

# DEEP LEARNING AND IMAGE PROCESSING FOR AUTOMATED CRACK DETECTION AND DEFECT MEASUREMENT IN UNDERGROUND STRUCTURES

F. Panella<sup>1,2,\*</sup>, J. Boehm<sup>1</sup>, Y. Loo<sup>2</sup>, A. Kaushik<sup>2</sup>, D. Gonzalez<sup>2</sup>

<sup>1</sup> Dept. of Civil, Environmental and Geomatic Engineering, University College London, Gower Street, London, WC1E 6BT UK – (fabio.panella.16, j.boehm)@ucl.ac.uk

<sup>2</sup> Ove Arup & Partners Ltd., 13 Fitzroy Street, London W1T 4BQ – (Yung.Loo, Akshat.Kaushik, David.Gonzalez)@arup.com

Commission II, WG II/10

**KEY WORDS:** Deep Learning, Automated Crack Detection, Photographic Tunnelling Surveys.

## ABSTRACT:

This work presents the combination of Deep-Learning (DL) and image processing to produce an automated cracks recognition and defect measurement tool for civil structures. The authors focus on tunnel civil structures and survey and have developed an end to end tool for asset management of underground structures. In order to maintain the serviceability of tunnels, regular inspection is needed to assess their structural status. The traditional method of carrying out the survey is the visual inspection: simple, but slow and relatively expensive and the quality of the output depends on the ability and experience of the engineer as well as on the total workload (stress and tiredness may influence the ability to observe and record information). As a result of these issues, in the last decade there is the desire to automate the monitoring using new methods of inspection. The present paper has the goal of combining DL with traditional image processing to create a tool able to detect, locate and measure the structural defect.

## 1. INTRODUCTION

### 1.1 Tunnel inspection: traditional methods and new technologies

This paper presents a novel pipeline for meaningful visual defect detection. The authors focus on tunnel surveys as they pose several challenges (McKibbins, Elmer, & Roberts, January 2010.). In order to keep the serviceability of such tunnels, regular inspections are needed to assess their structural status.

At the time of writing, the standard industry monitoring procedure consists of regular visual inspections carried out by an expert operator who has knowledge of the monitored structure and its material. This kind of inspection is simple, but slow and relatively expensive and the quality of the output depends on the ability and experience of the engineer as well as on the total workload (stress and tiredness may influence the ability in detecting information). As a result of these issues and progress in technology over the last decade, there has been the desire to automate the monitoring process using new methods of inspection; the main purpose is to reduce the workloads for the engineers and to obtain reliable outputs. The reasons are various and can be divided into several classes:

- a) Limitations of visual inspection
  - The reliance on visual inspections depends not only on the visibility of the defect but also on the subjectivity of observations;
  - To reduce the closure time of the infrastructure (particularly in the case of transportation tunnels), the investigations are often carried out during the night shift; because of that, another limit of the visual investigation may be the stress level of the operator.
  - Lack of continuity of inspectors or inspection methods can lead to reduction in the effectiveness of inspections and confidence in results.

- b) Safety reasons

- The underground structures represent in some cases, like sewer pipes, a confined space and operators need to be trained;
- Several hazards may put the health of the operators at increased risk.

- c) Speed and economic reasons

- Traditional visual inspection needs an engineer to go down the tunnel and record all the defects resulting in long shifts. The economic disadvantage is double: on one side the cost on site of the engineer (mainly during night shifts), on the other side the closure of the infrastructure.

All these factors might influence the results.

Recently many engineering companies have begun to use and develop sophisticated techniques to create high resolution 360° photographic datasets of their project sites for inspection purposes (McDonnell & Devriendt, 2017). However, the process is still applicable to a general scenario.

The use of immersive photographic tunnelling surveys has the great advantage of speeding up the inspection process, resulting in cost savings and shorter shifts in the tunnel (with the further benefit of improved health and safety implications). These photographic surveys can replace the visual inspection: the immersive tours give the possibility to spend a minimum amount of hours on the site (the time needed to collect pictures, with an appropriate mount, is less than 1 min/metre) and to process and analyse the inspection from the office.

To overcome other limitations, such as subjectivity of the observations and the dependency of the output on the stress level of the engineer, a new processing and analysis pipeline has been developed by the authors. The idea is to collect photos and extract a posteriori all the needed information with image processing and image analysis. In detail, the proposed pipeline outputs a meaningful defect detection providing the user with metrics, localisation and extents of the defect. The authors have chosen

---

\* Corresponding author

cracks as the defect of choice to be detected because in the underground structures this is one of the more prevalent defects which is possible to detect and the monitoring of the cracks can give important information about the structural behaviour of the structure.

Table 1 gives a qualitative comparison between the traditional Visual Inspection (VI), the 360° Immersive View Inspection (360-IVI) and the novel pipeline developed by the authors (CrackDet). The comparison is expressed with values from 1 (non-functional) to 5 (fully functional). Can be seen that the repeatability of the new pipeline is as limited as 360-IVI. This restriction might be resolved creating a 3D model using photogrammetric technique and applying the crack detection to the textured mesh.

Table 1: comparison between the inspection methods

Inspection Method	Visual Output	Objectivity	Repeatability	Automation	Quantitative Output
VI	3	2	1	0	0
360-IVI	3	3	3	1	0
CrackDet	4	4	3	4	3

## 1.2 Paper structure

The paper articulates as follows:

2. Related works and relative results;
3. Proposed pipeline:
  - 3.1. Introduction;
  - 3.2. Image acquisition;
  - 3.3. Image processing;
  - 3.4. Image classification;
    - 3.4.1. Comparison between two image classification networks.
  - 3.5. Outputs.
4. Future research.

In chapter 2 several state of the art methods for automated crack detection in underground structures or buried pipes are analysed. The costs in terms of computing time and accuracy are compared against CrackDet.

In chapter 3 the pipeline proposed by the authors is analysed step by step and the results are summarised.

Chapter 4. contains the final considerations and the proposals for future research.

## 2. RELATED WORKS

### 2.1 Introduction

Several works were proposed for the autonomous crack detection in underground structures or buried pipes. Three main workflow are identified:

- Image processing only;
- Image classification only (image processing is limited to image tiling);
- Combination of image processing and Machine learning.

#### 2.1.1. Image processing based crack detection

(Sinha & Iyer, 2005.) propose a high-level image processing based on an initial contrast enhancement to highlight the dark pixels, morphological transformations to clean the image from small connected objects, Laplacian or Gaussian as an object detector (considering that the crack intensity in the image has a Gaussian shape; so, it is possible to clear the image from all the connected objects with a different gradient of intensity);

combination of morphological transformation for a final cleaning.

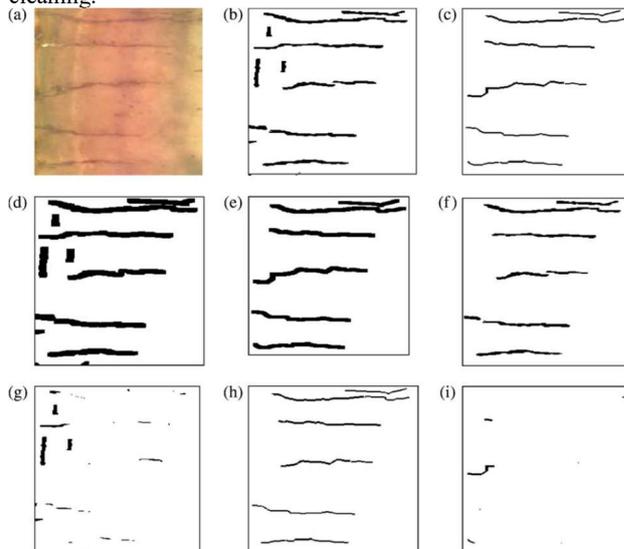


Figure 1: Matching procedure for detection of true and false pixels; (a) original image, (b) detected cracks, (c) ground truth, (d) and (e) detected and true cracks dilated by a 5X5 structuring element. (f) good points of the filter, (g) false +ve, (h) truly detected cracks and (i) missed cracks (false -ve).

(Sunil K. & Paul W., 2006) propose the two step algorithm: The first step is local and uses statistical properties to extract crack features from the segmented image, which are treated as crack segment candidates. In the second step, global cleaning and linking operations merge segments to form cracks.

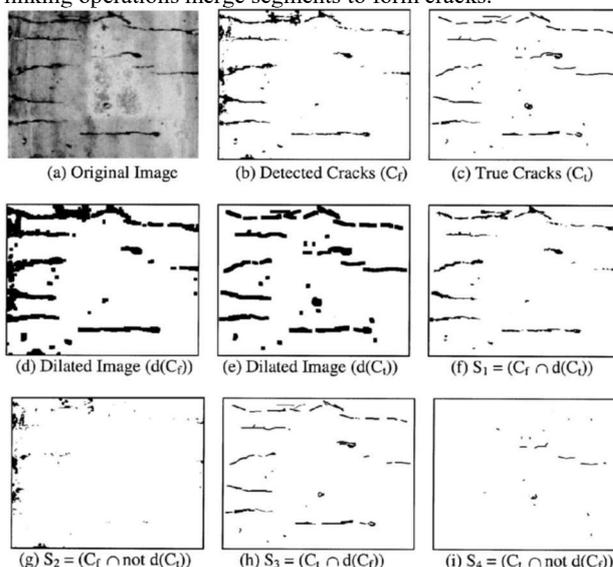


Figure 2: Illustration of the procedure of matching the images for detection of true and false pixels.

They also evaluate the probability of detection ( $P_d$ ) and false-alarm ( $P_{fa}$ ). This probability depends on the value of the threshold.

(Dapeng Qi, Yun Liu, Qingyi Gu, & Fengxia Zheng, October 2014.) focus on the importance of making the whole process fast and real time. For that purpose, the set-up consists of several linear CCD cameras mounted on the front of a train and all the equipment needed for the image processing and object classification is installed within the car itself. In order to be able to make the process fast enough, the image processing has to be simple. Their proposed algorithm is depicted below (Figure 3).

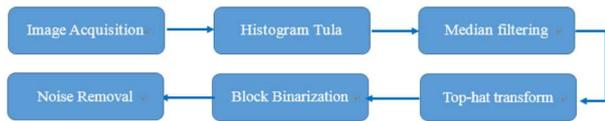


Figure 3: Procedure of the algorithm proposed by (Dapeng Qi, Yun Liu, Qingyi Gu, & Fengxia Zheng, October 2014.)

### 2.1.2. Image processing and ML based crack detection

(Wenyu, Zhenjiang, Dapeng, & Yun, 2014) consider the same set-up showed in (Dapeng Qi, Yun Liu, Qingyi Gu, & Fengxia Zheng, October 2014.) but determine the variables needed to classify correctly the object (Crack or Not-crack classes). They explain that only three features are effective for the classification:

- Standard deviation of shape distance histogram: irrelevant objects are patterns with irregular shapes, while cracks or crack-like objects have slender patterns. In general, the distance distribution of irrelevant objects with an irregular shape is heterogeneous. Thus, their standard deviations are larger than those of other classes.
- Pixel number: The pixel numbers of the irrelevant objects are always less than those of cracks. the statistical results of 200 candidate objects show that the 21 objects with the largest pixel numbers are all cracks.
- Average grey level: compared to cracks, many local dark regions are bright regions with high grey levels. Some misidentified objects look just like cracks in binary presentation, but they can be classified by their average grey levels in the original images.

The authors also test this pipeline with several classifier providing the respective Training and Test accuracy (Table 2).

### 2.1.3. Image classification based crack detection

(Young-Jin Cha, Wooram Choi, & Oral Büyüköztürk, 2017) this article proposes a vision-based method using a deep architecture of convolutional neural networks (CNNs) for detecting concrete cracks without calculating the defect features. As CNNs are capable of learning image features automatically, the proposed method works without the conjugation of IPTs for extracting features. The trained CNN is combined with a sliding window technique to scan any image size larger than  $256 \times 256$  pixel resolutions.

## 3. PROPOSED PIPELINE

### 3.1 Introduction

The proposed workflow combines the advantages of image processing, such as the possibility of extracting information from the segmented images, with the advantages of DL, such as the possibility to develop a classifier without providing any features. This last aspect is important because the vectors of features are highly dependent on lighting conditions and on the filtering

algorithm adopted. The present paper has the goal of creating a more generic purpose tool.

The pipeline proposed in the present paper is summarized in Figure 4.

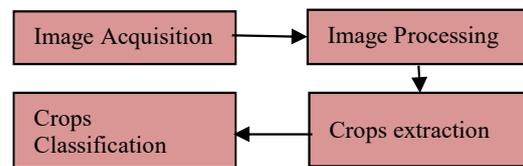


Figure 4: Proposed Pipeline

### 3.2 Image acquisition

At the time of writing the full set-up for image acquisition is being established, with trials and testing under development. Considerations for the setup include: Final measurement in the metric system (not in pixels) and attention during the acquisition process of the following parameters:

- Average distance from the wall: in the environment of underground structures it is related with the diameter of the structure itself and it is easy to keep it within a small range;
- Focal length: in order to fully automate the process, it is preferable to set the focal length at the beginning and to keep it fix during the survey;
- Horizontal sensor size and number of pixel; the ratio of these values give the dimension of a single pixel.

Through determination of these parameters, it is possible to realise a triangulation and to obtain a metric measurement of the detected cracks.

### 3.3 Image processing and Crops extraction

Several algorithms were tested. (Sinha & Iyer, 2005.) gives the best results in terms of isolating the cracks but it is highly time consuming: it is based on several geodesic reconstructions and each of them has to iterate 18 times (a linear structuring element rotating in the range  $0-180^\circ$  with a step of  $10^\circ$ ). Considering that, the authors decided to adopt an easy image processing algorithm based on:

- Low level contrast enhancement to highlight dark pixels;
- Median blur: this is highly effective against salt-and-pepper noise in the images;
- Adaptive Gaussian Threshold;
- Edge detection;
- Creation of the minimum area bounding box for each contour;
- Cropping out of the bounding boxes.

The low-level algorithm proposed is time effective. However, many misleading elements are segmented along with the true cracks. Therefore, there is a need for an image classifier.

Figure 5 depicts the output of the image processing step.

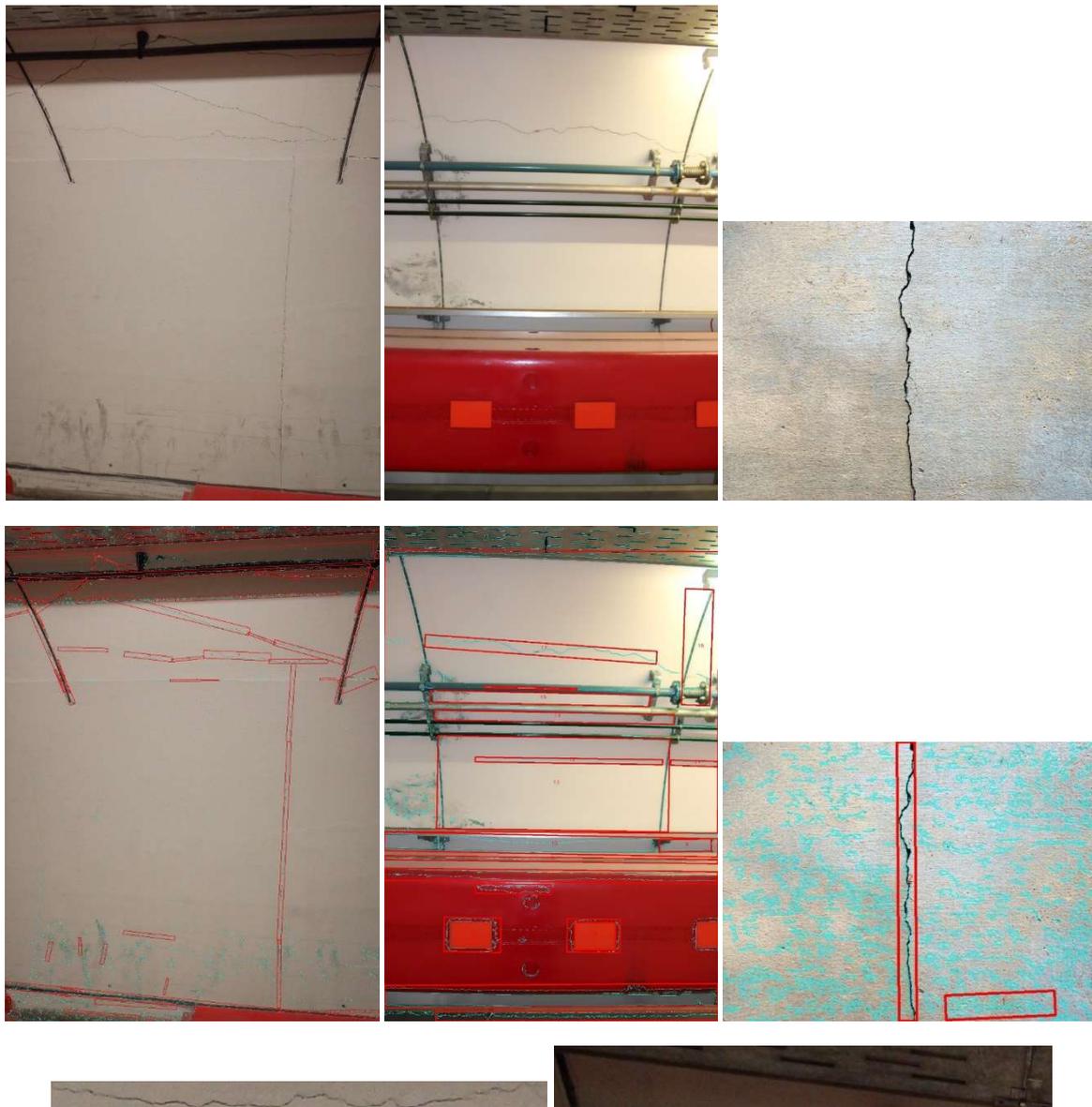


Figure 5: original images (top), image processing output (middle), example of cropped crack from the whole image (bottom)

### 3.4 Image classification

(Samuel, 1959) defined Machine Learning (ML) as the “field of study that gives computers the ability to learn without being explicitly programmed”. At its most basic it is the practice of using algorithms to parse data, learn from it, and then decide or predict something in the world. So rather than hand-coding software routines with a specific set of instructions to accomplish a particular task, the machine is “trained” using large amounts of data that gives to the machine the ability to learn how to perform the task. (Wenyu, Zhenjiang, Dapeng, & Yun, 2014) developed their algorithm based on feature of vectors. The challenge in this case is that ML neurons need to be fed with engineering features and it might be complex to find the right combination of features to correctly classify the objects. That task becomes even harder if the aim is to realise a general-purpose tool.

To overcome the aforementioned problem, Deep Learning (DL), a special discipline in ML is required for classification purposes. Neural Networks are inspired by our understanding of the biology of our brains – all those interconnections between the neurons. You might, for example, take an image, chop it up into

a bunch of tiles that are inputted into the first layer of the neural network which trigger the layers of hidden units, and these in turn arrive at the output units. Each unit receives inputs from the previous units, and the inputs are multiplied by the weights of the connections they travel along. Every unit adds up all the inputs it receives in this way and (in the simplest type of network) if the sum is more than a certain threshold value, the unit “fires” and triggers the units it’s connected to (those on its right).

In the present work the image processing is carried out in order to automatically crop out the object of interest from the whole images. The crops will feed the DL classification tool and they will be classified as Crack or No\_Crack.

To train a network from scratch requires hundreds of thousands of labelled images. In order to deal with the availability of a smaller dataset (188 crack\_objects and 188 no\_crack\_objects), pre-trained networks are adopted for the purpose. That process is called Transfer Learning. It is commonly used in deep learning applications. Fine-tuning a network with transfer learning is much faster and easier than constructing and training a new network. The advantage of transfer learning is that the pre-trained network has already learned a rich set of features. These features

can be applied to a wide range of other similar tasks. To use that networks for the purpose of crack detection, only the last layers (the fully connected ones) need to be changed introducing the desired labels.

Two architectures were used and compared by the authors: AlexNet and GoogleNet in order to compare their performances in terms of precision and accuracy as well as in terms of time consumed for training and testing.

According to (Culurciello, Canziani, & Paszke, 2017) AlexNet, the oldest architecture should achieve the lowest top-1 one-crop accuracy versus amount of operations required for a single forward pass whilst GoogleNet should provide a good compromise between these parameters (Figure 6).

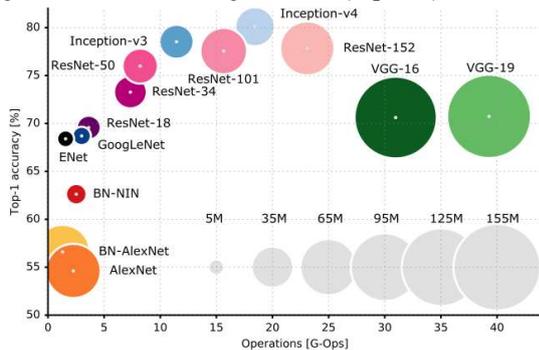


Figure 6: Comparison of different architectures (Culurciello, Canziani, & Paszke, 2017).

Upon comparison of the two architectures, with respect to the Authors' crack detection purposes, some unexpected results arose:

- With a such small dataset, the training time is around 10 seconds using a GPU and around 30 minutes using only the CPU for both architectures.
- The results in terms of precision and accuracy look to be in conflict with the comparison made by (Culurciello, Canziani, & Paszke, 2017): on a test dataset with 244 candidates of which only 27 cracks (the unbalanced test set was chosen to create conditions similar to the real environment: from a real survey the crack elements are certainly less than no\_crack elements) AlexNet achieved the higher accuracy and precision (Table 2).

The results are promising, considering the small and random dataset, with an average Accuracy (proportion of correctly classified instances) over 98%, a Precision (proportion of positives that are classified correctly) over 92% and a Recall (true positive rate) of 94%.

### 3.5 Outputs

After the classification, it is possible to ground truth the cracks with respect to each image (Figure 7). In addition to this, collect information about the location, the orientation and the dimensions in the object space (Table 3) can be collected. Such information is obtained as follows:

- Location: position of the centroid of the crack element in the image space;
- Orientation: assuming that the aspect ratio of the crack elements (length/height of the minimum area bounding box) is quite high, the orientation has being considered as the inclination of the maximum dimension of the bounding box itself (0° is horizontal);
- Dimensions: from image processing it is possible to extract Area and Perimeter of the connected objects in the image space. The assumption to obtain the length and the width of the crack is to consider the structural defect as a regular polygon. With this assumption, it can be said that the width has no influence, compared to the length, in the value of the perimeter. Therefore the length of the crack can be evaluated as half of the perimeter (Eq. 1). Knowing the length dimension, the width can be approximately evaluated as the area divided by the length (Eq. 1)

$$\begin{cases} l_{px} = P/2 \\ w_{px} = A/l \end{cases} \quad (1)$$

Where  $l_{px}$  = crack length [pixel];  
 $P$  = crack perimeter[pixel];  
 $w_{px}$  = crack width[pixel];  
 $A$  = crack Area [pixel].

To translate these values from the image space to the metric system, the following triangulation has being considered (Eq. 2):

$$\begin{cases} sf = WD \cdot px/f \\ l_{mm} = sf \cdot l_{px} \\ w_{mm} = sf \cdot w_{px} \end{cases} \quad (2)$$

Where  $sf$  = scale factor;  
 $WD$  = working distance from the wall [mm];  
 $l_{mm}$  = length [mm];  
 $w_{mm}$  = crack width [mm].

At the time of writing, the precision and accuracy of these types of readings is solely dependent on the capability and observation of the user in taking notes of the parameters presented in 3.2.

Table 2: comparison of different Algorithms/Networks in terms of accuracy of the prediction.

Algorithm/Network	Training set			Test set		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
<i>CrackDet (AlexNet)</i>	98.0%	92.0%	92.5%	93.9%**	72.4%**	77.8%**
<i>CrackDet (GoogleNet)</i>	-	-	-	80.3%***	35.3%***	88.9%***
<i>ELM*</i>	98.5%	-	-	91.6%	-	-
<i>RBF*</i>	96.5%	-	-	90.1%	-	-
<i>SVM*</i>	98.0%	-	-	91.3%	-	-
<i>KNN*</i>	-	-	-	88.7%	-	-

\* from (Wenyu, Zhenjiang, Dapeng, & Yun, 2014).

\*\* 21 True Positives out of 27, 207 True Negatives out of 223, 8 False Positives and 6 False Negatives.

\*\*\* 24 True Positives out of 27, 171 True Negatives out of 223, 44 False Positives and 3 False Negatives.

Table 3: Example of output

Object ID	angle [deg]	length [mm]	width [mm]	description
19_01_2018_12_57_Image_95_crop_10.jpg	-85.6	56.7	0.9	Discontinuous-cracking-parallel-to-circle-joints
19_01_2018_12_57_Image_95_crop_11.jpg	-89.3	43.1	1.2	Discontinuous-cracking-parallel-to-circle-joints
19_01_2018_12_57_Image_95_crop_12.jpg	-88.4	735.7	1.7	Circumferential-crack
19_01_2018_12_57_Image_95_crop_13.jpg	-8.7	65.0	1.3	Crack-parallel-to-cross-joints
[...]	[...]	[...]	[...]	[...]

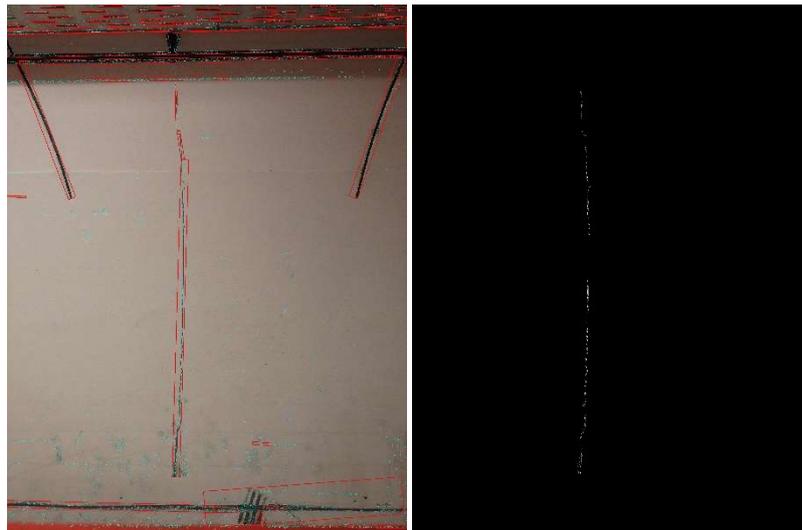


Figure 7: After image processing (left), Crack Ground Truth after image classification (right)

#### 4. CONCLUSIONS AND FUTURE RESEARCH

The pipeline is simple, and also fast and robust. It shows a good accuracy in the total length of the detected and correctly classified defects. Of note, the dataset is not homogenous, and it shows that the proposed tool can be considered a general purpose crack detector, provided that a bigger dataset will be given to train the DL algorithm.

The robustness of the proposed pipeline might be improved adding the depth channel to the processed images. To do this, without losing the time efficiency given by the single shoots, an ad hoc camera network has to be developed.

Lastly, in order to make the overall process faster and faster, the image classification has to be replaced by the image semantic segmentation. This process will make the process fully automated and independent of the human requirement to calibrate the image processing parameters. The process will enable segmentation and measurement of any other kind of defect within the underground structures, without the need for different scripts each time. Based on a trained FCN, the new process will detect the elements included in the training dataset and it will perform the segmentation of such elements.

#### 5. REFERENCES

- Culurciello, E., Canziani, A., & Paszke, A., 2017: An analysis of deep neural network models for practical applications. Retrieved from <https://towardsdatascience.com/neural-network-architectures-156e5bad51ba>
- Dapeng Qi, Yun Liu, Qingyi Gu, & Fengxia Zheng, 2014: An Algorithm to Detect the Crack in the Tunnel Based on the Image Processing.
- McDonnell, D., & Devriendt, M., 2017: The Application of BIM and Photogrammetric Methods to the Inspection and Management of the Sydney Harbour Tunnel. 16th Australian Tunnelling Conference.
- Mc Kibbins, L., Elmer, R., & Roberts, K., 2010: Tunnels: inspection, assessment and maintenance. CIRIA C671.
- Samuel, A. L. 1959: Some Studies in Machine Learning Using the Game of Checkers. IBM Journal of Research and Development.

Sinha, S., & Iyer, S., 2005: A robust approach for automatic detection and segmentation of cracks in underground pipeline images. *Image and Vision Computing*, vol. 23.

Sunil K., S., & Paul W., F., 2006: Automated detection of cracks in buried concrete pipe images.

Wenyu, Z., Zhenjiang, Z., Dapeng, Q., & Yun, L., 2014: Automatic Crack Detection and Classification Method for Subway Tunnel Safety Monitoring. *Sensors*, vol. 14.

Young-Jin Cha, Wooram Choi, & Oral Büyüköztürk., 2017: Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks. *Computer-Aided Civil and Infrastructure Engineering*.