Use Of Deep Learning For Automatic Detection Of Cracks In Tunnels

Vittorio Mazzia[1,2,3], Fred Daneshgaran[1] and Marina Mondin[1]

[1]California State University 5151 State University Drive, CA, Los Angeles, 90037
[2]Politecnico di Torino - Dept. of Electronic and Telecommunications Engineering
[3]PIC4SeR - Politecnico Interdepartmental Centre for Service Robotic
[4]SmartData@PoliTo - Big Data and Data Science Laboratory
Turin - Italy

Abstract. Tunnel cracks on concrete surfaces are one of the earliest indicators of degradation, and if not promptly treated, they could result in full closure of an entire infrastructure or even worse in a structural failure of it. Visual inspection, carried out by trained operators, is still the most commonly used technique and, according to the literature, automatic assessment systems are arguably expensive and still rely on old image processing techniques, precluding the possibility to afford a large quantity of them for a high frequency monitoring. To overcome some of these difficulties, this article proposes a low cost, automatic detection system, that exploits deep architecture of convolutional neural networks (CNNs) for identifying cracks in tunnels relying only on low resolution images. The trained model is obtained with two different methods: a custom CNN trained from scratch and a retrained 48-layer network, using supervised learning and transfer learning respectively. Both architectures have been trained and tested with an image database acquired with a first prototype of the video acquisition system.

Keywords: Deep Learning, Tunnels, Automation, Cracks

1 Introduction

The increased urbanization and population density in major urban centers of the world has led to a greater demand for new underground transportation infrastructures. Indeed, a well-organized network of tunnels in a densely built-up urban area is able to solve problems such as traffic congestion, noise and air pollution. Nevertheless, design, construction and maintenance of underground infrastructure still rely on empirical methodologies that most of the time result in a cost increase and extended handling times. In this context, where automatic assessments could be an important aid for maintenance of tunnels, the presented article proposes an affordable and reliable low cost, cracks detection system as replacement to traditional visual inspection techniques. Underground infrastructure are characterized by poor lighting conditions and filthy surfaces covered by dust and mold. Moreover, most of the time huge scuffs on the walls, due to the

aging of the lining, could easily resemble the shape of a crack. Due to these main reasons, automated inspection of tunnels is still an open challenge with not so many research studies and few enterprises that offer solutions. In the last few years, Deep Learning has demonstrated promising results in the image detection and recognition tasks. This intrinsic capability has been taken into evidence by the annual ImageNet Large Scale Visual Recognition Challenge that has been showing the potentiality of this technique since 2012 (1). For that reason, a convolutional neural network has been used as key element to obtain an automatic system able to rely only on low resolution, consumer grade equipment. A number of image processing techniques (IPTs) have been proposed for identifying civil infrastructure defects, but they all require high-resolution images, drastically increasing the cost of the overall system. Indeed, only a low cost device that allows for a high frequency monitoring could really be a valuable substitution for the actual manual assessment procedure. An image database has been obtained using a first prototype of the video acquisition system developed by our team and it has been exploited to train the two different solutions of the model that, considering the highly challenging task, achieved scores of test accuracy above 90%. Finally, artificially expanding the training data has been exploited in order to increase the dimension of the available dataset (2). Rectified linear units as neurons of the network, GPUs in order to speed up the simulation and Dropout technique in order to prevent the overfitting problem have been used extensively (3).

2 Related Works

As a matter of fact, as previously introduced, visual inspection of underground infrastructure is by far the most used technique for crack detection and tunnels maintenance. Several studies and solutions have been presented to automate and simplify this process but currently, these systems are known to be expensive and in many cases not reliable.

2.1 Cracks detection using image processing

Before 2010, Japan and South Korea were at the leading of research for crack detection in tunnels. Researchers in Japan proposed an automatic monitoring system to detect cracks in tunnels based on a mobile robot (4). They used edge enhancement and graph searching technology to extract the cracks on images. More recently, due to the high expenses which implies visual inspections, engineers were driven to develop more complete and reliable solutions. This is true especially for all those countries with a spread network of underground infrastructures. That is the case of China where for example the network of subway in Beijing has been rapidly developed and should be maintained. At present, urban rail transportation in China is still mainly dominated by manual checking, missing in efficiency and security. In this context, the Beijing Municipal Commission of Education Beijing Jiaotong University developed an algorithm

to detect cracks in tunnels based on image processing (5). The acquisition system is based on CCD cameras with high-frequency data collection. A Laser is used as light source and a custom IPT is used as core of the module. Toshiba Research Europe in 2014 tried to investigate a low cost system using a technique known as Structure from Motion (SfM) in order to recover a 3D version of the tunnel surface and find defects with comparisons (6). A different research project has been developed by Pavemetrics Systems Inc. making use of top grade and expensive equipment to acquire both 2D images, and high resolution 3D profiles, of infrastructure surfaces at speeds up to 30 Km/h (tests have been carried out at 20 Km/h) (7). More recently, in 2016 a research has been carried out in order to automatically collect and organize images taken from tunnels carrying high-voltage cables (8). The project is developed by making use of consumer DSLR cameras and high-power polarized LEDs, in order to improve image quality, mounted on a lightweight aluminium frame, which is studied to dampen vibrations during data capture.

2.2 Cracks detection using Machine Learning

As first attempt in 2002, during the golden age of Support Vector Machines, four researches proposed an efficient tunnel crack detection and recognition method (9). No further recorded attempts have been made using machine learning algorithms until 2016, when for the first time a deep learning model has been used within the detection algorithm (10). The system exploits the algorithm to devise an automatic robotic inspector for tunnel assessment. The robotic platform consists of an autonomous mobile vehicle with all sensors attached to a crane arm. Two sets of stereo-cameras are used for taking the necessary images and an Arduino Uno board is used as a pulse generator synchronizing the two cameras. The first stereo pair is responsible for the crack and the other defects detection that lay on the tunnel lining. The second stereo pair is used for the full 3D reconstruction of high fidelity models of the areas of cracks. A FARO 3D laser scanner is deployed when a crack is detected for a precise calculation of any tunnel deformation that could be present.

3 Overview of the Devised System

The basic architecture of the system is primarily composed of four core blocks. The first one is the video acquisition block. It is composed of three consumer grade camera with their related LEDs array as lighting sources. This composition ensures a coverage of about 180 of the tunnel lining with the minimum equipment cost. All the devices are mounted on a steel framework supported by a vibration isolation pad that increases the performances of the image stabilization system of the cameras. Finally, a series of sensors like accelerometer, encoder, help in tracking, inside a tunnel, a possible detected crack. The second block is part of the software and divides the input videos into their frames, which feed the model block. This is a pre-trained CNN that gives a prediction through the output

layer. It is not needed in a real time evaluation and so, the computational power required is very low. Finally, the resulting classification and all data coming from the sensors are stored and organized by the last block of the system.

4 Video Acquisition Block

The low equipment cost, which is more than an order of magnitude cheaper than the industrial one, the great portability and modularity are the key points of this module of the devised system. It has a huge impact on the price and its quality greatly affects the accuracy of the model block. As a matter of fact, the capability of the neural network to identify cracks of small dimension is highly determined by the quality of the input images and of the dataset for the training session. The camera to be chosen, as well as the light source and the mobile base, should be carefully weighted trying to have a compromise between quality and price. Naturally, the presented prototype, in the following sub-section, is one of several possible options and only further field experimentation can really determine what is the best one.

4.1 Acquisition system's first prototype

Two major reasons has led to the choice of the design of the prototype-1. First, it had to be easy to install on a vehicle, remotely controlled within it and not annoying for other drivers. Secondly, it has been used to test and compare the acquisition system with an unusual source of light. The prototype-1 was made of four components: a consumer-grade camcorder, an infrared light torch, a remotely controlled pan tilt and a bicycle rack as base. Infrared was selected in order to try a different spectrum of light, looking for advantages and disadvantages of it. The chosen infrared flashlight was the Evolva Future Technology T20 with a 38mm lens, emitting infrared light at 850nm and powered by one 18650 Li-Lon rechargeable battery (3.7-4.2V). Due to the chosen light source, a camera with a CCD sensor has been chosen. CCD sensors are very sensible to infrared light but with the drawback to present a lower frame rate than CMOS sensors. Consumer-grade cameras all have a hot filter for infrared light, which must be removed to detect light in this spectral region. It is not simple to manually remove this filter but some cameras have a night vision mode that simply remove it mechanically turning the RGB image into a gray-scale one. The total cost of the prototype, shown in Fig. 1, was less than $200. In order to have an order of magnitude, a leading manufacturer of industrial cameras, Imaging Source, has been contacted for a quotation. The cheapest camera without lens was around $850, four times more expensive than the entire prototype.

4.2 Dataset

The prototype, shown in Fig. 1, has been mounted on the roof of a minivan and carried through the major tunnels of Los Angeles. With the first data acquisition attempt, it has been possible to achieve a collection of videos lasting

Fig. 1. Prototype-1 mounted on the top of the vehicle.

approximately 40 minutes (total time: 40.09, total number of frames: 2405.4 x 30 = 72162 frames). After an accurate selection, it has been possible to obtain a dataset of 6494 images.

- Crack detection folder: 4094
- No-Cracks detection folder: 2400
- Total dataset images: 6494

Finally, using a the already introduced technique, artificially expanding the training data, through simple image modifications like cropping, rotating and scaling it has been possible to drastically increase the number of the available images. An example of the potentiality of this technique is presented in Fig. 2.

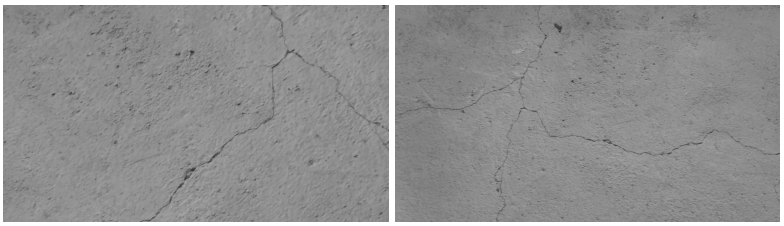

Fig. 2. Practical example of artificially expanding the training on a crack image of dataset. The picture on the left is the one on the right randomly rotated of 209° counterclockwise.

5 The Model

The model algorithm is the third block of the system and differs with almost all related works present in literature. It exploits recent achievements in computer

vision, brought by the application of Deep Learning. Indeed, this groundbreaking technique, all of a sudden, has exceeded all old IPTs and traditional machine learning algorithms, allowing, for this project, to use only images taken by low cost consumer-grade equipment. This has been made possible by the capabilities of convolutional neural networks to divide the problem in simpler ones and to identify increasingly more complex patterns from the input images. After a model is trained on the desired features, it is simple to embed it in an actual working system. It will analyze input images coming from the pre-processing block giving a probability prediction for each class set. In the next sections, we present two methodologies with related results that have been followed to produce models trained on the specific task of crack detection. The results of the two approaches are proposed with different values of hyper-parameters such as learning rate, η, number of training epochs, regularization parameter, λ ,and mini-batch size.

5.1 Re-trained CNN

Transfer Learning is becoming a very popular topic in machine learning community. Moreover, experimental evidence with Deep Learning techniques have demonstrated the possibility to successfully re-purpose an already trained deep convolutional network with novel generic tasks (11). Indeed, all convolution and pooling layers extract increasingly abstract features that can be used to classify different types of objects. Instead, the fully connected layers and the classifier need to be re-trained on the new task, using supervised learning with the proper image database. In light of this, as first step of the model generation, we have taken a state of the art CNN and retrained it to detect the presence of cracks. In this way, it has been possible to train a huge model in less than one hour (a model that usually required 2-3 weeks to be fully trained). The selected trained model is Inception-v4 model, a slimmed down version of the relative Inception-v3 model (12). Unlike the other, this novel architecture has been designed specifically for TensorFlow library. It has shown very good performance at relatively low computational cost and in order to achieve these performances, it exploits

Table 1. Transfer learning results applying different hyper-parameters.

Mark	**Iterations**	**Learning rate**	**Train b. size**	**Cross entropy**	**Test accurancy**
1	4000	0.01	100	0.143	95%
2	9000	0.01	100	0.109	97.5%
3	9000	0.001	100	0.132	96.4%
4	15000	0.001	100	0.147	95.1%
5	15000	0.01	100	0.103	97.6%
6	10000	0.01	100	N.A.	N.A.
7	9000	0.01	300	0.102	97.6%
8	9000	0.01	1000	0.106	97.5%
9	9000	0.01	1000	0.094	98.1%

inception blocks, batch normalization and residual connections in its sibling version Inception-ResNet-v2 (13). The model has been retrained with input images of 299x299 pixels using different settings and hyper-parameters. Final results are presented in (Table 1). The last training session has been performed using the artificially expanded dataset. The choice of hyper-parameters has been made taking into account common heuristic rules and the final accuracy has been computed with the test set, only presented at the last epoch. The test set was tuned at 10% of the dataset and containing 649 different pictures. The ninth test, that achieved 98.1% of test accuracy, classified in the correct class 637 images with only 12 misplaced. This result demonstrates the huge potential of transfer learning that already with the first simulations has proven to be able to reach high values of accurac

5.2 Custom CNN

A custom convolutional neural network, despite the lack of a large dataset, has been trained using supervised learning. Unlike Inception-v4, a CNN trained from scratch on a small number of classes is more suitable for the specific assigned task. Indeed, all weights and biases are specifically calibrated to recognize features of the selected classes. The definition of the network architecture does not have a closed solution. Many different approaches can be applied and usually, only the actual simulation can discern what is the most suitable framework. For this project, all decisions have been made trying to keep the number of parameters low. A high number of parameters not only increases the training process time but also makes the network more inclined to overfit input data. Fig. 3 gives a detailed overview of the designed architecture. It has two similar stages of convolution and pooling and a final soft-max classifier that has as output a certain prediction over two different classes. Going deeper with the network the number of parameters of a single feature decreases and after the second pooling the output matrix takes the shape of an array in order to be suitable for the fully connected layer. The input layer has output with dimension 128x128 pixels for each color channel. Images are in gray scale but the color model is RGB. First

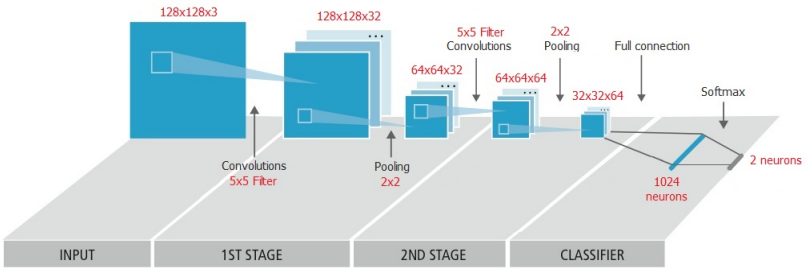

Fig. 3. Architecture layout of the convolutional neural network.

Table 2. Summary table of the network specifications.

Section	**Input**	**Output**	**Patch**	**Strike**	**Filters**
Input layer	1920x1080x3	128x128	N.A.	N.A.	3
1^{th} CL	128x128x3	128x128	5x5	1	32
1^{th} PL	128x128x32	64x64	2x2	1	32
2^{nd} CL	64x64x32	64x64	5x5	1	64
2^{nd} PL	64x64x64	32x32	2x2	1	64
FCL	64x64x32	1024	N.A.	N.A.	N.A.
Sofmax layer	1024	2	N.A.	N.A.	N.A.

convolutional layer (CL), like the second one, has a 5x5 filter with zero padding in order to analyze presence of cracks in the entire picture. Then, if a crack is detected, it is not important anymore its position within the image so that a max pooling (PL) is exploited to decrease the number of parameters. Finally, fully connected layers (FCL) analyze the extracted features and the softmax output layer generates a prediction. All neurons, except the ones of the last layer (sigmoid neurons), are rectified linear units (ReLu). This type of neurons have proved to be easier and faster to train (14). (Table 2) presents an overlook of the designed convolutional neural network. This is only one possible architecture. Many other frameworks can be devised and, as it is for hyper-parameters, only empirical and not deterministic rules are available for the designing process. In order to compare the results achieved with this model with the ones reached with Inception-v4 model, we selected an equal size of the test set (10%) and the same percentage of validation images (20%). In every simulation, the validation occurs every 150 steps. Unfortunately, it was not possible to maintain the same input dimension of the pictures due to a lack of the HW setting. Surprisingly, with only input images of 128px per side, it was possible to achieve 93.1% of test accuracy. So, it is highly probable that greater dimension of the input images, carrying more information, could greatly help the network learning more robust and useful features with a resulting higher accuracy. Again, the simulations have been carried out with different combination of hyper-parameters and only the ones of the last attempt are reported in (Table 3). The last simulation has been performed with 10000 steps and Dropout, L2 regularization and artificially expanding the training data in order to tackle the overfitting problem. Finally, the model has been trained with GPUs in order to speed up the entire process.

Table 3. Summary of hyper-parameters selected for the last simulation.

Image	**Hyper-Parameters**				
Size	***Train b.***	***Validation b.***	***Dropout***	***L. rate***	***L2 Reg.***
128 px	150	200	0.5	1e-5	10

5.3 Comparison between the two models

Both networks are easily implementable. After the training session it is possible to store the model in a binary format file. This file can be loaded and used in the software of the device in order to make predictions about new input images. Further works and improvements are needed, but the early results have been extremely promising. As expected, transfer learning, which needs less data for its learning process, has overtaken the custom neural network in terms of learning time and test accuracy. Considering the huge number of parameters that had to be trained, the result of 93.1% achieved by the custom CNN has to be considered as great. Little adjustments of the network, but especially a larger dataset and greater dimension of input images should increase the accuracy level around the value achieved by Inception-v4. Indeed, the graph presented in Fig. 4 shows that, due to the dimension of the image database, the custom deep neural network, after 7000 iterations, starts to overfit the data.

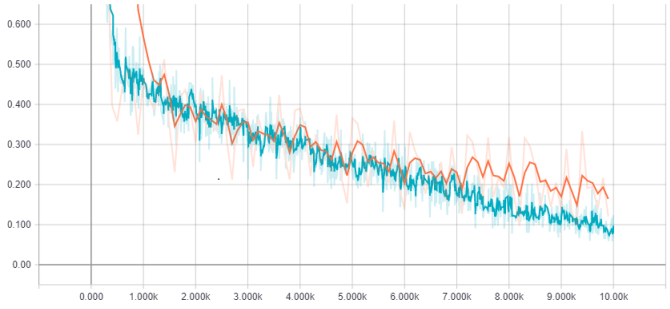

Fig. 4. Cross Entropy functions. Blue line is the training function and red line is the validation function, smoothing=0.4.

5.4 Comparison

In conclusion, both methodologies have successfully proven that convolutional neural networks, with a small effort, can overcome most of the traditionally complex and accurately calibrated image processing techniques. Moreover, libraries like TensorFlow have drastically simplified network design and implementation and, their tools and features have gradually opened the possibility to train deeper and more precise neural networks.

6 Conclusion

A low cost, easily enforceable automatic cracks detection system, based on recent developments in image recognition and computer vision, has been proposed as

a replacement to the commonly used manual inspection. Results have pointed out the suitability of deep learning architectures for the tunnel defect inspection problem relying only on low range, consumer grade equipment. Further works and researches may automate on a large scale the current tedious and unreliable work of underground infrastructures assessment.

References

1. DENG, Jia, et al. Imagenet: A large-scale hierarchical image database. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009. p. 248-255.
2. SIMARD, Patrice Y., et al. Best practices for convolutional neural networks applied to visual document analysis. In: ICDAR. 2003. p. 958-962.
3. SRIVASTAVA, Nitish, et al. Dropout: a simple way to prevent neural networks from overfitting. Journal of machine learning research, 2014, 15.1: 1929-1958.
4. T. Yamaguchi, S. Nakamura, S. Hashimoto, "An efficient crack detection method using percolation-based image processing," in Proceedings of 3rd IEEE Conference on Industrial Electronics and Applications (ICIEA 2008), pp. 1875- 1880, 2008.
5. Qi, D., Liu, Y., Wu, X., & Zhang, Z. (2014, August) - An algorithm to detect the crack in the tunnel based on the image processing. Intelligent In- formation Hiding and Multimedia Signal Processing (IIH-MSP), 2014 Tenth International Conference on (pp. 860-863).IEEE.
6. Stent, S., Gherardi, R., Stenger, B., Soga, K., & Cipolla, R. (2014) - Visual change detection on tunnel linings. Machine Vision and Applications, 27(3),319-330.
7. Fox-Ivey, R., Dominguez, F. S., & Garcia, J. A. R. (2015) - Use of 3D Scanning Technology for Automated Inspection of Tunnels.
8. Liu, Z., Shahrel, A., Ohashi, T., & Toshiaki, E. (2002, March) - Tunnel crack detection and classification system based on image processing. Proc. SPIE (Vol. 4664, pp. 145-152).
9. STENT, S. A. I., et al. A Low-Cost Robotic System for the Efficient Visual Inspection of Tunnels. In: ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction. Vilnius Gediminas Technical University, Department of Construction Economics & Property, 2015. p. 1.
10. Protopapadakis, E., Makantasis, K., Kopsiaftis, G., Doulamis, N., & Amditis, A. (2016) - Crack Identification Via User Feedback, Convolutional Neural Networks and Laser Scanners for Tunnel Infrastructures. VISIGRAPP (4:VISAPP) (pp 725-734).
11. DONAHUE, Jeff, et al. Decaf: A deep convolutional activation feature for generic visual recognition. In: International conference on machine learning. 2014. p. 647-655.
12. SZEGEDY, Christian, et al. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016. p. 2818-2826.
13. SZEGEDY, Christian, et al. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. In: AAAI. 2017. p. 4278-4284.
14. NAIR, Vinod; HINTON, Geoffrey E. Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th international conference on machine learning (ICML-10). 2010. p. 807-814.