

# Enhancing Machine Learning Pipelines on Industrial Applications

**Carlos Beltrán-González, Milind G. Padalkar, Alessio Del Bue**

Pattern Analysis and Computer Vision (PAVIS) - Istituto Italiano di Tecnologia  
{carlos.beltran, milind.padalkar, alessio.delbue}@iit.it

## Abstract

Frequently, the problem faced by machine learning projects in industrial applications is that of generating sufficient quality annotated data to feed the learning phase. This problem becomes critical when the data needs to be generated by highly specialized (and costly) human experts. In this contribution, we show one way of enhancing the human annotation performance by using a hybrid image fusion technique.

## 1 Introduction

Periodic inspection and maintenance are necessary in industrial components to ensure both quality and safety. The automation of such processes reduces costs and improves quality assurance. With the recent advancement of machine learning techniques (in particular, deep learning), there has been a steady progress in the use of image processing and computer vision based methods for automatic visual inspection. For example, use of photometric stereo to identify defects in casted steel [Landstrom *et al.*, 2013], convolutional neural networks for inspection of metallic components in nuclear power plants [Chen e Jahanshahi, 2018], examining flaws in concrete structures [Yang *et al.*, 2019], inspection with multiple lights [Aghaei *et al.*, 2020][Padalkar *et al.*, 2020], have been used to automate visual inspection. Similarly, defect detection in aeronautic components is discussed in [Martelli *et al.*, 2018][Biagio *et al.*, 2017],[Beltrán-González *et al.*, 2020].

However, building complete automatic systems remain a challenge in non deterministic scenarios. In such cases, men-in-the-loop solutions are frequently the only alternative. This is the case when data driven approaches (e.g. machine learning) are used as a technological solution. Human opinion remains necessary, particularly in the learning phase of the system in which massive training data is necessary to achieve relevant performance. A severe problem emerges when such humans are not common people but highly trained experts with profound expertise in their field. In such cases, the time and cost for annotating sufficient data becomes unaffordable for many industries.

## 2 Inspection enhancement

One of the key factors to improve the contribution of highly specialized humans for machine machine learning solutions comes from the enhanced digitization of the inspection process. For example, in one of our solutions, the human expert is presented with digital images that combine illumination and multisensorial fusion techniques to provide an enhanced inspection experience. In particular, we have transformed an industrial analog inspection process into a method that fuses several mutually registered images of ceramic tiles, into a single representative image, highlighting cracks to help the visual inspection process. The images are acquired with different illumination conditions using a customized illumination setup, to improve the visibility of cracks for remote inspection and for the creation of annotated datasets.

Our fusion method is based on training image generators with cycle-consistent losses motivated by [Zhu *et al.*, 2017], that allow transformation from one domain (acquired images) to another (fused) and back. Cracks are enhanced by constraining the image generators with loss networks that produce binary crack representations of the inputs. An example of fusion with crack enhancement using the proposed method is shown in Fig. 1. Our main contribution is a method to combine and enhance crack details into a single representative image from an image sequence acquired using different illuminations.

## 3 Image fusion with cycle-consistent losses

The pipeline of our proposed approach is shown in Fig. 2. First, we acquire mutually registered tile images in several different illumination conditions using a customized setup. We then obtain the binary crack annotations with the help of experts who carefully inspect the tile as well as the acquired image sequence. From the acquired sequence, we generate an initial estimate of the fused image with a modified version of exposure fusion [Mertens *et al.*, 2007] using morphological preprocessing and additional weight-criteria to improve the contrast and brightness. Co-located patches extracted from the image sequence, initial fusion estimate and binary crack annotations are then used for training our models.

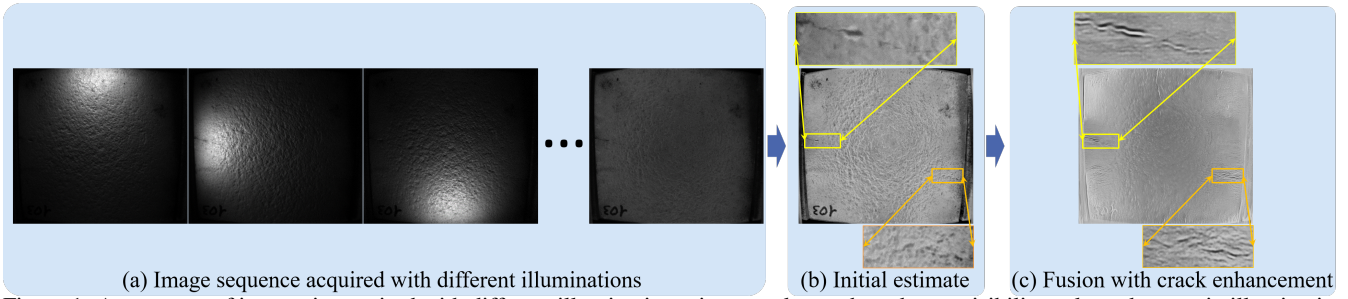


Figure 1: A sequence of images is acquired with different illuminations since cracks can have better visibility only under certain illumination directions depending on their location and orientation. Manual inspection of details in several images is exhausting and one can easily miss seeing a crack that is visible in one or more of those images. Our method generates one representative image (c) having better visibility of cracks, by fusing and highlighting the crack details present in the image sequence (a).

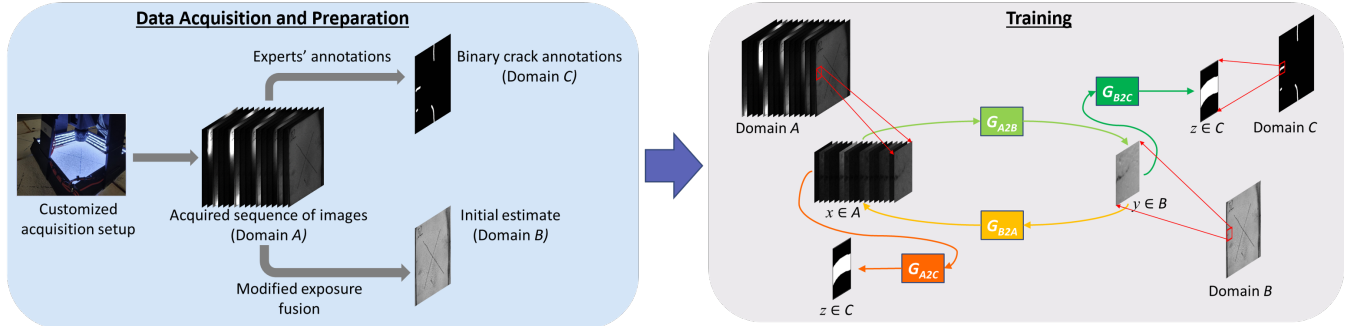


Figure 2: Proposed approach pipeline: An acquired sequence of images is annotated by experts to generate ground truth binary crack annotations. An initial estimate for fusion is obtained from the acquired sequence using a modified version of exposure fusion [Mertens *et al.*, 2007] to improve the contrast and brightness. Co-located patches from these images are then used for training image generators  $G_{A2B}$ ,  $G_{B2A}$  in a cycle consistent manner along with the crack generators  $G_{A2C}$ ,  $G_{B2C}$ .

## 4 Results

Our experiments are conducted on a real-world industrial dataset of 88 ceramic tiles, which consists of image sequences acquired using our customized multi-illumination setup [Padalkar *et al.*, 2020]. Each sequence has 65 different illumination images of size  $1944 \times 2592$ . Patches of size  $128 \times 128$  from 79 tiles are used for training, while 9 tiles are used for testing. Models are trained from scratch for 2 epochs on a single NVIDIA RTX-2080 GPU, with batches size of 8. Adam optimizer is used with learning rates 0.0001 and 0.00001 for training the image generators and crack generators, respectively.

An interesting aspect of this approach involves determining how to evaluate the results of the algorithms. Frequently, in order to facilitate the introduction of the technology in traditional industries, the human experts need to be involved to provide a subjective opinion of the quality of the results. This can be done by conducting cross-annotator studies by means of questionnaires or interviews. Additionally, objective metrics can be used to complement the experts assessments and compare different algorithm results. In the particular case of this work, we evaluate the performance using a measure that captures saliency in terms of edge strength. For an image  $I$  with salient regions  $\Omega \in I$  having stronger edges, the edge strength in  $\Omega$  should be higher than the global edge strength to make  $\Omega$  easily noticeable. Using Laplacian of Gaussian

Tabella 1: Performance comparison for test images with edge strength measure defined in Eq. (1) (The higher the better).

Image#	MEF	Initial estimate	Proposed ( $G_{A2B}$ )
1	1.2369	1.2354	<b>2.6695</b>
2	1.0719	1.1380	<b>3.1600</b>
3	1.0360	1.1272	<b>1.9077</b>
4	1.0825	1.1279	<b>1.7663</b>
5	1.0844	1.1723	<b>2.4617</b>
6	1.1637	1.0021	<b>2.3807</b>
7	0.9425	0.9081	<b>1.9220</b>
8	1.0956	0.9017	<b>2.4140</b>
9	1.0581	1.2385	<b>2.6135</b>

( $LoG$ ), the edge strength  $ES$  can be calculated as:

$$ES = \frac{mean(|L_p|)}{mean(|L_q|)}, \quad p \in \Omega, \quad q \in I, \quad L = LoG(I), \quad (1)$$

where  $L_p$  is value of  $L$  at pixel  $p \in \Omega$  and  $L_q$  is value of  $L$  at pixel  $q \in I$ . Table 1 presents the edge strengths for test images considering the crack annotations ground-truth as  $\Omega$ .

## 5 Challenges and Perspectives

An interesting challenge worth exploring comes from the introduction of multi-sensory imaging in the above described pipeline. In particular, we have been working with thermal images in the range of long-wave infrared (LWIR) using uncooled VOx microbolometer sensors. Although these sensors

become inefficient when the target object matches the ambient temperature, we have effectively used such sensors to enhance defects detection in objects under cooling processes. Although LWIR frequencies are invisible to the human eye the images collected by these cameras can be shown to human operators by applying color maps to present the temperature information. Such thermal images merged with optical images can be a powerful tool for the next generation of digital inspectors.

## 6 Publications

The work presented in this contribution can be further studied in our two works [Padalkar *et al.*, 2020] and [Padalkar *et al.*, 2021].

## Riferimenti bibliografici

- [Aghaei *et al.*, 2020] M. Aghaei, M. Bustreo, P. Morerio, N. Carissimi, A. Del Bue, e V. Murino. Complex-object visual inspection: Empirical studies on a multiple lighting solution. In *25th International Conference on Pattern Recognition (ICPR)*, 2020.
- [Beltrán-González *et al.*, 2020] C. Beltrán-González, M. Bustreo, e A. Del Bue. External and internal quality inspection of aerospace components. In *2020 IEEE 7th International Workshop on Metrology for AeroSpace (MetroAeroSpace)*, pages 351–355, 2020.
- [Biagio *et al.*, 2017] M. S. Biagio, C. Beltrán-González, S. Giunta, A. Del Bue, e V. Murino. Automatic inspection of aeronautic components. *Machine Vision and Applications*, 28(5):591–605, Aug 2017.
- [Chen e Jahanshahi, 2018] F. Chen e M. R. Jahanshahi. Nbcnn: Deep learning-based crack detection using convolutional neural network and naïve bayes data fusion. *IEEE Trans. on Industrial Electronics*, 65(5):4392–4400, May 2018.
- [Landstrom *et al.*, 2013] A. Landstrom, M. J. Thurley, e H. Jonsson. Sub-millimeter crack detection in casted steel using color photometric stereo. In *2013 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, pages 1–7, Nov 2013.
- [Martelli *et al.*, 2018] S. Martelli, L. Mazzei, C. Canali, P. Guardiani, S. Giunta, A. Ghiazza, I. Mondino, F. Cannella, V. Murino, e A. Del Bue. Deep endoscope: Intelligent duct inspection for the avionic industry. *IEEE Transactions on Industrial Informatics*, 14(4):1701–1711, 2018.
- [Mertens *et al.*, 2007] T. Mertens, J. Kautz, e F. Van Reeth. Exposure fusion. In *15th Pacific Conference on Computer Graphics and Applications (PG'07)*, pages 382–390, 2007.
- [Padalkar *et al.*, 2020] M. G. Padalkar, C. Beltrán-González, M. Bustreo, A. Del Bue, e V. Murino. A versatile crack inspection portable system based on classifier ensemble and controlled illumination. In *25th International Conference on Pattern Recognition (ICPR)*, 2020.
- [Padalkar *et al.*, 2021] Milind G. Padalkar, Carlos Beltrán-González, e Alessio Del Bue. Multi-illumination fusion with crack enhancement using cycle-consistent losses. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 2898–2902, 2021.
- [Yang *et al.*, 2019] L. Yang, B. Li, G. Yang, Y. Chang, Z. Liu, B. Jiang, e J. Xiaol. Deep neural network based visual inspection with 3d metric measurement of concrete defects using wall-climbing robot. In *2019 IEEE IROS*, pages 2849–2854, Nov 2019.
- [Zhu *et al.*, 2017] J. Zhu, T. Park, P. Isola, e A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017.