165–171.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Su, T.-C. and Yang, M.-D. (2014). Application of morphological segmentation to leaking defect detection in sewer pipelines. *Sensors*, 14(5):8686–8704.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., Rick Chang, J.-H., et al. (2015). Going deeper with convolutions. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Tran, H. D., Perera, B. J. C., and Ng, A. W. M. (2010). Markov and neural network models for prediction of structural deterioration of storm-water pipe assets. *Journal of Infrastructure Systems*, 16(2):167–171.

Xu, K., Luxmoore, A., and Davies, T. (1998). Sewer pipe deformation assessment by image analysis of video surveys. *Pattern Recognition*, 31(2):169 – 180.

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, Part 2):3562 – 3573.