17th IMEKO TC 10 and EUROLAB Virtual Conference
"Global Trends in Testing, Diagnostics & Inspection for 2030"
October 20-22, 2020.

Effect Of Measurement Uncertainty On Artificial Vision Methods, For Quality Control On Composite Components

Gaspari, Antonella[1]; Natale, Emanuela[2]; De Silvestri, Armando[3]; D'Emilia, Giulio[4]

[1] *University of L'Aquila, L'Aquila, Italy, antonella.gaspari@univaq.it, +390862434479*
[2] *University of L'Aquila, L'Aquila, Italy, emanuela.natale@univaq.it, +390862434344*
[3] *University of L'Aquila, L'Aquila, Italy, armando.desilvestri@student.univaq.it*
[4] *University of L'Aquila, L'Aquila, Italy, giulio.demilia@univaq.it, +390862434324*

***Abstract* – In this work, an inspection strategy by a vision system is analysed, for the identification of surface and aesthetical defects, with reference to composite components for automotive and aeronautical industrial sectors. Attention is paid to the background identification, since the specificity of the application requires particular care in order to avoid misunderstandings and false negatives during the detection phase. The evaluation of the parameters set-up effects is used for the identification of the main uncertainty contributions, which is a strong support for the most suitable choice of the monitoring strategy. The robustness of the approach is studied with reference to several laboratory datasets, representing some commonly found issues for an easy in-field transfer. To this aim, some commercial tools available in Matlab ® environment have been used. The obtained results encourage to monitor the variability of the performances rates, depending on the qualitative levels to be achieved during the operating conditions and on the desired reliability of the approach.**

Keywords – Measurement uncertainty; Artificial intelligence; Deep learning; Vision systems.

I. INTRODUCTION

The rapid development of artificial vision systems is gaining ground in the industrial field, thanks to a series of factors like reliability, robustness and relatively low-cost impact [1]. Their actual and potential trends of applications reside in those activities linked to the product and process quality control and to the autonomous driving of robot and motion systems. Nowadays, the measurement and inspection abilities are often integrated with some artificial intelligence tools, which are spreading into the industrial facilities since their degree of ease of use is gradually increasing [2]. Nevertheless, focusing, as an example, on the quality control, the success of this kind of applications requires to be fully aware of the vision process, especially in the analysis of complex components, in terms of defects' form and dimensions, their detectability – which may be compromised by the geometry of the component or by its surface characteristics, and the environmental conditions, which may influence the set-up choices [3-4]. In this work, an inspection strategy is analysed, for the identification of surface and aesthetical defects, with reference to composite components for automotive and aeronautical industrial sectors. In this type of materials, the warp and weft direction of fibres inherently may mislead the artificial vision system, pushing towards some tailor-made solutions for monitoring, which may also contribute to the optimization of the technological process [5-8].

The literature review offers several solutions, making the generalization difficult, also of very powerful solutions [9]; therefore, a measurement uncertainty based method is useful to define the instrumental, technical and functional aspects of architecture, together with data processing and management techniques to improve the in-field transfer of these solutions. A survey concerning the process on the whole, evidently requires effort and care: in this work, some methodological aspects are faced, useful to define the procedure best-practices. In general, a vision based quality control procedure is composed of the following main steps: (1) Definition of a cluster of defects; (2) Analysis of the physical principle for defect identification; (3) Optical configuration (4) Design of experiments for image acquisition; (5) Images pre-processing; (6) Features extraction; (7) Classification/Recognition.

One of the most critical step lies on step (5), which is not usually examined thoroughly, from the point of view of the variability that can affect the results. Traditionally, one of the main aim of the pre-processing phase is to focus on the Region of Interest (ROI), neglecting all parts of the image useless for the inspection of components, like the background. This outcome is used to reduce computational load and time, also improving the overall accuracy.

Furthermore, many tools, which are based on Deep Learning algorithms allow realizing both images pre-processing and defect detection [9-14].

Defect segmentation of textured surface images poses challenges such as ambiguous shapes and sizes of defects along with varying textures and patterns in the images. [15]. The variation of defect patterns is large; furthermore, the images can vary if the image acquisition condition changes quite a little. Additionally, the defect size must be considered in the determination of defects [16]. Therefore, it appears evident the usefulness of focusing on the surface of interest, avoiding wasting computational and timing efforts by processing areas of the images depicting unwanted and worthless background.

This paper refers to different techniques for the background detection and subtraction like semantic segmentation and box labelling and highlights limitations and benefits of each of them. Datasets have been specifically created for this work, suitably sized for *Transfer Learning* procedures, with the aim of emphasizing the peculiarities of the objects to be investigated more than improving the performances of the algorithms that are commercially available. The evaluation of the parameters set-up effects is a preliminary effort for the definition of the uncertainty budget, with reference to image analysis by Artificial Intelligence Techniques, according to previous work of the authors [17-18]. In particular, it allows to define the factors on which basing a correlation between the traditional metrics and the uncertainty causes, in order to delimit the area of investigation.

II. MATERIALS AND METHODS

Two different techniques for the background detection and subtraction are compared: Object Detection and Semantic Segmentation methods.

The dataset used for training, validation and testing is composed by 86 images referred to composite specimen of different size and geometry. The dataset can be divided into two groups: one, where the specimen covers more than 80% of the picture (Fig. 1a) and the other, where a thinner portion of the composite crosses the whole picture, with a width of around 20% of the total (Fig. 1b).

In the following, the operative procedure for the two approaches is illustrated, together with the parameters used for comparison purposes.

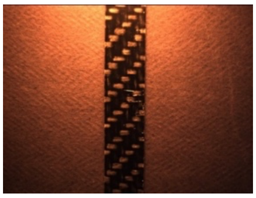

a)

b)

Fig. 1. Examples of pictures from the dataset: a) first group; b) second group

A. Object Detection: Bounding box

Object Detection consists of recognizing different objects inside an image, labelling them by means of rigid boxes. In this work, a Faster Region Convolutional Neural Network (Faster R-CNN) is used. It involves the Region Proposal Network (RPN) inside the CNN, useful to analyse the image to the aim of distinguishing background and foreground. This selection is made by means of some *Anchors*, i.e. the rectangular bounding boxes anchoring the object to be detected. The performance of the CNN is affected by the number and size of these *Anchors* and by the pitch along the vertical and horizontal directions, on which the *Anchors* are distributed onto the image to be processed. The number of *Anchors* has been preliminarily optimized on the base of the *MeanIoU* index (i.e. the one's complement of the Jaccard coefficient) and of the best trade-off between the computation time and the spatial resolution of the *Anchors*.

The CNN used for the features extraction is the Resnet50; in particular the layer for the feature extraction is the *Activation_40_relu*, showing a satisfactory trade-off between the spatial resolution and the extracted functionalities. The general training options and hyperparameters are as follows: stochastic gradient descent with momentum, as the iteration method for the optimization of the differentiable functions, 10 maximum epochs, 2 maximum samples of the dataset contemporarily processed and an initial learning rate of 10^{-3}.

By default, the inspection net considers a *Negative Overlap Range* of 0 to 30% and a *Positive Overlap Range* of 60% to 100%.

B. Semantic Segmentation

Semantic Segmentation allows labelling different categories of the image pixel by pixel [19]. The architecture is based on two neural networks: one for the identification of the Region of Interest (ROI) and the other one for the features extraction. Once the dataset is ready, i.e. the initial manual labelling of the pixels is completed, the necessary inputs to create the segmentation net are: the size of the images to be analysed (350*350 pixels), the number of classes (n.2) and the CNN for the feature extraction step (*Resnet18*). The segmentation net used is the DeepLab-V3.

The two categories that are claimed to be recognized are: the Fiber Reinforced Thermoplastic Composite (FRTC) and the Background (BG).

A mandatory pre-processing step before training, is balancing the effect of different number of pixels between the two classes by suitably modifying the weight of the pixels contained in each class, in the final layer of the net.

For comparison purposes of results, commonly used indices are used; the class parameters are [20]:

- *Accuracy (A)*: ratio between the total pixels correctly identified of a class over the number of pixel of that class;

- *Intersection over union (IoU)* or Jaccard similarity coefficient: measure of the statistical precision penalizing the false positives;
- *Boundary F1 countour matching score* (*BFScore*): gives an indication of how close the predicted boundary of a class matches the ground truth boundary.

The global parameters are defined as follows:

- *Global Accuracy (GA):* ratio between the total pixel of one class on the total number of pixel;
- *Mean Accuracy:* average of the *Accuracy* values of all the classes;
- *MeanIoU*: average od the *IoU* values of all the classes;
- *WeightedIoU:* average of the IoI of each class, weighted with respect to the number of pixel of that class;
- *MeanBFScore:* average of the BFScore of all the classes.

III. RESULTS

In this Section, the effect of the choice of different combinations of training options and typology of data augmentation is described, in order to individuate the best strategies to be applied during the training phase, thus avoiding undesired variability.

A. Object Detection: Bounding box results

The net has been trained splitting the dataset into the following: 60% training, 20% validation, 20% testing. The reliability assessment of the net has been carried out with reference to random rotations of the pictures. The resulting independent training datasets and corresponding nets are:

- Base: no image has been modified;
- 90°: half of the pictures have been rotated by 90°;
- Gradual: evenly spaced rotations of the images in the range 0°-90°.

Fig. 2 gives an illustrative example of the obtained results, with reference to the two different group of images of Fig. 1. As it could be expected, a more robust net is obtained if the configurations are varied during training.

Despite the general performance of the *Gradual* net is over 90%, it cannot be assumed as totally satisfactory, since the specimen is not framed on its whole.

As already mentioned, the factors that mostly influence the net's behaviour are the use of a variable learning rate and the tuning of the *Negative* and *Positive Overlap Ranges*. The Learning Rate Schedule has been therefore modified and set to 2, the Factor for dropping the learning rate to 0.3, the initial learning rate used for training is chosen as 10^{-4} and a L2 regularization factor of 0.005. Furthermore, using a Positive Overlap Range of 0.5 to 1, the average precision obtained is 98%.

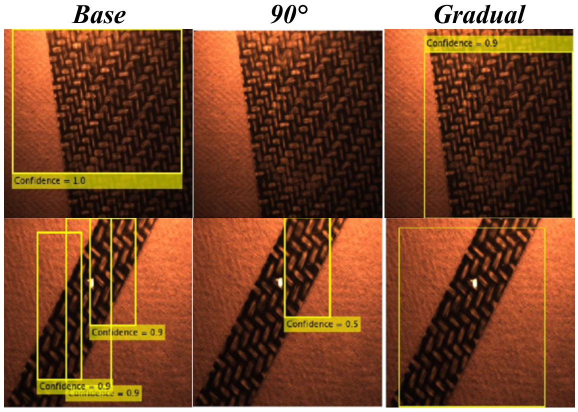


Fig. 2. Examples of the output for Base, 90° and Gradual nets, for the first group of images (top line) and second group of images (bottom line)

Nevertheless, failure to recognize a part of the component being analysed, makes it impossible to identify defects in the area that is not considered, leading to a change in the investigation strategy (e.g. increasing the number of acquired images to be processed or the number of anchor boxes) or using different background recognition techniques, as shown below.

B. Semantic Segmentation results

The following factors have been studied and compared in this subsection: the use of constant *versus* variable learning rate (LR), the training duration, the data augmentation and the dataset balancing.

Constant Learning Rate vs Variable Learning Rate. The effect of LR has been studied considering a constant value set at 10^{-4} and a variable one, constant at 10^{-3} up to the first 10 epochs and decreasing by a factor of 0.3 for the remaining epochs. Two independent datasets are tested and compared, namely *Original* and *Filtered*. Both consist of images that have been randomly rotated and flipped on the X/Y directions. *Filtered* refers also to contrast-enhanced images.

The effect of different strategies for LR is shown in Tables 1 and 2, with reference to the Original dataset, being the behavior of all the indices similar for both datasets (*Original* and *Filtered*).

Table 1. Global indices for constant and variable LR.

	Constant LR	Variable LR
GlobalAccuracy	0.8717	0.8977
MeanAccuracy	0.8785	0.904
MeanIoU	0.7726	0.8144
WeightedIoU	0.7323	0.8143
MeanBFScore	0.5327	0.6145

Table 2. Class indices for constant and variable LR.

	Class	Accuracy	IoU	BFScore
Constant LR	FRTC	0.9657	0.7763	0.4804
Variable LR	FRTC	0.9848	0.8162	0.5514
Constant LR	BG	0.7914	0.7688	0.5389
Variable LR	BG	0.8232	0.8127	0.6418

A variable LR improves the performances by 2% in terms of accuracy and up to 25% in terms of BFScore. The increase of BFScore is very useful, as for the separation of the sematic areas is concerned. Fig. 3 gives a visual representation of the of the results of semantic segmentation with constant and variable learning rates, for both testing datasets. The positive effect of variable learning rate is confirmed.

Input image Constant LR Variable LR

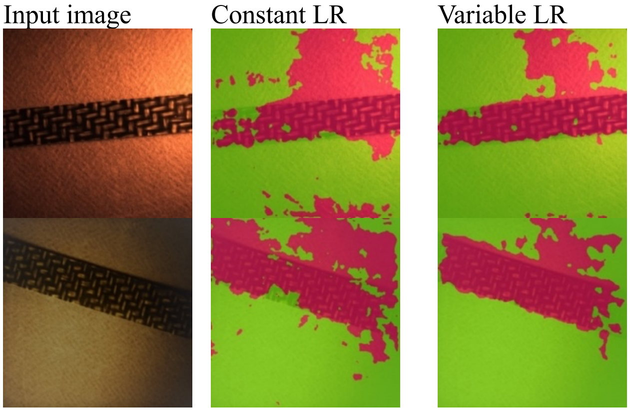


Fig. 3. Segmented areas for Original (top line) and Filtered (bottom line) images: pink and green areas means FRTC and Background, respectively.

Training duration. Generally speaking, training can be considered over if the difference between *Validation Loss* and *Training Loss* is "small". It is common practice to check that the magnitude of both *Loss* values is the same, in order to avoid underfitting or overfitting conditions. All other parameters being equal, Training and Validation Losses are reported in Table 3, when the number of maximum epochs is changed in the range [15-50].

Table 3. Training and Validation Losses for different number of maximum epochs.

Maximum epochs	Training Loss	Validation Loss
15	0.0456	0.0745
30	0.0143	0.0404
40	0.0150	0.0405
50	0.0156	0.0266

Besides the gap of 0.03 moving from 15 to 30 epochs, no substantial reduction in Loss is registered, although the 25% increase from test to test. This result is confirmed by the values obtained in terms of Global Accuracy, Accuracy, IoU, BFScore, whose relative variations are negligible. Furthermore, using a too long duration of training there is the risk of overfitting.

Data Augmentation. One of the traditional practice for overfitting elimination is modifying the training dataset images. Data augmentation has been studied according to two different automatic procedures (net_aug e net_aug1). Among the many possible automatic methods, the following have been chosen: reflection along X and Y directions, translation along X and Y direction, of a quantity in the range (-20,+20) pixels, rotation of +/-90° (Table 4).

Table 4. Operations for data augmentation.

	net_aug	net_aug1
X-Reflection	YES	YES
Y-Reflection	NO	YES
X-Translation	[-20; 20]	NO
Y-Translation	[-20; 20]	NO
Rotation	NO	0-90°

Looking at Tables 5 and 6, in both testing datasets (*Original* and *Filtered*), data augmentation remarkably improve the performances of segmentation in terms of GlobalAccuracy IoU and BFscore. As for BFscore behavior is concerned, variability of results can be attributed to specific areas of images, where the composite covers most of the picture. The net without data augmentation is called net_noaug.

Table 5. Data Augmentation indices for the Original dataset.

Net	GA	Accuracy	IoU	BFScore
net_noaug	89.1%	98.9%	80.7%	54.5%
net_aug	+7.89%	-0.32%	+13.1%	+14.4%
net_aug1	+9.73%	-0.02%	+17.5%	+27.2%

Table 6. Data Augmentation indices for the "Filtered" dataset.

	GA	Accuracy	IoU	BFScore
net_noaug	88.3%	97.1%	78.00%	41.6%
net_aug	+7.51%	-2.19%	+12.7%	+22.7%
net_aug1	+8.16%	-1.04%	+11.9%	+11.1%

Training dataset size and balancing. Further improvement techniques refer to the increase of the number of images used for training and the correct balancing among the different conditions of interest. The dataset is composed by 450 images with specimen randomly oriented, also changing the illumination conditions; four lighting set-ups have been realized.

In Fig. 4 the performance parameters are compared, with reference to 5 different nets, tested on 3 different testing datasets. The trained net are called as follow:

- net_86 is the same net as aug_1, previously cited;

- net_214 is trained on 214 images, homogenously distributed in the 4 light conditions;
- net_414 is trained on 414 images, homogenously distributed in the 4 light conditions;
- net_214+86 is trained on 300 images, combining those of net_86 and net_214;
- net_414+86 is trained on 500 images, combining those of net_86 and net_414;

The testing phase has been carried out on the previous datasets *Original* and *Filtered*, plus an additional one (*Additional*), composed by 50 images extracted from the new 450 images acquired.

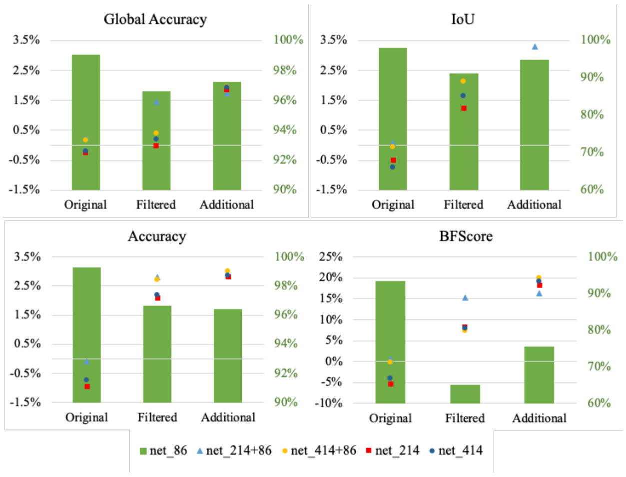


Fig. 4. Effect of dataset size and balancing on performance indices for Original, Filtered and Additional testing datasets

In Fig. 4, net_86 shows the best results among all the nets, with reference to the dataset *Original*; it is quite obvious, being the situation where correspondence between training and testing is maximized. The behavior of the other nets is satisfactory, as well. The comparison with dataset *Filtered* and *Additional* is important, since these datasets are more representative of the in-field operative conditions, taking into account different illuminations and orientation of the specimen. It is important to notice that the score of all nets with increased number of images are better with respect to net_86. The best net on the whole is net_214+86, being the well-balanced one. The net_414+86 suffers of overfitting: increasing the number of images is useless if the balancing is lost.

As an example, Fig. 5 gives a visual representation of the previously described results, highlighting how the improvement of results derives from a balanced use of all the parameters involved in the procedure.

IV. CONCLUSIONS AND OUTLOOK

In this work, a particular step of an inspection strategy by a vision system has been analysed, for the background identification of images related to Fiber Reinforced Thermoplastic Composite components. In facts, this step appears as a crucial one, due to the specificity of the material, which may easily lead to misunderstandings and false negatives during the detection phase of surface and aesthetical defects in the automotive and aeronautical industrial sectors. Two commonly used techniques for the background detection and subtraction have been compared: Object Detection and Semantic Segmentation methods.

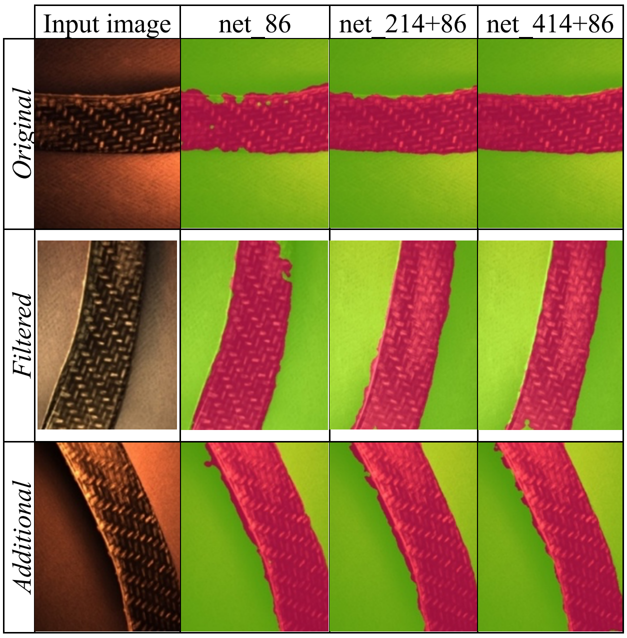


Fig. 5. Qualitative comparison of the obtained results.

Concerning the first one, the following parameters resulted of primary importance: the training dataset enlargement, carried out by means of geometric transformation of input images; the choice of a variable learning rate and the positive overlap range. Although general performance is quite satisfactory, reaching a level of accuracy of 98%, the inability of covering the entire surface of interest leads to makes preferable the use of other techniques. On the other hand, with reference to Semantic Segmentation, the factors determining the wider variability are ascribable to the choice of a variable learning rate, the training duration, the data augmentation and the enlargement and balancing of the training dataset. The qualitative and quantitative analysis carried out allows taking under control the effects of these parameters, on the obtained performances rates, thus leading to set-up the most robust and reliable strategy, depending on the quality levels required during the operating conditions.

Future work will be devoted to the further investigation of the effects of the image quality on both the pre-processing phase of the approach and on the approach itself.

REFERENCES

[1] **Würschinger, H.; Mühlbauer, M.; Winter, M.; Engelbrecht, M.; Hanenkamp, N.**: Implementation and potentials of a machine vision system in a series production using deep learning and low-cost hardware, *Procedia CIRP*, Vol. 90, 2020, pp. 611–616.

[2] **Alonso, V., Daca-Nieto, A., Barreto, L., Amaral, A., Rivero, E.**: Industry 4.0 implications in machine vision metrology: an overview, *8th Manufacturing Engineering Society International Conference, Procedia Manufacturing,* Vol. 41, 2019, pp. 359–366.

[3] **Muñoz, A.; Mahiques, X.; Solanes, J.E.; Martí, A.; Gracia, L.; Tornero, J.**: Mixed reality-based user interface for quality control inspection of car body surfaces, *Journal of Manufacturing Systems*, Vol. 53, 2019, pp. 75-92.

[4] **J. Kunze, T.; Marholtz, M.; Sinapius, M.**: Identification and quantitation of processing parameters controlling the surface quality of carbon fibre-reinforced composites, *Journal of Reinforced Plastics and Composites*, Vol. 35, No. 8, 2016, pp. 638 – 654.

[5] **D'Emilia, G., Di Ilio, A., Gaspari, A., Natale, E., Perilli, R., Stamopoulos, A.G.**: The role of measurement and simulation in additive manufacturing within the frame of Industry 4.0, *Proceedings of the 2019 IEEE International Workshop on Metrology for Industry 4.0 and IoT,* June 2019, pp. 382-387.

[6] **Stamopoulos, A.G., Spitilli, P., D'Emilia, G., Gaspari, A., Natale, E., Di Ilio, A.**: Assessment of the measurments contribution on composites thermoforming processes: A case study of an automotive component, *Proceedings of the 2020 IEEE International Workshop on Metrology for Industry 4.0 and IoT*, June 2020, pp. 299-303.

[7] **Di Nardo G., Fabbiano L., Tamborrino R., Vacca G.**: Automatic defect detection and characterization by thermographic images based on damage classifiers evaluation, *Metrology and Measurement Systems*, Vol. 27, No. 2, 2020, pp. 219-242.

[8] **Amirkhanov A.; Salaberger D.; Kastner J.; Gröller M.E.; Heinzl C.**: Visual Analysis of Defects in Glass Fiber Reinforced Polymers for 4DCT Interrupted In situ Tests, *Computer Graphics Forum*, Vol. 35, No. 3, 2016, pp. 201-210.

[9] **Tabernik, D.; Šela, S.; Skvarc, J.; Skocaj, D.**: Segmentation-based deep-learning approach for surface-defect detection, *Journal of Intelligent Manufacturing*, Vol. 31., 2020, pp. 759–776.

[10] **Racki D., Tomazevic, D., Skocaj D.**: A compact convolutional neural network for textured surface anomaly detection, *IEEE Winter Conference on Applications of Computer Vision*, Lake Tahoe, NV, United States ,March 12-15, 2018.

[11] **Long J., Shelhamer E., Darrell T.**: Convolutional Networks for Semantic Segmentation, Proceedings of the *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 07-12, 2015, pp. 3431-3440.

[12] **Long J.; Shelhamer E.; Darrell T.**: Fully Convolutional Networks for Semantic Segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 4, 2017, pp. 640-651.

[13] **Qian K.**: Automated detection of steel defects via machine learning based on real-time semantic segmentation, *ACM International Conference Proceedings Series, 3rd International Conference on Video and Image Processing, ICVIP 2019*; Shanghai; China; December 20-23, 2019.

[14] **Liu, Y., Yang, Y., Wang, C., Xu, X., Zhang, T.**: Research on Surface Defect Detection Based on Semantic Segmentation, *2019 International Conference on Artificial Intelligence, Control and Automation Engineering (AICAE 2019)*, 2019, pp. 403-407.

[15] **Xie, Y., Zhu, F., Fu, Y.,** Main-Secondary Network for Defect Segmentation of Textured Surface Images, *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2020, pp. 3531-3540.

[16] **Kumarasinghe, A.S.N., Nakajima, A., Motegi, K., Shiraisi, Y.**: A Combination of Deep Learning and Segmentation Algorithms applied to Appearance Inspection Problem, *2019 8th International Congress on Advanced Applied Informatics (IIAI-AAI)*, 2019, pp. 489-494.

[17] **D'Emilia, G.; Di Gasbarro, D.; Gaspari, A.; Natale, E.:** Managing the uncertainty of conformity assessment in environmental testing by machine learning, *Measurement: Journal of the International Measurement Confederation,* Vol. 124, 2018, pp. 560-567.

[18] **D'Emilia, G., Di Gasbarro, D., Gaspari, A., Natale, E.:** About the role of uncertainty assessment in environmental testing, *15th IMEKO TC10 Workshop on Technical Diagnostics 2017 - "Technical Diagnostics in Cyber-Physical Era"*, 2017, pp. 89-94.

[19] **Hao, S.; Zhou, Y.; Guo, Y.**: A Brief Survey on Semantic Segmentation with Deep Learning, *Neurocomputing*, Vol. 406, 2020, pp. 302–321.

[20] **Csurka, G., Larlus, D., Perronnin, F.**: What is a good evaluation measure for semantic segmentation?, *Proceedings of the British Machine Vision Conference*, 2013, pp. 32.1-32.11.