An effective approach to defect detection

Antonio Terpin
Electronic Engineering, Scuola Superiore
Università degli Studi di Udine
Udine, Italia
Email: terpin.antonio@spes.uniud.it

Claudio Verardo
Electronic Engineering, Scuola Superiore
Università degli Studi di Udine
Udine, Italia
Email: verardo.claudio@spes.uniud.it

***Abstract*—Quality control is a main issue in any industry. The need of assuring a human-comparable evaluation during products quality control has reflected on an intense research aiming to develop an automatic defect detection scheme. In this paper an effective solution to defect detection on steel surfaces from images is presented. Firstly, a preprocessing step aiming to spot plausible imperfections areas is discussed. Secondly, a localization and classification of defects is made. Lastly, a proper pixel-wide segmentation scheme is proposed.**

***Index Terms*—Computer Vision, Image Processing, Defect Detection.**

I. Introduction

A fundamental aim in any industry is to achieve highly efficient and effective quality control processes. Therefore, automatic defect detection systems have become an active research area in the last decades.

There is a lot of previous work on defect detection on steel surfaces, even though many ideas [1–3] are as interesting as they are poorly tested and very dependent on background surface uniformity assumptions. This leads to weak generalized results, when one wants to analyze different motif of steel surfaces. Moreover simple thresholding tecniques poorly perform with more complex defect shapes and different light exposure. There are solution using deep learning architectures [4, 5], although with ridicolously small datasets, which provide an attempt to a robust defect segmentation. Other more refined approaches rely on wavelets to detect abrupt changes in the surface, and in textile industry it is possible to find this kind of previous work [6–9]. The core of this paper is based on wavelet analysis and on deep learning, due to two main reasons.

Firstly, multi-resolution analysis (MRA) based on wavelets have been proven effective in facing localization both in spatial and in frequency domains [10–12]. This because of their mathematical properties, compared to Fourier's transform.

Secondly, deep learning has outperformed in the last years any human-designed classificator, indeed computer vision and image processing are increasing in popularity and they are being used ever more in many fields, from autonomous driving vehicles to retail and retail security. Hence there has been an appreciable improvement in the effectiveness of defect detection based on visual systems. There has been a lot of work since the rise of deep learning applications [13] and three main computer vision tasks have been outlined: classification, object localization and object detection.

Classification face the supervised learning problem of identifying to which of a set of categories a given object belongs to. In computer vision this means assigning one of the available labels to an image. This is the simplest of the three tasks and recognizing the category of the principal object in a picture is the standard application of Convolutional Neural Network (CNN), from handwritten characters [14, 15] and house numbers [16] to traffic signs [15].

The main reason why CNNs have become so popular since LeCun originally introduced them [13, 14, 17, 18] is that they represent a black box from raw pixels to categories labels, therefore they overcome the intrinsic difficulties of designing tailored features extractors. Although properly structured CNNs have enviable classification accuracy, they need a cornucopia of data to avoid overfitting. Therefore, to reduce the required size of hidden layers, some preprocessing is tipically made, although they need a smaller dataset compared to a fully-connected multi-level deep learning architecture, and they are also more likely to be shift and scale invariant [18].

A classification task in defect detection field is accounted when objects, e.g. steel surfaces, need to be binary classified as defective or flawless. When visual systems are considered and pictures are taken to classify a particular object, this would be negligently in practical applications. Indeed, monitoring locally the product concerned would be overly expensive, whereas a single global visual system is patently appetible. Moreover, a local analysis may miss some global features of a particular defect.

Object localization sight to find a given number of items in a given context, predicting both their position and their class. Object detection removes the constraint on the number of items, allowing either zero or any finite number of objects. In computer vision, in particular in 2D images, the position is described by a bounding box.

CNNs have been used along with sliding window and multiscale approaches for object detection [19–21], and there is a lot of work aiming to improve performances and bounding boxes accuracies, both by designing different neural network architectures [19] or by tailoring existing one [17]. Regarding to scale-dependence in the object detection task, a solution is given by either brute-force learning (and CNN oversizing) or image pyramids [19], whereas the bounding boxes accuracies can be optimized by combining different scale sliding windows results, taking into account activation confidence in a particular area of the image and applying thresholding tecniques.

In this paper a further refined system is presented, since the purpose of the defect detection algorithm is not only to globally mark an image as picturing a flawless or defective steel surface, but both to highlight flawed regions inside the image and to label it as belonging to a particular defect class.

Pixel-wide classification is known in literature as image segmentation task, and there are three main families of tecniques: hysteresis thresholding, edge-based and region-based [22]. Thresholding exploits a previously known function from the pixels space and classifies pixels through comparison with some discrete values (thresholds) [23], but it is tipically used within other tecniques rather then alone. Region-based approaches use graph algorithms [24, 25], watersheds analogies [26], . Edge-based tecniques, instead, use an edge detection filter [27–29], along with denoising and thresholding considerations, to solve the boundary detection problem. Remark that although similar, boundary detection aims to describe changes in pixel ownerwhip from one object or surface to another, whereas an edge is an abrupt change which can be a subdomain of a border. There are also more advanced tecniques [30] boundary-related which rely on energy minimization and are embedded on region-based approaches. Indeed, all these tecniques can be mixed both together and with learning algorithms, either unsupervised [22] or supervised [31].

The approach here described merges the more effective and efficient ideas of previously described work, balancing the drawbacks of different tecniques. Since segmentation is needed, an edge-based contour detector is here presented, to reach high speed segmentation. Wavelet are used along with image preprocessing and alpha-shape [32] to identify proposals, i.e. regions of interest for the classificator, which may contain a defective area. To overcome the bias introduced from hand-crafting the edge-detection filter, the hyperparameters of the algorithm are tuned with Bayesian Optimization [33–35]. A multi-column CNN (MC-CNN) [15] is then used to combine the segmentation information with a well-known classificator architecture, exploiting both local information and global information. The proposed architecture is shown to have [[[[state-of-the-art]]]] **HOPEFULLY** performances on a Kaggle competition dataset **come facciamo??? non possiamo veramente pubblicare dati a riguardo.... o si? TODOOOO **.

II. Architecture overview

The defect detection system architecture proposed in this paper is shown in figure 1.

Steel surfaces pictures of 1600×256 pixels are taken at the input of the process. Since they may be taken one with a different light exposure condition then the others, some preprocessing is made to enhance the quality of the image, e.g. histogram equalization or linear scaling. Moreover, the images considered have three equal colors levels, therefore they are converted into gray levels, to save space. This first step is further described in III.

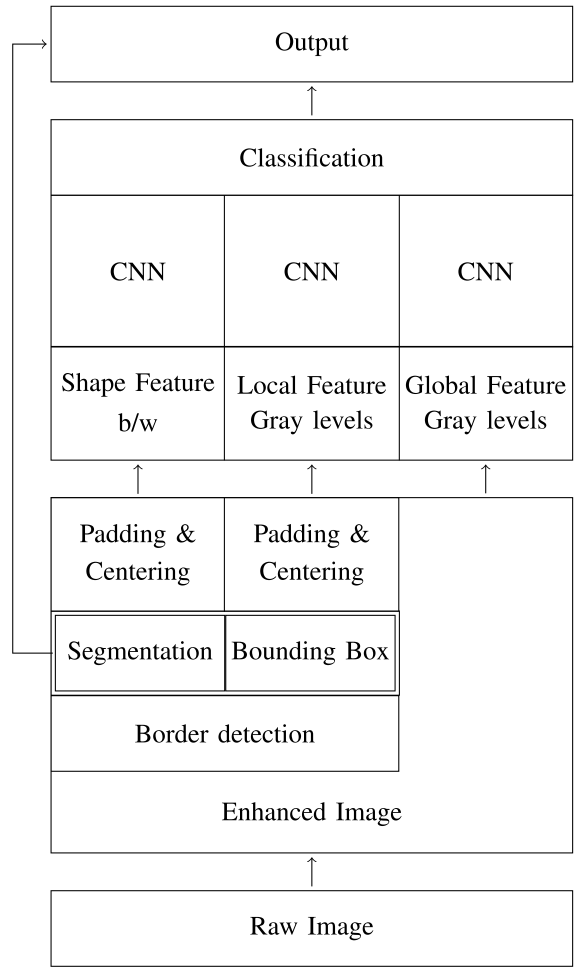

Fig. 1. Proposed defect detection system architecture

The aim of this paper is both detecting pixels representing steel imperfections and classify those regions. Therefore, image segmentation is either obtained as an output of the system or it is needed in some step during the process. To achieve this, a brute-force multi-scale sliding window on the picture could be used, but to improve performances without reducing accuracy a MC-CNN is proposed in V. This MC-CNN allows to combine and consider separately interesting regions, which are called proposals and described in IV, to reduce the number of evaluations. Moreover, both local and global information, are combined to improve classification accuracy. *** IN REALTAá SAREBBE DA VERIFICARE SE SI AVESSE TEMPO, INOLTRE AGGIUNGERE TABELLA DELLE VALUTAZIONI DI DIVERSI SISTEMI, COMPRENSIVA DI TEMPO, ACCURATEZZA, F-MEASURE, ... *** This approach avoid the complexity of combining different scale information and handling windows with different class of defects.

The column of the MC-CNN concerned with global information is fed with the full enchanced image. Conceptually, this CNN learns to evaluate the probability of presence of some kind of defects in the whole surface picture. The other two

TABLE I
TODO RESULTS WITH SOME METRICS

column consider local information instead. This local information is obtained from a further processing step, described in IV. Firstly, a contour detection algorithm (IV-A) is used to spot proposals. Secondly, image segmentation (IV-B) is done, to feed the MC-CNN only with some interesting regions. This segmentation results in a black and white (b/w) map describing plausibles defects. One column of the MC-CNN is fed with this map, therefore it learns to classify regions only observing their shapes. The other column is fed with the portion of original image enveloped in the bounding box (IV-C) of the map, therefore it is trained to consider local features inside, outside and on the border of the considered proposal.

Since defects may have different dimensions, the local information are centered in a 1600×256 pixels black image.

The two MC-CNN columns concerning local information ends with a soft-max layer, because they focus on a small enough region. Instead, the one dealing with global information has a different output layer. Their results are combined in order to properly classify the local regions, and the approach is described in V.

Finally, if the classification outcome labels the region as defective, segmentation coordinates are kept. When all the proposals of the considered image have been processed, defective pixels are encoded with Run-Length-Encoding (RLE) algorithm. The output given is the RLE-encoding of all defective pixels grouped by class. If the surface is flawless, all this encodings are empty.

III. Image preprocessing

In this section image preprocessing is introduced, and the raw image is enhanced to improve learning quality. *** TODO POTREBBE ESSERE INTERESSANTE CONFRONTARE RISULTATI CON E SENZA ***

Firstly, since given images have three equal levels of colours, they are in gray levels. Therefore it is possible to shrink the space occupied on disk by discarding hue and saturation information and using only luminance.

Rec.ITU-R BT.601-7 calculates luminance ($E\left[y\right]$) as:

$$E\left[y\right] = 0.299 * R + 0.587 * G + 0.114 * B$$

where R, G, B are the three image channels. Observe that since $R = G = B$, also $E\left[y\right] = R = G = B$, which justifies the assumption that discarding hue and saturation does not affect effectiveness of the system, whereas improving space and computational efficiency. Luminance is denoted by $E\left[y\right]$ since brightness is named y in literature, therefore the luminance, or the physical intensity expected is labeled this way.

Secondly, since pictures may be taken under different light exposure conditions, and since learning has heuristically be proven to be more efficient and effective if input assumptions

Fig. 2. Histogram of luminance distribution on sample image before linear scaling.

Fig. 3. Histogram of luminance distribution on sample image after linear scaling.

are always the same, linear scaling and hystogram equalization are done.

Linear scaling ensure that all images gray levels spread over all the range of possible values. $\mathcal{I}\left(x, y\right)$ refers to the luminance level of pixel (x, y) of image $\mathcal{I}$. Therefore, denoting with G_{max} the greater luminance level (tipically $2^k - 1$ for some k), the luminance scaled image is obtained as:

$$\mathcal{I}_{new}\left(x, y\right) = G_{max} \frac{\mathcal{I}\left(x, y\right) - \mathcal{I}_{min}}{\mathcal{I}_{max} - \mathcal{I}_{min}}$$

$$\mathcal{I}_{max} = \max_{x,y} \mathcal{I}\left(x, y\right) \ ; \quad \mathcal{I}_{min} = \min_{x,y} \mathcal{I}\left(x, y\right)$$

Histogram equalization is a technique for adjusting image intensities to enhance contrast.

TODO MATH ....

In figures 4 and 5 the effects of equalization on a sample image hystogram of luminance distribution are visible.

The difference between hystogram equalization and linear scaling (or stretching) is that ....

Pictures 6 and 7 show preprocessing output on a sample image.

IV. Region proposals

A. Contour detection

B. Image Segmentation

Introduction to alpha shapes and cite article describing proper segmentation usign alpha shapes.... describe parameteres and present limitations of such an approach in this practical application.... Therefore, bayesian optimization is proposed for alpha value.... Present table comparing different alpha values.... Explain evaluation scheme for alpha value optimization....

C. Bounding box

V. Defects area localization and classification

Introduction to problem, machine learning, ...

A. Proper data augmentation

To keep proper spatial information ....

B. Region of interests and proposals

Explain how to pick region of interests and proposals....

Fig. 4. Hystogram of luminance distribution on sample image before equalization.

Fig. 5. Hystogram of luminance distribution on sample image after equalization.

Fig. 6. Sample image before preprocessing.

Fig. 7. Sample image after preprocessing.

C. Region based Convolutional Neural Network (R-CNN)

1) Convolutional filters dimensioning: Stats on defects shapes....

2) Pooling layers: ....

3) ...: ....

D. Thresholding and location aggregation

Explain how from confidence value on different region of interests one can build the localization of defects area .....

E. Multi-task loss

Both localization and classification. If not classificated correctly but detected as defective .....

VI. Conclusion

Review the article, make some considerations on results and provide suggestions to further work...

References

[1] M. Sharifzadeh, S. Alirezaee, R. Amirfattahi, and S. Sadri, "Detection of steel defect using the image processing algorithms," in *2008 IEEE International Multitopic Conference*, December 2008, pp. 125–127.

[2] G. K. Nand, Noopur, and N. Neogi, "Defect detection of steel surface using entropy segmentation," in *2014 Annual IEEE India Conference (INDICON)*, December 2014, pp. 1–6.

[3] L. Chen and J. Deng, "Research on surface defects detection of stainless steel spoon based on machine vision," in *2018 Chinese Automation Congress (CAC)*, November 2018, pp. 1096–1101.

[4] Hongbin Jia, Y. L. Murphey, Jinajun Shi, and Tzyy-Shuh Chang, "An intelligent real-time vision system for surface defect detection," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, vol. 3, August 2004, pp. 239–242 Vol.3.

[5] J. Masci, A. Giusti, D. Ciresan, G. Fricout, and J. Schmidhuber, "A fast learning algorithm for image segmentation with max-pooling convolutional networks," in *2013 IEEE International Conference on Image Processing*, September 2013, pp. 2713–2717.

[6] A. Kumar and G. K. H. Pang, "Defect detection in textured materials using gabor filters," *IEEE Transactions on Industry Applications*, vol. 38, no. 2, pp. 425–440, March 2002.

[7] Y. Li and X. Di, "Fabric defect detection using wavelet decomposition," in *2013 3rd International Conference on Consumer Electronics, Communications and Networks*, November 2013, pp. 308–311.

[8] V. V. Karlekar, M. S. Biradar, and K. B. Bhangale, "Fabric defect detection using wavelet filter," in *2015 International Conference on Computing Communication Control and Automation*, February 2015, pp. 712–715.

[9] H. Y. Ngan, G. K. Pang, and N. H. Yung, "Automated fabric defect detection—a review," *Image and Vision Computing*, vol. 29, no. 7, pp. 442 – 458, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0262885611000230

[10] M. Vetterli and J. Kovačevic, *Wavelets and Subband Coding*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1995.

[11] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 1992.

[12] R. Bernardini, "Wavelets for differential equations and numerical operator calculus [online first]," February 2019. [Online]. Available: https://www.intechopen.com/online-first/wavelets-for-differential-equations-and-numerical-operator-calculus

[13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–44, May 2015.

[14] Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, and L. D. Jackel, "Handwritten digit recognition with a back-propagation network," in *Advances in Neural Information Processing Systems 2*, D. S. Touretzky, Ed. Morgan-Kaufmann, 1990, pp. 396–404. [Online]. Available: http://papers.nips.cc/paper/293-handwritten-digit-recognition-with-a-back-propagation-network.pdf

[15] D. Ciregan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, June 2012, pp. 3642–3649.

[16] P. Sermanet, S. Chintala, and Y. LeCun, "Convolutional neural networks applied to house numbers digit classification," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, November 2012, pp. 3288–3291.

[17] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, November 1998.

[18] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, "Object recognition with gradient-based learning," in *Shape, Contour and Grouping in Computer Vision*. London, UK, UK: Springer-Verlag, 1999, pp. 319–. [Online]. Available: http://dl.acm.org/citation.cfm?id=646469.691875

[19] R. Girshick, "Fast r-cnn," in *2015 IEEE International Conference on Computer Vision (ICCV)*, December 2015, pp. 1440–1448.

[20] X. Wang, H. Ma, and X. Chen, "Salient object detection via fast r-cnn and low-level cues," in *2016 IEEE International Conference on Image Processing (ICIP)*, September 2016, pp. 1042–1046.

[21] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," *arXiv e-prints*, December 2013.

[22] J. Kuruvilla, D. Sukumaran, A. Sankar, and S. P. Joy, "A review on image processing and image segmentation," in *2016 International Conference on Data Mining and Advanced Computing (SAPIENCE)*, March 2016, pp. 198–203.

[23] "Picture thresholding using an iterative selection method," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 8, no. 8, pp. 630–632, August 1978.

[24] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, November 2012.

[25] Jianbo Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, August 2000.

[26] L. Vincent and P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 6, pp. 583–598, June 1991.

[27] R. Klette, *Concise Computer Vision: An Introduction into Theory and Algorithms*. Springer Publishing Company, Incorporated, 2014.

[28] P. Kovesi, "Phase congruency detects corners and edges."

[29] M. Morrone and D. Burr, "Feature detection in human vision: A phase-dependent energy model," *Proceedings of the Royal Society of London. Series B, Containing papers of a Biological character. Royal Society (Great Britain)*, vol. 235, pp. 221–45, Genuary 1989.

[30] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, January 1988. [Online]. Available: https://doi.org/10.1007/BF00133570

[31] D. R. Martin, C. C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 5, pp. 530–549, May 2004.

[32] P. Stelldinger, U. Köthe, and H. Meine, "Topologically correct image segmentation using alpha shapes," in *Discrete Geometry for Computer Imagery*, A. Kuba, L. G. Nyúl, and K. Palágyi, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 542–554.

[33] P. I. Frazier, "A tutorial on bayesian optimization," *arXiv e-prints*, p. arXiv:1807.02811, July 2018.

[34] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," *arXiv e-prints*, p. arXiv:1206.2944, June 2012.

[35] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.