Dross attachment estimation in the laser-cutting process via Convolutional Neural Networks (CNN)

Luca Franceschetti[1], Matteo Pacher[2,3], Mara Tanelli[1], Silvia C. Strada[1],
Barbara Previtali[2] and Sergio M. Savaresi[1].

Abstract—Laser cutting of metals offers the advantage of high precision and accuracy. Dross attachment, measured as the length of the re-solidified material perpendicular to the surface, has definitely the highest impact on the overall process quality. Dross attachment is commonly judged by skilled technicians that evaluate the cut quality. Process parameters are optimized to maximize the cutting speed while keeping an acceptable level of dross attachment. However, in practice, increased levels of dross may occur due to different processing conditions. In this framework, a real-time dross attachment monitoring system is desired. Within the stream of vision based monitoring systems, in this work we use high frequency images generated by a precision camera, mounted on the laser head, to capture the cutting process light emission. A CNN-based classification system is developed, where captured images are fed into the trained network with the aim of automatically recognize if a predetermined dross attachment level is exceeded. To our best knowledge, this is the first work where a CNN is used for monitoring the quality of laser cutting process via dross attachment classification.

I. INTRODUCTION

Among the many uses of laser there is the one for cutting metal plates and tubes. On metals, the laser cutting process is highly accurate, offers a great productivity and makes it possible to cut very intricate shapes, [1], [2]. The laser beam can be thought as a column of high intensity infra-red light, of a single wavelength in the range of some hundreds of microns in diameter and travels from the laser resonator through the machine's beam path. The beam has to be precisely focused so that the power density at its waist diameter is extreme and the material either melts, burns, vaporizes and is eventually blown away by a jet of gas, leaving an edge with a high-quality surface finish. Among others, cutting speed, laser power, beam focal position and gas pressure are reported as the main influential parameters of the process, [3]. Nozzle diameter and distance between the end of the nozzle and the material surface are usually fixed to reference values. Finally, two kinds of assisting gas, namely, oxygen and nitrogen, are commonly used; according to the assisting gas type, two cutting modes are defined, that are *oxidation* and *fusion* cutting. The oxidation cutting mode benefits of the energy content given by the oxygen: the gas jet not only blows away the material but participates actively to the cutting operation thanks to an exothermic reaction. In the fusion cutting mode all the energy

[1]Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, via G. Ponzio 34/5, 20133 Milan, Italy. Email: {luca.franceschetti, mara.tanelli, silvia.strada, sergio.savaresi}@polimi.it

[2]Dipartimento di Meccanica, Politecnico di Milano, via la Masa 1, 20156 Milan, Italy. Email: {matteo.pacher, barbara.previtali}@polimi.it

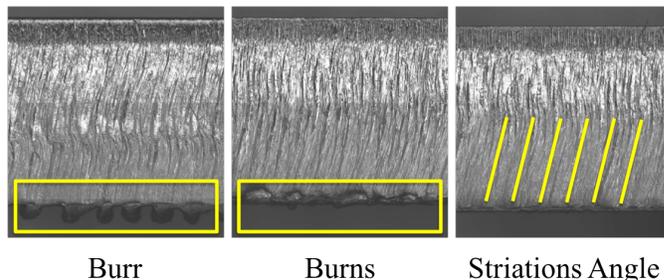[3]Adige S.P.A., Via per Barco 11, 38056, Levico Terme (TN), Italy.

Fig. 1. Possible defects in the laser-cutting process.

required to melt the material is given by the laser beam and the role of the gas consist in removing the molten material only. Due to the larger field of application in terms of material type and the increased productivity that it offers, the fusion cutting mode is replacing the oxidation cutting one in most of the applications and this study is accordingly focused on this kind of process. Considering the fusion cutting mode, the main defects are given, according to both industrial practice and relevant standards, by (see also Fig. 1):

- dross attachment measured as the length of the re-solidified material perpendicular to the material surface;
- surface roughness of the cut edges;
- striation's pattern
- presence of a burned bottom profile due to a too high energy input.

Dross attachment has definitely the highest impact on the overall quality of the laser cutting process, [3], [4], [5], [6]. Commonly, process quality is not a quantitative measure but it consists of a judgement given by skilled technicians that brings together the mentioned defects in a single synthetic mark. In practical applications, process parameters are a priori optimized with respect to quality, productivity and process reliability. Manufacturers try to increase cutting speed as much as possible, to reduce process time as far as the process quality for all cutting conditions is ensured. It is observed that, within a suitable range, the roughness decreases with the cutting speed. Therefore, dross attachment becomes the real limiting factor to the increase of cutting speed. The online estimation of dross attachment represents a challenging and innovative issue in the laser cutting context. Related works employ photodiodes [7], [8], to transduce the cut image from an optical signal to an electrical one. More recently, camera based monitoring systems are used, which guarantee a much more high quantity of information of the cut, [9], [10], [11],
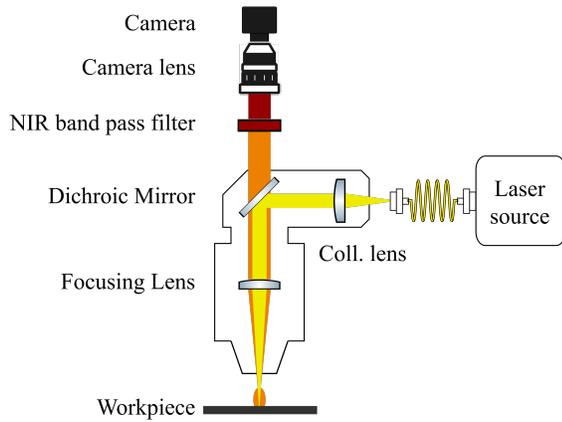
Fig. 2. Experimental architecture for process monitoring.

[12]. The present study addresses the problem of determining whether a significant and non negligible dross attachment is present or not in a performed cut, based on the high resolution camera images, which are collected at high speed during the cutting process. We estimate the dross attachment by resorting to a suitable Convolutional Neural Network (CNN), [13], [14], starting from the camera images. We will use the trained CNN in "binary classification" mode: having chosen a set of geometrical descriptors below which the dross height is considered to be of acceptable quality, then the CNN will classify "dross-yes" or "dross-no" based on the fact that this quality level is exceeded. To our best knowledge, this is the first work where a CNN is used for dross attachment estimation in the monitoring of the laser cutting process. We will show in detail how, a proper on-line training of the CNN, can yield satisfactory and promising results, with the significant advantage of avoiding the time-consuming design phase inherent in other estimation techniques.

II. Problem Statement and Experimental Setup

The goal of our work is to tackle the problem of dross attachment estimation from indirect kerf characteristics' measurement. In order to perform the indirect estimation of dross formation a dedicated experimental setup has been developed, which is depicted in Fig. 2. All cutting experiments were performed with a BLM Group LC5 machine [15], equipped with an IPG YLS-6000 laser source with a maximum available power of 6kW and a fiber diameter of 100μm. The cutting head was a Precitec HPSSL, that has a collimation lens of 100mm and a focal lens of 200mm. Therefore, the laser spot in the bore of the nozzle has a dimension of 200μm. A microcontroller was adopted to monitor inputs (process parameters) at a sample rate of 10kHz. The microcontroller then outputs a TTL signal that triggers the camera acquisition and permits to synchronize videos with signals. This platform was chosen considering also the possible future implementation within a feedback control scheme. The monitoring equipment is composed of a Ximea [16] camera, mounted coaxially with respect to the laser beam. A proper camera lens was chosen

based on a target framerate of 1500fps and a target spatial resolution of at least 10μm/px. Then, different tests using external illumination, or simply acquiring the process emission, were performed to choose proper optical filtering. To design and test the estimation methodology, a dedicated experimental campaign was needed, to obtain cuts with different levels of dross attachment, for different materials and thicknesses. All the performed cuts were square sided specimens of side 45mm.

III. CNN based Dross attachment Estimation

In the present contribution we will describe a CNN custom architecture for dross attachment estimation from indirect measurements coming from the high-speed camera. This section is organized as follows: we firstly describe how the ground truth of the dross attachment in laser cuts is normally defined. Then, we provide in Section III-B and III-C some theoretical background on CNNs and our custom CNN structure. We then describe the dataset used in this work in Section III-D. Section III-E and III-F outline the network training phase and the validation stage. Furthermore, Section III-F includes a detailed description of the CNN-based algorithm testing process together with its performance discussion.

A. Dross attachment Ground Truth Definition

After the cuts are carried out, pictures of the specimens are taken. A single specimen side is represented by three consecutive pictures, that are then stitched together. Stitching does not affect the estimation performances because is univocal. Stitched images are processed with an edge-detection tools, used to reconstruct the perimeter of each side of the specimen and to reconstruct the overall specimen dross measurement. The latter will be used as ground truth in the following estimation process, as shown in Fig. 3.
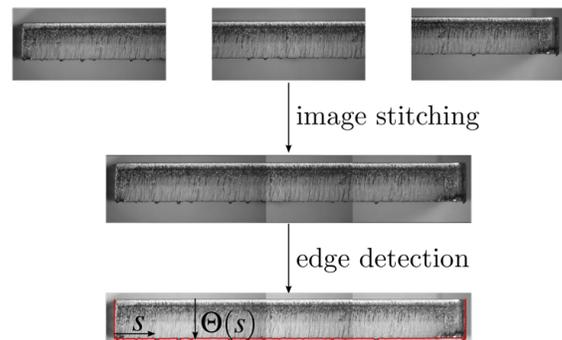

Fig. 3. Spatial measurement $\Theta(s)$ of the specimen profile after the cut is performed.

Once the above process is applied, a measure of the whole thickness is obtained, then an estimation of dross attachment is calculated.

$$h(s) = \Theta(s) - \Theta_{real} \qquad (1)$$

In Eq. 1, $h(s)$ represents the measure of the dross profile, $\Theta(s)$ represents the measured thickness obtained by image stitching

and edge detection algorithms and Θ_{real} is the real thickness of the specimen. All this variables are function of the spatial coordinate s, along the perimeter of the considered specimen.

B. CNN Fundamentals

The Convolutional Neural Networks (CNN) are a type of Artificial Neural Networks (ANN). The CNN can automatically learn a hierarchy of features from the input image matrices, which prove to be better than those hand-crafted features extracted by carefully designed complex algorithms [13]. CNNs automatically assign importance to various aspects and objects in the images. By discovering and differentiating those aspects and objects, CNNs can learn filters to isolate a specific characteristic inside images. A typical CNN architecture consists of several nested convolutional and pooling layers followed by fully connected layers at the end. One of the first efficient example of CNN is AlexNet [17], followed by many others in the consecutive years. A compact presentation of this kind of network [Input - Conv - ReLU - Pool - FC] consists of the following five layers:

- **Input**: the input of the CNN is the image. More specifically, in our application inputs are 210×210 pixel 1-channel grayscale images, containing the intensity values of each pixel.

- **Conv**: the convolutional layers apply a set of learned filters each of which is connected to only a small region of the output. So, the filters are usually learnable matrices, updated during training, of small size (like 3×3 or 5×5). The convolution operation is applied over the whole input image, across the spatial dimensions, to extract one feature, performing a matrix multiplication between the filter and the portion of the image over which the kernel is hovering. It is still needed to specify manually some numerical parameters such as the number of filters, the size and the architecture of the whole network.

- **ReLU**: ReLU (Rectified Linear Units) is the most widely used activation function by adding non-linear transformations to the output response of the convolutional or fully connected layers. It is an element-wise operation applied per pixel. The formula of this function is $f(x) = max(0, x)$ and it replaces all negative pixel values with a zero. ReLU can effectively prevent the gradients from saturating, expedite convergence of the training procedure while at the same time keeping the original value to the most extent, which proves to be experimental better than conventional sigmoid-like activation functions [13], [17], [18]. Convolution is a linear operation since it applies an element-wise matrix multiplication and additions. Introducing non-linearity in the CNN accounts for most of the real-world data, that contain non-linearity, otherwise unrecognisable by the convolution operation itself.

- **Pool**: the pooling layer performs a form of non-linear down-sampling along both spatial dimensions, leading to reduced spatial size of the convolved features. It aims to reduce the amount of the network parameters and the computational cost, keeping the most important information inside the convolved features. The pooling layer is commonly placed between two successive convolutional layers. The most common pooling strategy is "max" pooling, which outputs the max value from the neighbourhood of the input feature map. Extracting dominant features results in convolved features which are rotational and positional invariant. Simple ANN have little invariance to shifting, scaling, and other forms of distortion because they process images directly with pixel values as inputs. Results will turn out to be very bad with input image slightly modified (for example translated by 1 pixel). Pooling layers allow CNN to be almost completely invariant to forms of distortion.

- **FC**: The fully connected layers are the last part of the neural networks. All the neurons in the fully connected layers are connected to all the units of the last layer. And the last fully connected layer generates the output of the whole network.

These layers are stack together to form a CNN, the input is fed forward into the network, and the hyperparameters are updated by the back propagation algorithm.

C. Structure of the CNN for dross attachment estimation

As shown in Fig. 5, the custom CNN structure used consists of 24 layers. To summarize, the architecture can be described as: input layer, $C(32,3,3)$, $S(2,2,2)$, $C(32,3,3)$, $C(64,3,3)$, $S(2,2,2)$, $C(128,3,3)$, $C(128,3,3)$, $S(2,2,2)$, $FC(1024)$, $FC(1024)$, $FC(2)$, softmax layer. $C(n, 3, 3)$ represents a convolutional layer with n filters of kernel size 3×3, $S(2, 2, 2)$ represents a pooling layer with a sub-sampling factor of 2×2 by stride 2 in both dimensions, $FC(n)$ represents a fully-connected layer with n neurons. The pooling strategy adopted in all the pooling layers is max-pooling, which is robust to distortions. After all the convolutional layers, a ReLU layer is applied. The last fully-connected layer generates the output, composed by 2 numbers, corresponding to the classes adopted (dross present or not), meaning the estimation of each image's class membership probability. Softmax layer is used to select the class with higher probability.

D. Images' Dataset Description

The dataset is composed by 210×210 pixel 1-channel grayscale images. Images are captured by an high-speed camera (1500fps), mounted coaxially with respect to the laser beam, as described in Section II. Images shows the real-time

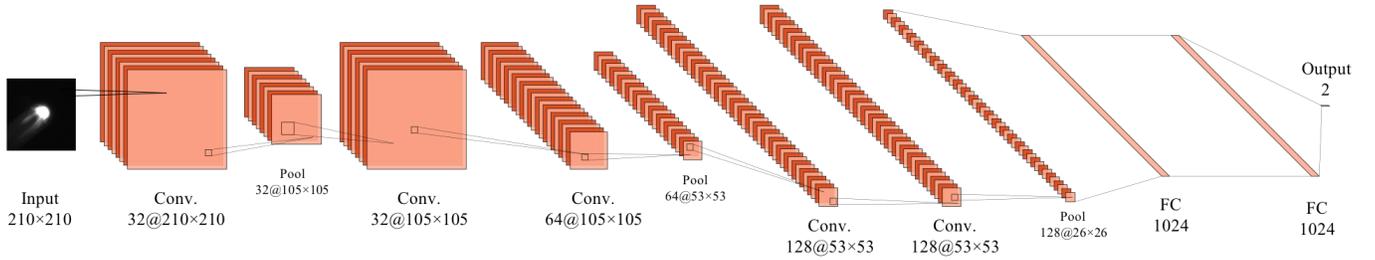
Fig. 4. Example of image in the dataset.

Fig. 5. Schematic architecture of the custom CNN.

emission of the laser irradiated zone, as can be seen in Fig. 4, where the black part is simply the background and the grayscale values are related to the material temperature. We will concentrate on shape and dimensions of the white part, called "blob". For each image, we also have the real dross profile $h(s)$, as described in III-A. Images are captured from a fusion-cutting process of stainless steel (AISI 304) and mild steel. Both materials are considered with different thickness, in particular 3, 5, 8 and 10 millimetres. We want to achieve a general purpose recognition of dross so the CNN is trained to recognize defected images for specimen of both materials and the different thicknesses. The whole dataset has more than 10000 images. To better understand the information collected from the camera, let us consider Fig. 6, which shows a camera image during a cut of AISI 304 with different thicknesses, ranging from 3 to 8 mm.
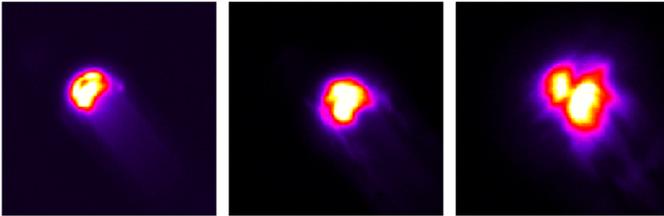

Fig. 6. Cut of AISI 304 with a different thicknesses, ranging from 3 to 8 mm, starting from the left image. Shown images are modified with pseudo-color transformation.

Of course, the camera has a very high frame rate (1500fps), so that the overall cutting process yields many images that can be processed. It can be seen that the different cuts generate different blobs in the images because different thickness produces different cutting process. So our CNN is trained to recognize defect in different kind of cutting processes, as our goal is to achieve a general purpose dross recognition.

E. CNN Training and Testing

CNNs work with classification problems, so two class had to be created: dross present and not present. In this way, expressing a threshold in millimetres, is possible to detect the presence or the absence of defects. The threshold was discussed with expert technicians and set to 0.03mm. With this specific threshold, all the images are divided into the
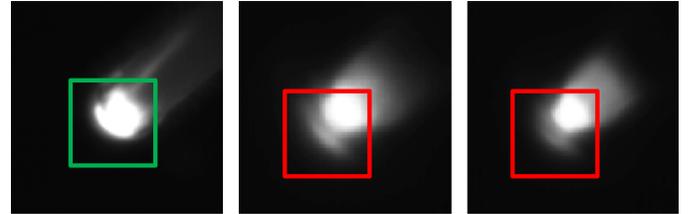

Fig. 7. Difference between well-classified and misclassified images, highlighted with red square. First image is an example of well-classified, second and third images are examples of misclassified images.

two classes. The two created classes are perfectly balanced in number of elements; half of the images belong to the class with defect and the other half belong to defect-free class. During training, the whole dataset is divided in three different sets:

- training set, used during training to update weights and biases of CNN;
- validation set, used for validating updates during training;
- test set, used for testing accuracy results with CNN after the whole training procedure is finished.

In our experiments, we randomly choose 70% of the data as the training set, 20% of the data as the validation set and the rest 10% as the test set. The output of the trained CNN is a binary indicator $h_{th}(s)$. An example of a real output signal is reported in Fig. 8, case b). First training sessions showed that image and dross profile $h(s)$ frequencies were decoupled; the final accuracy results were always around the "random guess" percentage. This means that CNN could not learn anything from images. Analysing the spectrum of $h(s)$, we discovered that the maximum spatial frequency was 150Hz. For this reason, we created an averaged dataset, using 10 of the original frames, captured at 1500fps, to create an averaged image from 10 subsequent ones. The average is made pixel-by-pixel. The created average dataset frequency is approximately 150Hz and coupled with the dross signal. Therefore, a deep knowledge of the laser-cutting process is required to correctly train the CNN.

F. Training Results and Misclassification

Training results show a final accuracy of 93% over the validation set and a final accuracy of 92% over the test set. This means there is no observation of over-fitting in the

training process. A misclassification of 7-8% is acceptable in this application field because our CNN is trained to recognize defect in different cutting processes. Images of different materials and thicknesses are quite different, so the CNN has to learn a wide set of features to classify them correctly. On the other hand, we decided to understand and investigate why there is such misclassification. In particular, we aimed to understand if there exists particular conditions of material and/or thicknesses related to misclassification. It turns out that there was an even distribution of misclassifications over thickness and materials, meaning that characteristics of specimen do not influence training results and misclassification. We also visually investigated the images, and we discovered that misclassified ones tend to have a very particular characteristic in the front part of the *blob* (Section III-D). Specifically, as can be seen in Fig. 7, the front of the *blob* (red square in the images) is not well defined in misclassified images. Conversely, the first image shows a well-classified image, in which the front of the *blob* is coincident with laser ray. So, the common characteristic of misclassified images is the antecedent "scarce illumination" of *blob*, identified in front of direction of laser ray during cutting.

IV. Enhanced CNN Dross attachment Estimation via Regression

In order to have a continuous output variable, that indicates the quality of the laser cutting process, the binary indicator $h_{th}(s)$ (introduced in Section III-E) is converted from spatial domain to time domain in $h_{th}(t)$ using a spatial/temporal relationship, that is acquired from machine's process parameters. Using the same relationship $h(s)$ (see III-A) is converted in $h(t)$. Then $h_{th}(t)$ is manipulated with a sliding window. A windows of 150ms is selected and the sliding window output $y(t)$ is calculated using

$$y(t) = \frac{1}{SW} \sum_{k=1}^{SW} h_{th}(t-k) \qquad (2)$$

where SW indicates the number of frames included in 150ms of data. With our experimental setup, in 150ms are acquired

225 frames with a framerate of 1500fps (see Eq. 3).

$$SW_{frames} = 0.15\,[s] \cdot 1500 \left[\frac{frame}{s}\right] = 225\,[frames] \quad (3)$$

The whole transformation is depicted in Fig. 8. The output $y(t)$ is a continuous variable, varying between 0 and 1, representing the likelihood of having dross attachment over time in a specific cut. Thus, a regression output is obtained from the binary output of the CNN. With $y(t)$ as a continuous variable is possible to continuously control cutting parameters, i.e. speed or pressure (see Section I). Furthermore, $y_{truth}(t)$ is defined as the measurement of the dross profile processed with sliding window method. More precisely, starting from $h(t)$ and using the sliding window method in the same way as is done for $h_{th}(t)$, $y_{truth}(t)$ is obtained. In this way, $y(t)$ and $y_{truth}(t)$ are comparable, and the difference between the two represents the error in the estimation of dross with the CNN. Therefore, an estimation performance analysis is carried out, comparing $y(t)$ with $y_{truth}(t)$, for each material and thickness. Overall, the performances are satisfying. More precisely, dross attachment estimation is perfectly accurate in dross-free cuts and in cuts with very high-dross cuts, that means the error on defects estimation is approximately zero. An accurate analysis of cuts with intermediate-dross values shows that high-frequency changes in $h(t)$ lead to a wrong classification with the CNN. In particular, when $h(t)$ is around the chosen threshold and switches frequently above and below it, the CNN-based classification contains a non negligible number of false positives. This is obviously visible also from $y(t)$. An example of false positive detection in an intermediate dross cut is shown in Fig. 9. The overall performances of the enhanced CNN dross estimation via regression on the whole dataset are summarized in Table I.

V. Conclusions

A CNN-based method has been presented for the estimation of dross attachment in laser cutting experiments performed by an industrial laser cutting machine, starting from the images collected through a high-speed camera. This kind of vision based monitoring of the cutting process on the one side
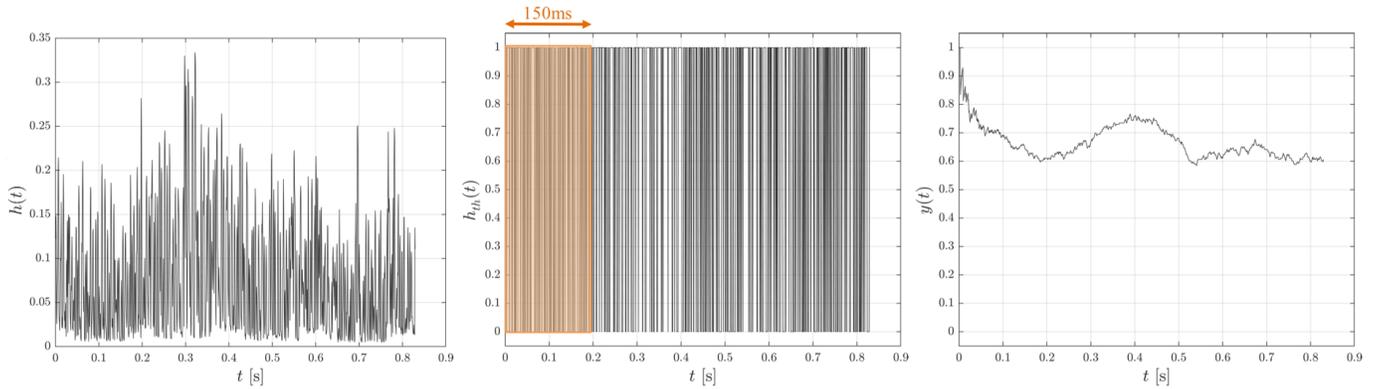

Fig. 8. a) Real dross attachment profile of an illustrative laser cut. b) Binary indicator. c) Sliding window of the binary output.
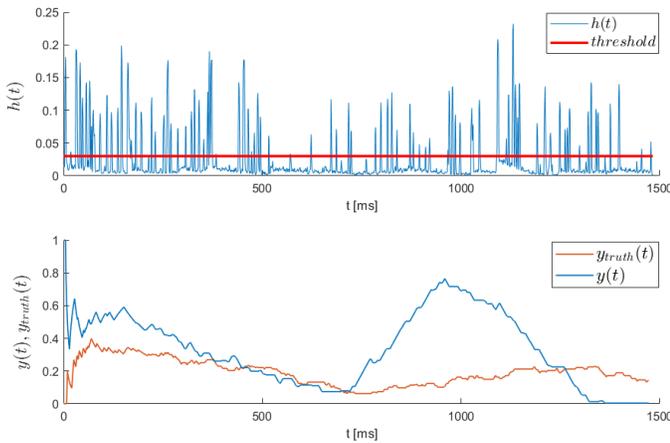
Fig. 9. a) $h(t)$ and quality threshold in illustrative intermediate-dross cut. b) $y(t)$ and $y_{truth}(t)$ in the same illustrative intermediate dross cut.

TABLE I
RESULTS OF THE ENHANCED CNN DROSS ATTACHMENT ESTIMATION VIA REGRESSION.

Cut Type	Classification	Regression
dross-free cuts	98%	99.8%
low-dross cuts	95%	97%
intermediate-dross cuts	85%	87%
high-dross cuts	99%	99.9%

tries to overcome the need of an a-posteriori measurement of dross while, on the other side, aims at developing a real-time algorithm for the detection of a significant dross level and hence at making possible the development of a future closed-loop control system, able to read the sensed dross presence and to act, if needed, reducing the cutting speed, one of the main process variables. We show in the paper how a custom CNN architecture, once suitably trained with precaution based on laser-cutting knowledge, can automatically differentiate images captured during the laser cutting process. The real-time images exhibit a black background with a white part (the laser irradiated zone), of different geometry and shape depending on the material and its thickness, from which the network is able to recognize the presence of dross attachment that exceeds an accepted maximum level. The CNN, operated on the validation and test sets, shows an accuracy of about 92% against both datasets. A detailed investigation was also performed on the misclassified images letting us identify the situations where the CNN recognition task is more challenging. In order to generate a continuous dross estimation, the original binary CNN response (presence or non-presence of dross, according to a suitable threshold) is weighted over the total length of a sliding time window while the cutting is run, so that an enhanced regression-adapted CNN is built. Future work will focus on the better understanding of those intermediate-dross sections where the continuous output signal leads to erroneous dross attachment estimates.

REFERENCES

[1] ISO Central Secretary, "Thermal cutting — classification of thermal cuts — geometrical product specification and quality tolerances," International Organization for Standardization, Geneva, CH, Standard, 2017.
[2] W. M. Steen and J. Mazumder, *Laser material processing.* springer science & business media, 2010.
[3] J. Pocorni, D. Petring, J. Powell, E. Deichsel, and A. F. Kaplan, "The effect of laser type and power on the efficiency of industrial cutting of mild and stainless steels," *Journal of manufacturing science and engineering*, vol. 138, no. 3, p. 031012, 2016.
[4] P. Sforza and V. Santacesaria, "Co2 laser cutting: analytical dependence of the roughness of the cut edge on the experimental parameters and process monitoring," in *Laser Materials Processing: Industrial and Microelectronics Applications*, vol. 2207. International Society for Optics and Photonics, 1994, pp. 836–847.
[5] M. Schleier, B. Adelmann, C. Esen, and R. Hellmann, "Cross-correlation-based algorithm for monitoring laser cutting with high-power fiber lasers," *IEEE Sensors Journal*, vol. 18, no. 4, pp. 1585–1590, 2017.
[6] M. Pacher, L. Monguzzi, L. Bortolotti, M. Sbetti, and B. Previtali, "Quantitative identification of laser cutting quality relying on visual information," in *Lasers in Manufacturing (LiM) 2017*, 2017, pp. 1–11.
[7] J. De Keuster, J. Duflou, and J.-P. Kruth, "Monitoring of high-power co2 laser cutting by means of an acoustic microphone and photodiodes," *The International Journal of Advanced Manufacturing Technology*, vol. 35, no. 1-2, pp. 115–126, 2007.
[8] B. Adelmann, M. Schleier, B. Neumeier, and R. Hellmann, "Photodiode-based cutting interruption sensor for near-infrared lasers," *Applied optics*, vol. 55, no. 7, pp. 1772–1778, 2016.
[9] P. Wen, Y. Zhang, and W. Chen, "Quality detection and control during laser cutting progress with coaxial visual monitoring," *Journal of Laser Applications*, vol. 24, no. 3, p. 032006, 2012.
[10] I. Eriksson, P. Gren, J. Powell, and A. F. Kaplan, "New high-speed photography technique for observation of fluid flow in laser welding," *Optical Engineering*, vol. 49, no. 10, p. 100503, 2010.
[11] J. Duflou, E. Fallahi Sichani, J. De Keuster, and J.-P. Kruth, "Development of a real time monitoring and adaptive control system for laser flame cutting," in *Proceedings of the 28th International Congress on Applications of Lasers and Electro-Optics, November 2-5, 2009-Orlando, Fl, USA*, vol. 102. Laser institute of America, 2009, pp. 527–536.
[12] E. F. Sichani, J. De Keuster, J.-P. Kruth, and J. Duflou, "Monitoring and adaptive control of co2 laser flame cutting," *Physics Procedia*, vol. 5, pp. 483–492, 2010.
[13] T. Wang, Y. Chen, M. Qiao, and H. Snoussi, "A fast and robust convolutional neural network-based defect detection model in product quality control," *The International Journal of Advanced Manufacturing Technology*, vol. 94, no. 9-12, pp. 3465–3471, 2018.
[14] D. Weimer, B. Scholz-Reiter, and M. Shpitalni, "Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection," *CIRP Annals*, vol. 65, no. 1, pp. 417–420, 2016.
[15] *Macchina taglio laser 3d per tubi e profilati di grandi dimensioni — BLM GROUP*. [Online]. Available: https://www.blmgroup.com/it/lasertube/lc5
[16] *XIMEA website*. [Online]. Available: https://www.ximea.com
[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
[18] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.