Smart Image Based Technology and Deep Learning for Tunnel Inspection and Asset Management

Fabio PANELLA[1,2], Yung LOO[1], Michael DEVRIENDT[1], David GONZALEZ[1], Akshat KAUSHIK[1], Richard OLLERHEAD[1], Jan BOEHM[2]

1 Ove Arup & Partners Ltd., 13 Fitzroy Street, London W1T 4BQ, UK – (Yung.Loo, Fabio.Panella, David.Gonzalez, Akshat.Kaushik, Richard.Ollerhead@arup.com)

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

Abstract:

Tunnel inspection and asset management is typically a labour-intensive process where engineering judgement and experience is heavily relied upon to identify and assess tunnel condition over kilometres of homogeneous structures. Novel work flows and digital applications have been developed by the authors to create various smart image-based inspection and analysis tools that reduce the potential subjectivity and inconsistency of these inspections. This has resulted in significant improvements to existing tunnel inspection practices and structural health assessment.

Current advances in image capture technology and computational processing power has enabled high integrity data to be easily captured, visualised and analysed. The work flows and tools developed take advantage of existing low-cost image capture hardware, open-source processing software and couples this with the creation of unique machine learning algorithms and analytics.

Core innovations include: (i) use of low-cost photographic equipment for high quality imagery capture (ii) use of automated inspection vehicles for data capture (iii) Deep learning for automatic defect object recognition and defect classification (iv) Creation of immersive dashboards and 3D visualisations.

This results in a suite of image based service offerings and deliverables, relevant to specific tunnel engineering issues and asset management aims. Thanks to deep learning, defect detection and asset condition metrics are automatically created, enabling: (i) the tunnel owner to gain greater insights into their asset resilience and operations, (ii) the tunnel engineer to focus on key issues aided by machine learning.

Keywords: Deep Learning, Automated Crack Detection, Photogrammetric Tunnelling Surveys, Underground Infrastructure, Safety and Human Factors, Technical Approaches and Innovations.

1. Introduction: Benefits of the smart image based tunnel inspection approach

The Institution of Structural Engineers defines structural engineering as "*the science and the art of designing and making, with economy and elegance, buildings, bridges, frameworks, and other similar structures so that they can safely resist the forces to which they may be subjected for a specified life span.*"

(Brian, 1997) augmented this definition to account for durability and serviceability of structures, proposing that in order to quantify these entities and to keep the structure into the service zone, a monitoring and maintenance strategy is needed to confirm the designed prediction made at design stage and to assess the timed and targeted maintenance.

The development of a strategy needs the knowledge of the present and historical conditions of the structure. This process can be schematised as shown in Figure 1.

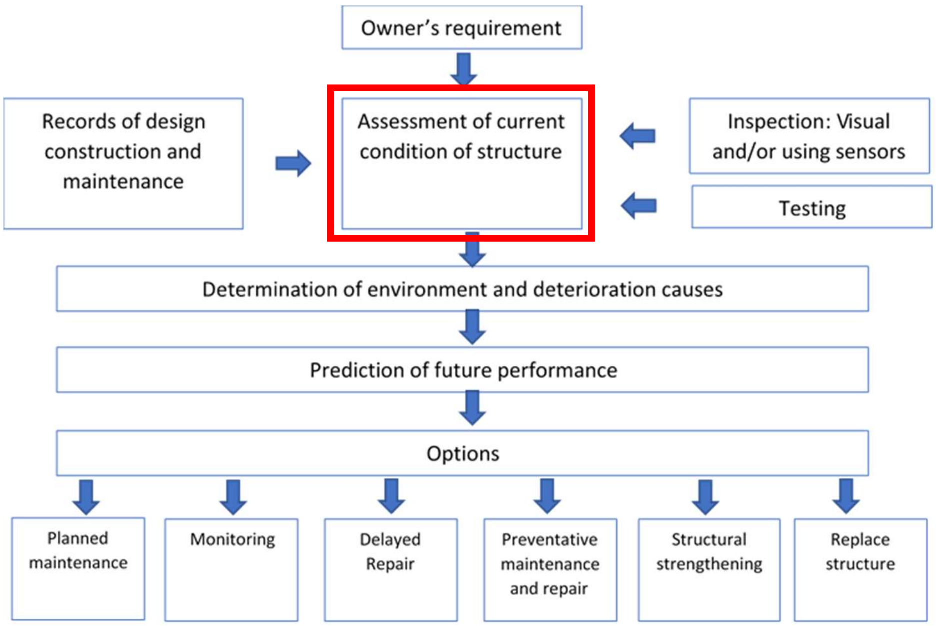


Figure 1: Development of a maintenance strategy for a single concrete structure (Menzies, 1997)

The authors' research has focused on the development of a strategy to automate the monitoring of railway tunnels to assess the current conditions of the structure (notwithstanding its applicability to other tunnel use types e.g. power, highways). At the time of writing, the standard industry monitoring procedure consists of regular visual inspections carried out by an expert operator who has the knowledge of the monitored structure and its material. The operator also has to attach photographic evidence to the detailed report to the historical documents of the structure in order to keep a record of condition along the designed life of the structure itself. This technique is relatively expensive and slow (time of closure of the asset being a further significant cost and operations loss for the client). Furthermore, it has a low level of repeatability and it strongly relies on the subjectivity and sensitivity of the operator.

The authors present a novel pipeline for meaningful visual defect detection. The authors focus on tunnel surveys as they require regular inspections to assess their structural status, and pose several challenges (Mc Kibbins, et al., January 2010) whose issues can be widely expanded upon across other linear infrastructure asset types.

The purpose of the research is to overcome the limitation of the traditional approach; the authors aim to reduce the time of closure of the infrastructure, with direct benefits in terms of costs, health and safety, and to generate consistent output after every survey so that the subjectivity and sensitivity of the engineer, as well as the continuity of the inspectors will not adversely impact the level of detail and conclusions of the survey outcomes.

Recently some engineering companies have begun to use and develop sophisticated techniques to create high resolution 360º photographic datasets of their project sites for inspection purposes (McDonnell & Devriendt, 2017). 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 occupy the site for a fewer number of hours (i.e. the time needed to acquire images, 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 workflow involves collection of images and extraction a posteriori of all the required 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 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.

The novel pipeline is outlined in the following sections. This has been used successfully on a number of tunnel assets, providing the asset manager with high levels of repeatability, automation and objectivity, as well as detailed reporting in terms of key metric information.

2. Related works

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

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

Sinha & Iyer (2005) proposes 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. Sunil & Paul (2006) proposes 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. They also evaluate the probability of detection (P_d) and false-alarm (P_{fa}). This probability depends on the value of the threshold. Dapeng, et al. (2014) focuses 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, so it is based on histogram enhancement, median blurring, dark pixels picking (top-hat transform), block binarization and noise removal. Wenyu, et al. (2014) considers the same setup proposed in (Dapeng, et al. 2014) but determines 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, pixel number and average grey level. The authors also test this pipeline with several classifiers providing the respective training and test accuracy (Table 1). Young-Jin, et al. (2017) proposes a vision-based method using a deep architecture of convolutional neural networks (CNNs) for detecting concrete cracks without extracting handcrafted features. As CNNs are capable of learning image features automatically, the proposed method works without the conjugation of image processing techniques for extracting features. The trained CNN is combined with a sliding window technique to scan any image size larger than 256×256 pixels resolutions and produce the automatic crack detection (Figure 7). Panella, et al. (2018) developed their first approach to the topic blending the classic image processing and the convolutional network power in order to produce an automatic classification of cracks and their segmentation. The authors decided to adopt a time effective low-level image processing which, therefore, needs a convolutional network classifier to eliminate the outliers.

3. Proposed pipeline

The end to end pipeline proposed in the present research is depicted in Figure 2.

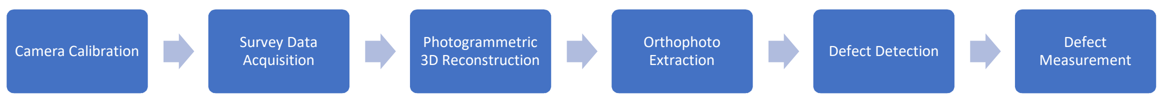


Figure 2: Workflow proposed by the authors

The data acquisition component is repeatable and low-cost, with the authors having implemented a standardised process for data collection adaptable to conventional tunnel dimensions and layouts. To meet all these requirements, point and shoot action cameras with fisheye lenses (120° field of view) have been mounted on a rigid aluminium rig. An operator is required to push the rig along the tunnel and trigger all the cameras simultaneously using a wi-fi remote control. This is conventionally undertaken manually, semi-manually (with the aid of an in-situ operating aid e.g. tunnel crown I-beam fixing), or remotely operated (with the use of a remotely operated vehicle). The direction of travel for the operation is towards automatic and remotely-operated, however this is highly reliant on infrastructure within the tunnel, and asset management procedures allowing for this possibility.

The camera network has been designed in line with standard photogrammetric good practice for corridor mapping, as is the case in tunnels. Pix4D (undated) recommends 85% frontal overlap and 60% side overlap (if the corridor is acquired using more than one flight line). The designed camera network consists of a set of 4 GoPro cameras rigidly connected as shown in Figure 3. In order to avoid changes in the camera parameters, the camera calibration has to be performed soon before or after the survey, being careful not to provoke shocks to the cameras. In the case study the camera calibration has been performed with the self-calibration approach.

The geometry of the inspected infrastructure does not need the knowledge of the markers' coordinates if the calibration has been correctly performed. The exact location in global or local coordinate of the targets is needed to give a scale to the model and to increase the performances of the reconstruction in terms of global accuracy. The first test of the process has been performed on a virtual tunnel created with a 3D computer graphics software (Blender, undated). The calibration parameters of a real GoPro camera were used in the virtual environment. Figure 4 shows the output of the photogrammetric reconstruction and the unwrapped mesh of the model. Once unwrapping of the lining has been performed, it can be exported as an image and used as input for the automatic crack detection.

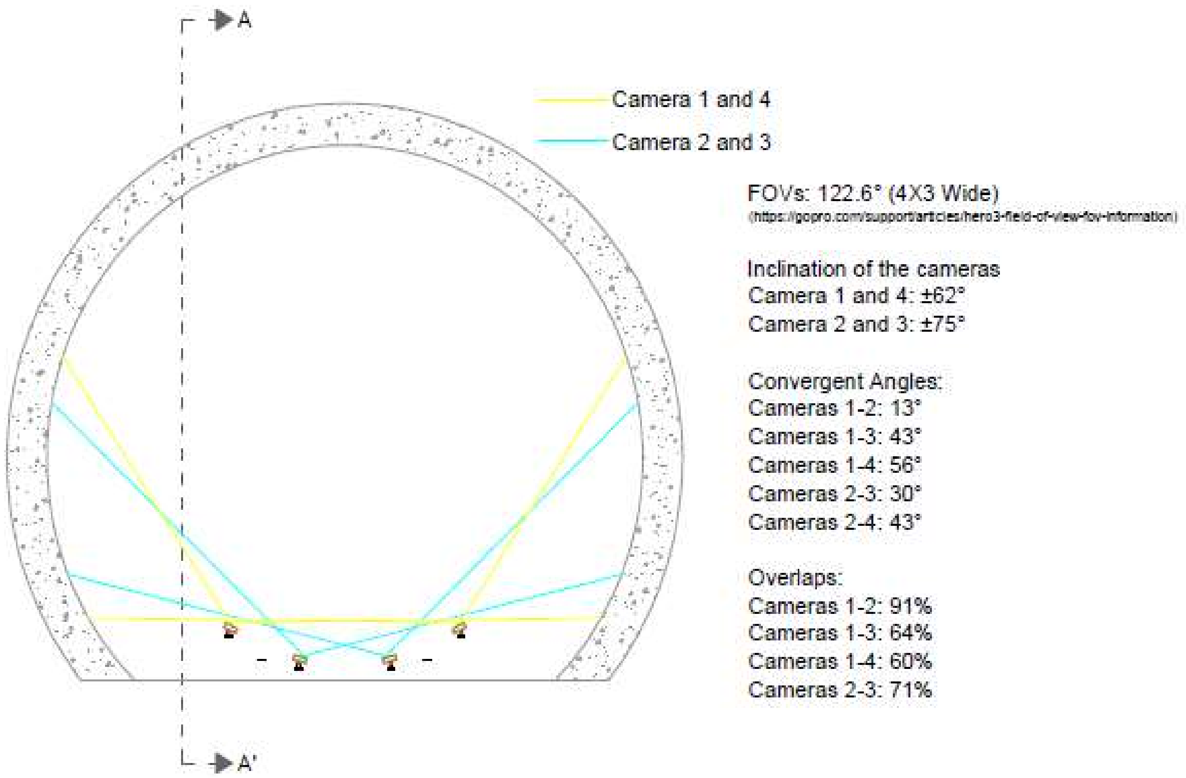


Figure 3: Camera network setup: the shooting step is 0.50m in order to have an overall overlap of 75%

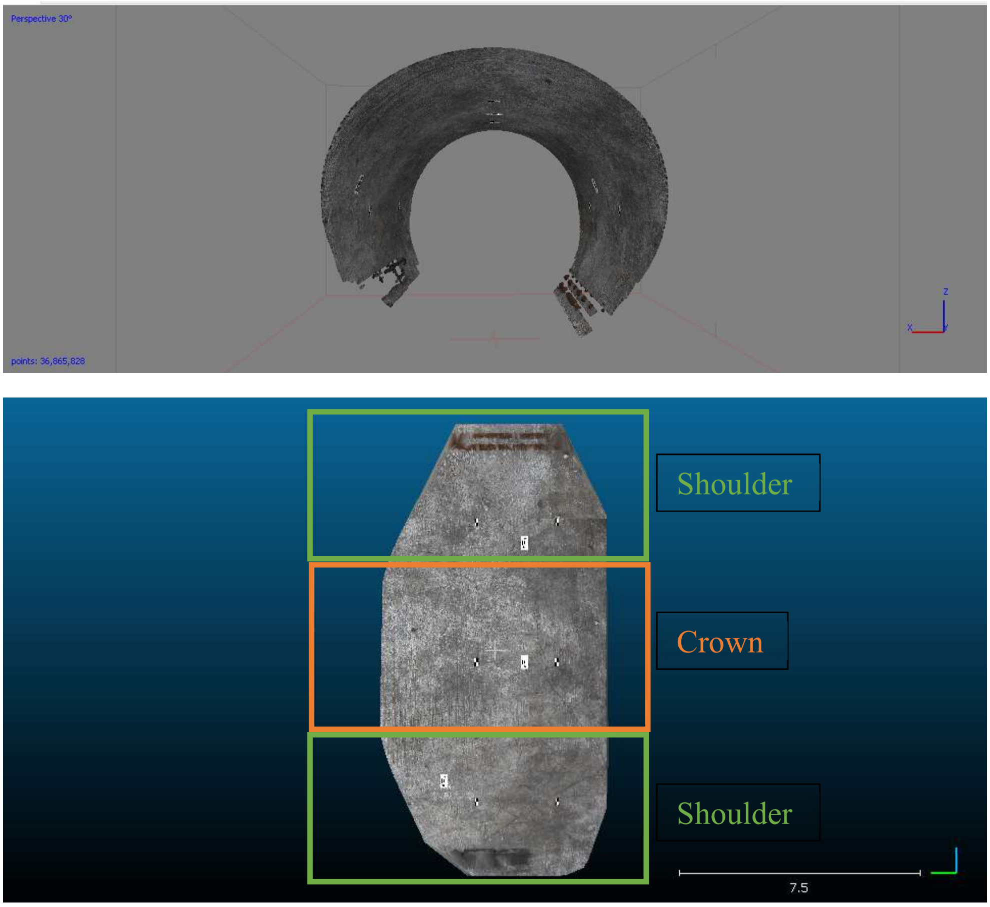


Figure 4: (Top): 3D reconstruction of the virtual tunnel front view; (Bottom) unwrapped texture of the tunnel lining reconstruction

Semantic segmentation represents the last evolution in the progression from coarse to fine inference, providing the user with a per pixel prediction classification. Its architecture generally consists of two networks: the encoder (pre-trained classification network) followed by the decoder. The decoder performs the dense pixel classification projecting the discriminative features learnt by the encoder to the pixel space.

There are several mechanisms to perform the projection of the pixel inference onto the pixel space. In this paper the author will focus on the Fully Convolutional Networks (FCN) method. FCNs, as their name implies, consist of only convolutional and pooling layers. This provides a global characterisation of the image, associating each pixel with a label. In that way, it is possible to overcome the restriction of the CNNs to accept and produce labels only for specific sized inputs having the ability to make predictions on arbitrary-sized inputs. This approach was originally proposed by (Long, et al., 2015).

To replace a fully-connected layer with an equivalent convolutional layer, the size of the filters is set to the size of the input to the layers, and as many filters as there are neurons are used in the fully-connected layer (Heinrich, 2016). On the other hand, the resolution of the inference of the feature map is limited (see Figure 5) due to the alternation of the convolutional and pooling layers.

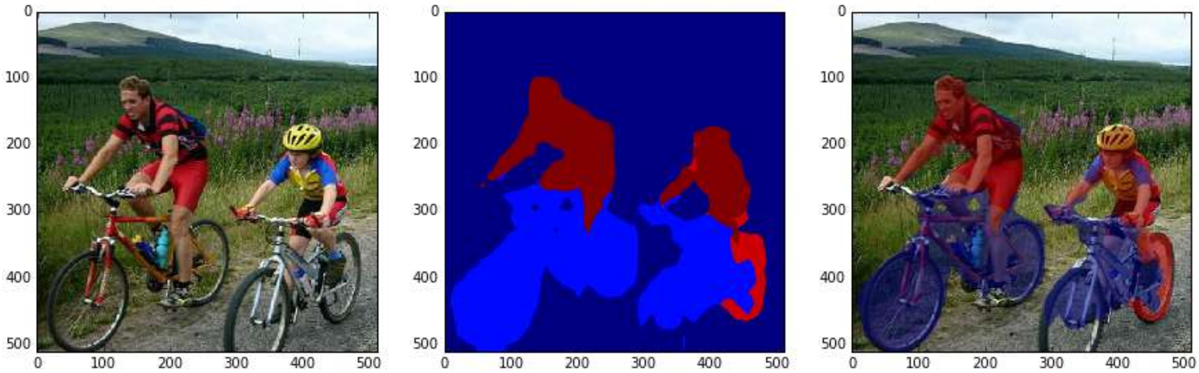


Figure 5: Example of low resolution inference.

To refine the inference resolution of the semantic segmentation and have results closer to the ground truth (particularly important in the case of crack detection), several approaches have been proposed (e.g. (Chen, et al., 2015), (Yu & Koltun, 2016), (Badrinarayanan, et al., 2017)). In the present research, the authors implement the FCN-8 architecture (Figure 6): combining predictions from final layer, pool4 layer, at stride 16 and additional predictions from pool3, at stride 8, provide further precision up to 8x up sampled prediction (Long, et al., 2015).

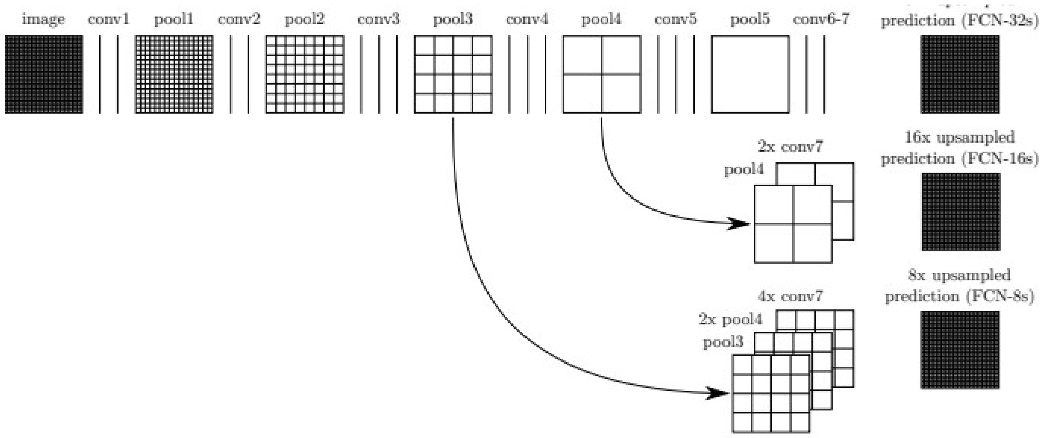


Figure 6: FCN-8 architecture

The training of the semantic segmentation has been performed using on-site pictures from tunnel surveys, as well as publicly available images from the internet. A subsample of the on-site image database was kept for testing purposes.

The performances of the developed ML algorithm are measurable through the following descriptors (Eqs. 1 to 3 (Geron, 2017)):

- Accuracy is a metric for evaluating classification models. Informally, accuracy is the fraction of correct model predictions (Eq. 1).
- Precision is the proportion of positives identified correctly (Eq. 2);
- Recall represents the proportion of actual positives identified (Eq. 3);

$$Accuracy = \frac{TP+TN}{TOT} \quad \text{Eq. 1}$$

$$Precision = \frac{TP}{TP+FP} \quad \text{Eq. 2}$$

$$Recall = \frac{TP}{TP+FN} \quad \text{Eq. 3}$$

Where:

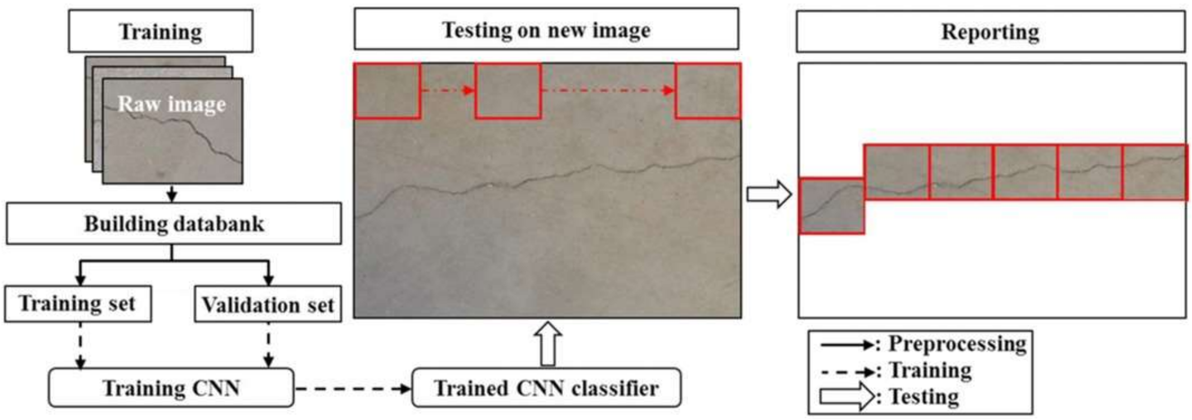
- True Positives (TP) are the pixels part of the crack correctly labelled;
- True Negatives (TN) are the background pixels correctly labelled;
- False Positives (FP) are the background pixels classified as crack;
- False Negative (FN) are the crack pixels classified as background;
- TOT is the total number of analysed pixels.

In the present study case, the performances of the network are promising with an accuracy value over 99%. Anyway, accuracy is evaluated considering both the foreground and the background (see Eq. 1). As such, it is appropriate to consider recall (Eq. 3) as the parameter to evaluate the performances of the network together with the accuracy. The recall value gives information about the proportion of the crack length correctly labelled

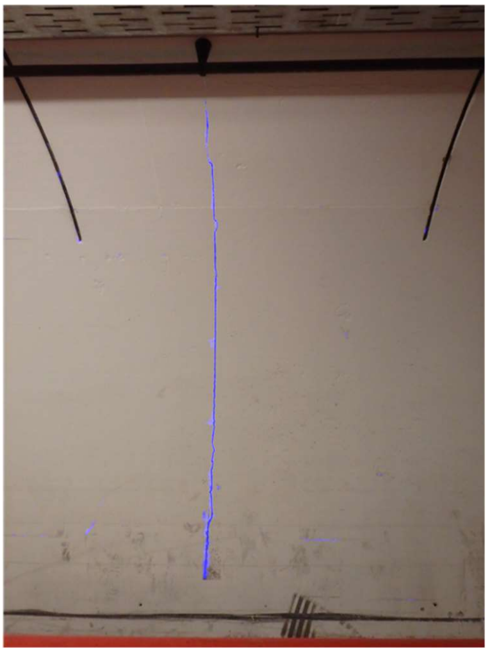
To assess the suitability of the process and to be sure that no overfitting of the data occurred, the algorithm has been validated successfully against on-site surveys from concrete tunnel linings. This is illustrated in Figures 7b and 7c, from a tunnel at CERN in Geneva, and Figure 7d, from a UK Power Network tunnel in London.

The algorithm has also been tested on images outside of a tunnel domain in which the machine has never been trained on, such as cracking in roads (Figure 7e). The test dataset counted a total of 60 images. Table 1 summarises the performance descriptors mentioned before evaluated on the test dataset. The following considerations can be made:

a) Accuracy over 99% achieved, shows the robustness of the model in detecting cracks;
b) Recall ≈ 80% achieved, indicating that a very high proportion of actual cracks have been detected – see Figures 7b) to 7e).
c) Precision ≈ 27%: Note that this metric is expected to be of a low percentage as it is highly influenced by false positives. Cracks are tiny elements in the image space, so an error of few pixels around the edges of the cracks classified as cracks will produce a low value of precision (Figures 7b to 7e)).
d) The results are promising considering that only 90 images were used to train the algorithm and that the image test dataset has been chosen to have the lowest level of similarity with the image training dataset.
e) Overall, this novel approach based on semantic segmentation outperforms the image classification based approach (Panella, et al., 2018) tested on this particular dataset.



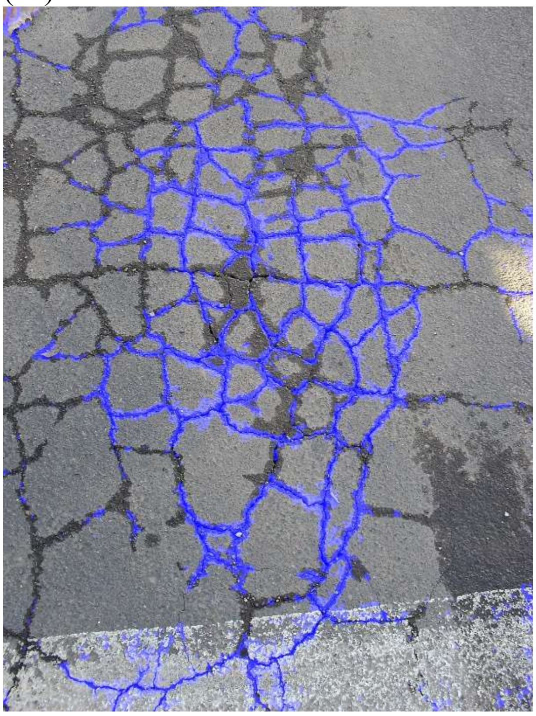
(6a)



(6b)

(6c)

(6d)



(6e)

Figure 7: examples of semantic segmentation on a test dataset:
(7a): (Young-Jin, et al., 2017) pipeline proposal and relative output
(7b, 7c) Concrete lining in a tunnel at CERN, Geneva
(7d) Concrete lining from a UK Power Networks tunnel, London
(7e) Cracking in road asphalt "Creative Commons Asphalt Deterioration" by Bidgee is licensed under CC BY-SA 3.0.

Table 1: comparison of different Algorithms/Networks in terms of accuracy of the prediction

Algorithm / Network	Training set			Test set		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Novel Semantic Segmentation	99.9	-	-	99.6%*	26.1%*	78.1%*
CrackDet [i] *(AlexNet)*	98.0%	92.0%	92.5%	93.9%**	72.4%**	77.8%**
CrackDet [i] *(GoogleNet)*	-	-	-	80.3%**	35.3%**	88.9%**
ELM[ii]	98.5%	-	-	91.6%	-	-
RBF[ii]	96.5%	-	-	90.1%	-	-
SVM[ii]	98.0%	-	-	91.3%	-	-
KNN[ii]	-	-	-	88.7%	-	-

[i] (Panella, et al., 2018)
[ii] (Wenyu, et al., 2014)
* per pixel evaluation
** per crop evaluation

Following classification, the original mask of each crack is kept in order to realise the crack map per image. This also enables a first pass estimation of the crack dimensions in terms of length and width. The weights are saved after model training which are then used to refine and improve on further models and tunnel surveys, thus improving the accuracy over time. An example of a crack density metric output is illustrated in Figure 8. This can be used by the asset manager to improve their understanding of the structures condition, and a decision aid on where to focus further investigations or potential interventions.

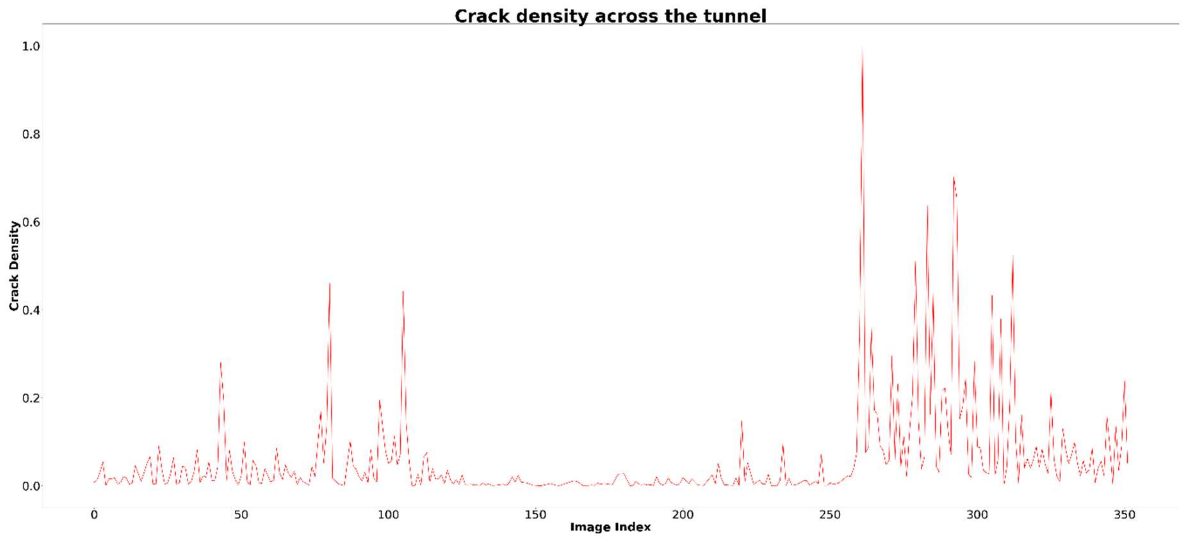


Figure 8: example of a normalised crack density metric across tunnel chainage (i.e. image index).

4. Future research

The automated assessment of tunnel crack defects is possible with a high accuracy. The tests showed a recall of ~80% and an accuracy >99% on test images (images with a low level of similarity with the training images). Future research is focused on improving the scores of the segmentation using improved algorithms and larger datasets, across varied tunnel lining types.

Other aspects to be explored in the future research are the analysis of several cameras and lenses to select the most suitable combination for each situation taking in consideration the lighting conditions on site and calibration of the optimal camera setup based on the specifications of the camera type defined in the previous step. The authors are also undertaking research to investigate alternative faster remote data capture means, such as through the use of drones in tunnels, although the viability of this in underground confined spaces is an early challenge, as compared to its more prevalent use at surface.

These examples demonstrate the efficacy of these digital solutions and the utility to tunnel owners with respect to repeatability, cost, health and safety improvements. These simple yet powerful innovations have provided step change improvements to conventional industry practice, highlighting a path for how future digital developments within tunnelling and civil engineering works can succeed.

5. Acknowledgements

The authors would like to thank the European Centre for Nuclear Research, CERN, and UK Power Networks for authorisation for use of tunnel asset images in this paper.

References

Badrinarayanan, V., Kendall, A. & Cipolla, R., 2017. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Blender, undated. Blender. [Online], Available at: https://www.blender.org/ [Accessed 08 12 2017].

Brian, S. N., 1997. The Consequences of Poor Serviceability and the Way Forward. In: Management of Concrete Structures for Long-Term Serviceability. s.l.:Thomas Telford Publishing, pp. 1-8.

Chen, L. C. et al., 2015. Semantic Image Segmentation With Deep Convolutional Nets and Fully Connectec CRFs. s.l., s.n.

Dapeng, Q., Yun, L., Qingyi, G. & Fengxia, Z., 2014. An Algorithm to Detect The Crack in The Tunnel Based on Image Processing.

Geron, A., 2017. In: Hands-on Machine Learning With Scikit-Learn & Tensorflow. s.l.:O'Reilly Media, Inc. 1005.

Heinrich, G., 2016. Image Segmentation Using Digits 5. [Online] Available at: https://devblogs.nvidia.com/image-segmentation-using-digits-5/ [Accessed 28 05 2018].

Long, J., Shelhamer, E. & Darrel, T., 2015. Fully Convolutional Networks for Semantic Segmentation. s.l., s.n.

Mc Kibbins, L., Elmer, R. & Roberts, K., January 2010. Tunnels: Inspection, Assessment and Maintenance (C671, s.l.: CIRIA.

McDonnell, D. & Devriendt, M., 2017. The Application of BI and Photogrammetric Methods To The Inspection and Management of The Sydney Harbour Tunnel. s.l., s.n.

Panella, F. et al., 2018. Deep Learning And Image Processing for Automated Crack Detection and Defect Measurement In Underground Structures. Riva del Garda, s.n., pp. 829-835.

Pix4D, undated. Pix4D-support. [Online] Available at: https://support.pix4d.com/hc/en-us/articles/203756125-How-to-verify-that-there-is-Enough-Overlap-between-the-Images [Accessed 30 08 2018].

Sinha, S. & Iyer, S., 2005. A Robust Approac for Automatic Detection and Segmentation of Cracks In Underground Pipeline Images.

Sunil, K. & Paul, W., 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, Volume 14.

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, C., Wooram, C. & Oral, B., 2017. Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks. Computer-Aided Civil and Infrastructure Engineering.

Young-Jin, C., Wooram, C. & Oral, B., 2017. Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks. Computer-Aided Civil and Infrastructure Engineering.

Yu, F. & Koltun, V., 2016. Multi-Scale Context Aggregation by Dilated Convolution. s.l., s.n.