IEEE INSTRUMENTATION & MEASUREMENT SOCIETY

Received 2 February 2023; revised 8 March 2023; accepted 10 March 2023. Date of publication 3 April 2023; date of current version 27 April 2023.

Digital Object Identifier 10.1109/OJIM.2023.3264042

Vision-Based System for Measuring the Diameter of Wood Logs

MARCO CARRATÙ [1] (Member, IEEE), VINCENZO GALLO [1] (Graduate Student Member, IEEE), CONSOLATINA LIGUORI [1] (Senior Member, IEEE), JAN LUNDGREN [2], MATTIAS O'NILS [2], AND ANTONIO PIETROSANTO [1] (Senior Member, IEEE)

[1]Department of Industrial Engineering, University of Salerno, 84084 Fisciano, Italy

[2]STC Research Centre, Mid Sweden University, 852 30 Sundsvall, Sweden

CORRESPONDING AUTHOR: M. CARRATÙ (e-mail: mcarratu@unisa.it).

ABSTRACT Detecting and measuring objects with vision-based systems in uncontrolled environments is a difficult task that today, thanks to the development of increasingly advanced artificial intelligence-based techniques, can be solved with greater ease. In this context, this article proposes a novel approach for the vision-based measurement of objects in uncontrolled environments using a specific type of convolutional neural network (CNN) named you only look once (YOLO) and a direct linear transformation (DLT) process. The case study concerned designing a novel vision-based system for measuring the diameter of wood logs cut and loaded onto trucks. This problem has been occurring in the Swedish forestry industry. In fact, this operation is not carried out with computer vision algorithms because of the high variability of environmental conditions caused by the changing position of the sun, weather conditions, and the variability of truck positioning. To solve this problem, the YOLO network is proposed to locate logs while attempting to maintain a high Intersection over Union (IoU) value for the correct estimation of log size. Furthermore, in order to obtain accurate measurements, the DLT is used to convert into world coordinates the dimensions of the logs themselves. The proposed CNN-based solution is described after briefly introducing today's methodologies adopted for wood bundle analysis. Particular attention is paid to both the training and the calibration steps. Results report that for 80% of cases, the error reported has been smaller than 4 cm, representing only 8% of the measurement, considering a mean log diameter for the application of 50 cm.

INDEX TERMS Convolutional neural networks (CNNs), forest industry, novel measurement application, smart industry, you only look once (YOLO).

I. INTRODUCTION

SEVERAL businesses deal with applications on a regular basis where measuring processes might represent a high expenditure, either in terms of money or infrastructure costs. This can be because the parameters to be measured may be complex or have correlations that are difficult to discern and analyze. One of these businesses is the Swedish forest industry. In fact, in this industry sector every year, hundreds of trucks transport wood logs from harvesters to industries for secondary processing. So, measurements of the raw material are necessary to both decide the process destination (paper production, furniture construction, or others) and to set the selling price.

The sale of wood and its price is subject to a quality assessment. These assessments are regulated by the "Quality Rules" defined in the official documentation [1]. The rules mainly concern the volume of the load to be sold, the species of the individual logs in the load, the amount of rot, and especially the diameter of the individual logs, also because too large logs may be incompatible with the operating ranges of papermaking machinery, tables, furniture, and other processes to which these logs may be subjected.

In the recent past, operators had to manually assess the volume, species, rot level, and any log with a diameter greater than a certain threshold, reported in Table 1, to evaluate any operations carried out on the load before its departure

TABLE 1. Size limits of logs, under bark.

Minimum diameter	50 ±10 mm
Maximum diameter	700 ±10 mm

and estimate the future selling price. The operation is today still human-made but through remote visual inspection.

With the manual assessment, the estimation's goodness depends on the image quality, the operator's experience, and the bundle's specific characteristics, such as the space between the logs, their shape, and the angle at which the photo was taken. These issues require highly skilled operators and represent a significant cost to the companies operating in the sector.

In order to automate the mentioned measurement process, the authors, in collaboration with the Swedish company Biometria [2], [3], responsible for forestry measures in Sweden, studied the possibility of assessing the size of a timber bundle transported by truck and also had the possibility of classifying individual logs according to quality [4].

A. LITERATURE REVIEW

The log identification techniques available in the literature, which is an essential step for subsequent measurement, are based on traditional computer vision and artificial intelligence techniques.

For example, the commercialized product "Logmeter," presented in [5], is a system for log volume estimation based on five fundamental steps. The first step concerns the truck's arrival at the site, which is identified through an RFID system. Then, the images are captured. The whole process takes less than 30 s. Next, the speed and load profile are estimated using various sensors, and data is captured simultaneously using three HD IP Cameras, 11 Laser Scanners, and four LED Floodlights.

After scanning the entire load, the system removes, through a segmentation process, unwanted elements, such as the truck, wheels, cab, drawbars, and platform. The length, diameter, volumes, curvatures, and stacking of the various logs in the bundle are then calculated. The entire solid volume is recreated by identifying the characteristics of each external log.

The approach of this commercial system is primarily devoted to general purpose, as it is programmed more to segment and then calculate the volume of the log load from the shape of the truck trailer. It is, therefore, unable to identify individual logs and measure some of their characteristics as diameter. Another limitation of the approach is the cost and footprint of the hardware. In fact, it requires all the sensors mentioned above, mounted on a gantry under which the truck with the log load must pass. This causes problems at the logistic level, as the instrumentation must be mounted in often remote areas and at the cost of the necessary equipment.

Other solutions did not realize a complete system but presented some phases of the measurement procedure of interest.

As for camera-based measurements, there are two types of approaches in the literature: 1) those based on a single camera and 2) those based on a stereo camera.

In the stereoscopic case, used for acquisition and measurement in three dimensions, the methods are based on the calibration of the camera with two checkerboard target objects of known dimensions [6], [7], [8]. On the other hand, the single-camera approach, which is more similar to the problem at hand, finds several solutions in the literature. One of these is based on Zhang's algorithm [9], which can also correct lens distortions thanks to the separate estimation of the intrinsic and extrinsic parameters of the camera. However, this approach requires multiple shots of the calibration checkerboard target from different angles and distances; this was not possible in this application due to the camera's fixed position. In addition, this algorithm alone does not convert pixel coordinates to world coordinates. The other approach is based on the direct linear transform (DLT) and the homographic projection and has been employed and characterized in the past for industrial applications [10], [11], [12]. Another approach used in the literature does not use the DLT but a sheet of paper side by side with the log to calibrate the system. The proposed approach does not measure cut logs but the diameter of live trees and does not have the robustness of a deep learning-based algorithm [13], [14].

Other algorithms aim only at segmenting the visible logs in an image. For example, in [15], an algorithm for the detection of log edge, based on a combination of the histogram of oriented gradients (HOGs) and local binary patterns (LBPs), is proposed.

The first step merges the two sets of initial detections (HOG and LBP) and applies the Gaussian mixture model to represent the resulting color distribution. This result is then used as a model for calculating the probability map, pixel by pixel, for whether the objects belong to the background or the foreground. Then, the combination of the HOG and LBP algorithm is applied again. This process is then repeated over 15 times.

An Alternative algorithm in the literature is based on a combination of HOG and random forest decision algorithm and can find log edges in bundle images taken from the rear [16], [17]. This method shows better results than the previous one because it uses a machine learning technique, the random forest, instead of simple computer vision techniques.

These methods, primarily based on nonintelligent algorithms, do not require significant computational power to identify logs but are less robust. There is still a dependence on environmental, weather, and lighting conditions. Also, the positioning of the camera can strongly influence the results.

A similar problem regarding cell segmentation has been solved with a generalized framework [18] based on iterative erosion and image segmentation. It can distinguish the

background and the foreground (represented by the cells), counting them and creating precise masks based on edge detection.

Another category of systems developed to analyze logs more robustly involves artificial intelligence algorithms.

These methods [19] have also been reviewed and employed for defect detection and segmentation in wood parts and even wood types. Here, the authors concluded that different data sources must be combined to increase the performance of all the machine learning algorithms, particularly the convolutional neural networks (CNNs, multimodal deep learning).

To summarize, computer vision algorithms are mainly used for these types of industrial applications in the literature, especially in applications of indirect vision-based measurements; these techniques have been around for a long time and have been applied extensively in the industrial environment on production lines but have weaknesses related to their reliance on camera positioning and scene lighting [20], [21], [22], [23].

On the other hand, deep learning can completely change the underlying paradigm of vision-based measurement systems. For the segmentation and classification of objects, CNNs are going to be the best choice for researchers in this field. CNNs are artificial feedforward neural networks inspired by the visual cortex of the animal world. Neurons act as local filters on space and help to identify strong spatial correlations in images. The brain then uses these correlations to classify objects and scenes [24], [25], [26]. These approaches are becoming possible thanks to the huge amount of data collected using the increased computational and storage capacity and due to the development of distributed intelligent sensor networks, which have been crucial to overcome the past limits of deep learning approaches. For example, computer vision uses CNNs for object recognition, classification, and nonoverlapping instance segmentation [27]. However, research related to this type of application has yet to be found in the literature, particularly using only surveillance camera images with quality unsuitable for traditional computer vision in environments with uncontrollable lighting and positioning. Therefore, the type of application can be considered novel.

B. SYSTEM PROPOSAL

Considering the techniques presented, it was decided to design a novel vision-based system for a novel application: the measurement of the diameter of wood logs using nonstereoscopic images. The operation of the proposed system is structured in two steps: the first consists of an object detection and segmentation algorithm designed specifically for the object of interest to be extracted and measured, and the second of converting pixel measurements to real measurements. The second phase is highly mature, robust, and reliable nowadays, to the point that it is used without problems in industrial environments for different purposes, such as quality assessment.

The object detection task has been possible thanks to the measurements provided by harvesters and the company at different times: these enabled the creation of a new object detection dataset, with the object being the individual cut logs loaded on trucks' trailers for transportation. The dataset was created solely with images from surveillance cameras at the cutting locations.

To summarize, this article presents an innovative application for a vision-based measurement system tasked to detect individual logs in a bundle and measure their diameter to check their compliance with the quality rules for sale. More in detail, the contributions of this article are the following.

1) *Innovative Application:* A vision-based measurement system aimed at improving robustness in automatically detecting the object of interest in a highly noisy background characterized by uncontrolled illumination.
2) Development of a novel and extensive dataset for the identification of logs cut and loaded onto trucks in industrial environments.
3) Optimization of the you only look once (YOLO) network based on the case study in order to achieve the best possible detection of each log, aiming to minimize measurement error.
4) Conversion of pixel dimensions to meters for each log using reference targets already in the scene and defined by law.
5) Evaluation of the performance of the proposed system by comparison with other traditional methods based on edge detection algorithms employed in the literature.

This article is structured as follows. The first part will describe the dataset's creation and the YOLO network's training activity. The contactless measurement system will be described in the second part, and its performance in assessing the diameter of trunks from the inference results will be compared with classical computer vision algorithms.

II. THEORETICAL BACKGROUND

State-of-the-art did not present answers that completely solve the object of this article: a methodology for the automatic detection of the object and its measurement, such as the case under study. Indeed, some of these need to be more robust for the application, requiring expensive architecture.

The proposed methodology consists of three steps: 1) the first involves log detection using the neural network YOLO; 2) the second involves DLT calibration; and 3) the third consists in measuring the detected logs. More in detail, the operating framework can be summarized as follows: the image acquired by the security camera in the log cutting and loading area is sent to the neural network. The latter uses the previously trained model to perform an inference that results in both the number of logs detected in the scene and their position and size in terms of pixels. Then the operator selects specific points whose position in terms of real measurements (centimeters) is known to perform a DLT process to retrieve the detected log's diameter.

A. NEURAL NETWORKS CHOICE

CNNs were chosen for object detection; the advantage of this approach is that it does not require any traditional segmentation algorithm, whose performance is significantly worse in changing environmental conditions and the noisy, not uniform background, like the condition of the analyzer scenario. In particular, YOLOv4 architecture has been used since it outperforms concerning other objects detection networks, such as R-FCN, SSD, RetinaNet, and DSSD, both in inference time and in terms of mAP (these are, respectively, better than 100 ms on average and 5%, considering a reference Intersection over Union (IoU) equal to 0.5 [28], [29]).

YOLO network detects specific objects in an image as quickly as possible. This network has become the standard for object detection applications both for its speed of inference and high performance in terms of IoU, precision, and recall. Furthermore, the number of true positives, true negatives, false positives, and false negatives are also monitored to evaluate the goodness of the results. YOLO has also demonstrated great robustness to varying scaling of image compression [30].

Other problems that worsen the object detection performance are overfitting and underfitting, each due to an inadequate dataset or training or network configurations unsuitable for the dataset itself.

B. DATASET FUNDAMENTALS

The critical aspect of deep neural networks is the dataset. This must be as representative as possible of the objects to be classified to obtain a correct and free prediction from false detections. The dataset of an object detection network such as YOLO is very articulated as it must provide a classification for the object found and contain information relative to the position and size of the bounding boxes of object detection on the test images. For example, the Microsoft COCO (Common Objects in Context) dataset contains 91 categories of ordinary objects, with 2 500 000 labeled instances in 328 000 images. ImageNet, on the other hand, has fewer categories but more instances per category.

These can be used in applications that include detecting and classifying images similar to the classes contained in these datasets through a transfer learning approach. However, there are currently no datasets similar to the problem under consideration.

For this reason, it was decided to take advantage of the experience of Biometria operators to create an innovative object detection dataset employable in multiple environmental conditions and with different camera qualities. For this dataset, approximately 1500 images of trucks loaded with logs were used; 600 images were taken on which bounding boxes enclosing the individual logs were drawn from this set of images. A "Label what you can see" approach was chosen for the automatic inference phase of the network. In fact, only the bounding boxes on logs that were clearly visible and not superimposed on other structures have been drawn, such as mechanical arms on the truck, buildings in the background, and other disturbing elements.

FIGURE 1. Example of images labeled from the dataset.

FIGURE 2. Example of incorrect image labeling.

During the image labeling process, attention was also paid to the size of the bounding box drawn on each log present. The size of the bounding box cannot be smaller or larger than the actual size of the log; in that case, an incorrect prediction would be made in the inference phase, as illustrated in Figs. 1 and 2.

The last considered aspect in the manual creation of the dataset has been the overlap between neighboring logs: in order to prevent the network from identifying two overlapping logs as a single log, thus invalidating the size measurement, as can be seen in Fig. 3, only the logs in the foreground were selected, and consequently, all the overlapping logs in the background were not selected.

III. EXPERIMENTAL RESULTS

After planning and choosing the dataset and neural network to be used, the focus was on the training and deployment of the system. First, the optimizations and augmentation of the dataset will be presented, then the different training experiments, and finally, the overall operation of the proposed methodology.

A. DATA AUGMENTATION

After the image annotation process, the focus has been on the data augmentation process. In detail, two sets of modifications were chosen.

TABLE 2. Number of logs detected on different images for two YOLO architectures with three different CLs.

Image	*v1 CL 10%*	*v2 CL 10%*	*v1 CL 20%*	*v2 CL 20%*	*v1 CL 30%*	*v2 CL 30%*
#1	186	228	165	203	140	150
#2	120	201	94	161	76	97
#3	181	220	155	195	144	140
#4	150	175	130	158	123	126
#5	154	181	138	167	130	142
#6	40	54	38	51	37	46
#7	62	106	50	94	45	70
#8	125	148	112	129	100	106
#9	157	187	135	159	116	120
Mean	**130**	**166**	**124**	**146**	**100**	**111**
Std	**47.5**	**52.9**	**41.9**	**45.5**	**37.7**	**32.9**

FIGURE 3. Example of overlapping bounding boxes.

1) Horizontal tilting and distortion [−5°, +5°].
2) Horizontal tilt and rotation [−5°, +5°].

Finally, the images were resized to reduce the computational load on the GPU during training. The dataset thus created can count on a collection of more than 1500 images, after the data augmentation process, for a total of about 150 000 instances. The resulting dataset has been split into three sets: training, validation, and testing, using a 70/20/10 ratio.

The result was the largest existing dataset for detecting cut wood logs from surveillance camera images.

It is necessary to use appropriately sized hardware to manage the dataset's size and meet the computational demands of neural network training. In particular, the two crucial elements for the good management of a neural network are the RAM and the GPU. After careful analysis, 32 GB of RAM was chosen, and an RTX QUADRO 8000 with 48 GB of VRAM was chosen as the video card. The video card is also compatible with CUDA version 7.5, which is helpful for use in conjunction with PyTorch and the cuDNN libraries.

The reasons behind this choice are the large video cards' VRAM availability, which increases the dataset's storage capacity for training, and CUDA cores numbers employed for the tensor calculation of convolutional nodes.

Although the YOLO neural network has a defined basic architecture, it is possible to optimize the object detection results by changing some hyperparameters such as the batch size, the type of data augmentation, the type of activation layers, and the training optimizer. In addition, it is also possible to insert other layers, such as the spatial pyramid pooling (SPP), in the architecture [31]. The number of training epochs, equal to 150, has been chosen as a trade-off between execution time, recognition ability, and false positives.

After the first training, some optimization techniques were applied to improve the results. In the first instance, the focus was on comparing performance in terms of varying the dataset, particularly the type of data augmentation.

Indeed, it is well known in the literature that data augmentation brings many benefits to all Deep Learning algorithms by increasing the number of images available for training by applying random transformations to them [32]. Images are always augmented consistently with the reality of the target; for example, for a picture of a person, it is unrealistic to apply a 180° rotation, while for a log, it is realistic.

The first network (**v1**) was trained using a batch size of 32 and the dataset with horizontal tipping and rotation. In contrast, the second network (**v2**) was trained using a batch size of 32 and the dataset with horizontal tipping and distortion. A number of epochs of 150 were chosen for both. To verify and compare the performance of the two trained networks, the inference was performed on the testing set of 150 images, choosing a confidence level (CL) of 10%, 20%, and 30%. All results in this article are reported in comparison to a baseline. The baseline is the number of logs Biometria operators can detect during work tasks.

A demonstration of how the CL can vary detection performance is shown in Table 2; in particular, this also demonstrates that there is an optimum CL peculiar to each dataset. More in detail in the table, the mean number of detected logs increases using the v2 network thanks to the data augmentation previously described in all the reported CLs.

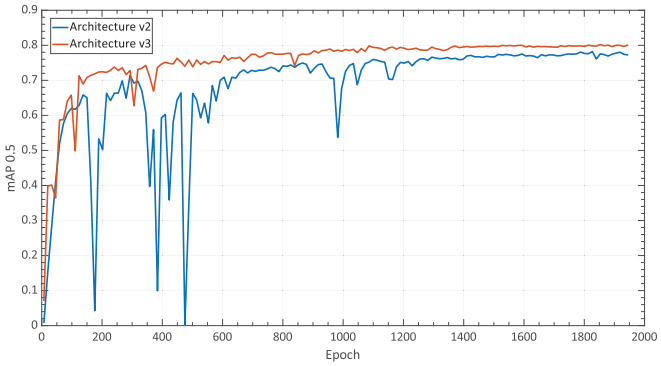


FIGURE 4. Comparison between v2 (orange line) and v3 (blue line) networks (precision and mAP).

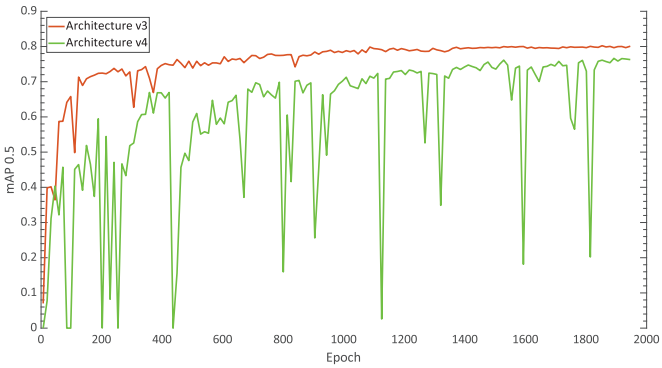


FIGURE 5. Comparison between v3 (orange line) and v4 (green line) networks (precision and recall).

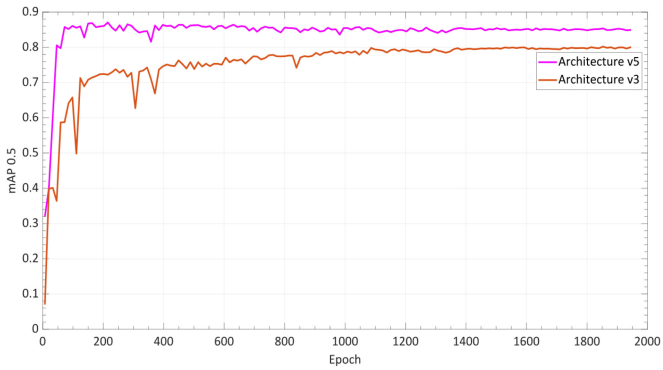


FIGURE 6. Comparison between v3 (orange line) and v5 (blue line) networks (mAP).

Considering that the detection performance of multiscale objects is related to the receptive fields of the network, the use of an SPP layer was evaluated to address the problem of scale variation effectively.

The SPP module is added to improve the feature extraction capability of the network. As a result, SPP offers better performance in terms of accuracy, recall, and mAP. In particular, it has clear advantages in mAP, which is 5% higher than the v2 network.

As well as for data augmentation, the performance improvement resulting from replacing max pooling with the SPP layer is also reported in [31], [33], and [34]

In Fig. 4, it is possible to observe how the mAP of the network with SPP level is higher than the previous network (v2). Moreover, in Fig. 4, an increase in precision is appreciable, motivated by the fact that the **v3** network (with SPP) can recognize more true positives. This aspect is important because although the v2 network can detect more objects categorized as logs, the v3 network detects more "true" logs and generates fewer false positives.

Some valuable strategies to speed up the training (and reach a better solution) were then adopted: using a good activation function (YOLO uses Softmax), using batch normalization, and reusing parts of a pretrained network (i.e., doing transfer learning, using the standard YOLO weights). In addition, another speed-up can be achieved by using an optimizer that is faster than the normal stochastic gradient descent (SGD) optimizer used by the YOLO network.

The training was then carried out using the adaptive moment estimation (Adam) optimizer, as this algorithm has been shown in the literature to be faster than SGD and achieve a higher mAP. In addition, the use of Adam also allows the removal of a hyperparameter (SGD momentum) from the list of unknowns.

These optimizations have been applied in version 4 of the network but were not able to improve the overall mAP. The comparison of the precision and recall parameters using Adam optimizer (**v4** network) is visible in Fig. 5.

The last version of the network (**v5**) has been trained with images that have been squared with zero padding, using the same data augmentation as the previous versions, and reducing the batch size to four. As reported in the literature, square images improved network performance [30].

In Fig. 6, the results of this new training are compared with the results of the v3 network, which particularly stood out in the previous tests. The latter was trained using a dataset consisting of lower resolution images, which increased the mAP metric and, consequently, improved performances.

B. LOG DETECTION

After the different types of training performed on the YOLO network, a set of weights was chosen to check the network's error when identifying logs in an image (network v5). More specifically, a set of 60 images never used by the network (in training or validation) was used to verify the error committed by the network. The operators manually labeled the logs present on each of these images, using the LWYS strategy (presented previously), supplying, therefore, the real bounding boxes expected after the inference process and representing the baseline for the further evaluation of the results. These images were then inferred, selecting only the bounding boxes with a CL greater than 40%. The overlapping bounding boxes have also been identified and discarded in detecting the logs. For this purpose, a nonmaximum suppression (NMS) algorithm has been used: first, the algorithm calculates the areas of the two overlapping or contiguous bounding boxes and detects the overlapping area; then, if the ratio between the intersection and the union of the two bounding boxes is

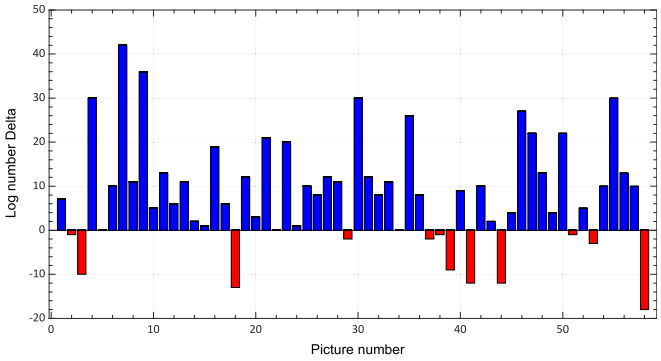


FIGURE 7. Difference between the predicted and baseline number of logs for each bundle image.

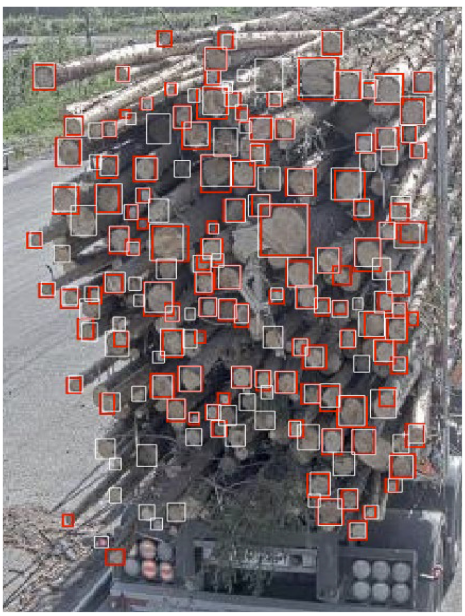


FIGURE 8. Example of detected logs (white bounding boxes) more than the baseline (red bounding boxes).

equal to or greater than 45%, the algorithm selects only the bounding box with the higher confidence (resulting from the inference phase of the YOLO network).

Finally, the number of logs detected by the network was compared with the number of logs detected by the operators, and the delta was estimated as follows:

$$\text{Delta} = \text{No.Predicted} - \text{No.Baseline}. \quad (1)$$

Fig. 7, which presents the results of (1) applied to each log bundle image tested, shows that the network rarely commits a negative error, while it often commits a positive error. This aspect is very reassuring, as it shows that the network can identify more logs than have been reported by the operators in the first place (baseline). In fact, in some images, it is able to detect more than 150% of the baseline, as can be seen in Fig. 8, where the white box indicates the logs detected automatically by the network and the red box the baseline. The images where fewer logs were detected than the baseline had, as the main issue, the presence of backlight sun or snow on the load obstructed the view excessively, preventing the network from detecting all the logs in the scene.

Given the nature of the problem, a network that can detect logs better than it does as a human can is a pro and not a con. For this reason, the "incorrectly detected" logs inside the bundle cannot be considered errors. In this case, the estimated relative error based on the 60 images under review was 1.4%.

C. LOG DIAMETER MEASUREMENT

The actual measurement process consists of two steps: 1) the DLT calibration and 2) the measurement of the detected logs.

The operator is asked to select specific points whose position is known in terms of real measurements (centimeters). With this information, an algorithm calibrates the camera, calculating the homography matrix. Finally, the algorithm converts, for each log detected by the YOLO neural network, the coordinates of the bounding box from pixels to centimeters and estimates the diameter of the log using the maximum dimension of the same.

A camera calibration algorithm based on homographic approximation has been used to measure the diameter of the different logs in a bundle.

Based on the DLT theory, this algorithm relies on the knowledge of the absolute coordinates of some points of the scene being imaged, all belonging to a specific plane called the homography plane. Using these points called "targets," it is possible to obtain the camera calibration parameters, which are mandatory for a pixel coordinate to the world coordinate transformation. If the number of known points is the minimum necessary to obtain the solution, a linear calibration procedure by direct calculation of the solution with a closed formulation has to be used; if it is higher, a nonlinear calibration procedure, which involves the minimization of an objective function, has to be applied [35], [36], [37], [38]. In this way, assuming that the rear of the logs is not too far from the homographic plane, it is possible to obtain the logs' diameter. In particular, the homographic plane used has been one of the rear vertical planes of the trailer on which the logs are loaded.

A set of six targets has been used to calculate the homography matrix, which is necessary to pass from pixel coordinates to world coordinates (in centimeters). These targets were chosen to take advantage of the presence of some reflective hazard triangles installed on the truck's rear bumper. The size of the triangles is standardized and is 16 cm for the base, 17 cm for the larger side of the triangle, and 198 cm for the inner distance between the two triangles. The world coordinates are known, as the distance between the two extreme vertices of the truck's rear bumper, 260 cm, is also available. The targets used are shown in Fig. 9.

The verification of the camera calibration was then carried out using measurements taken from 30 different log bundle images. These measurements were carried out on specific known points present in the back part of the truck (e.g., the

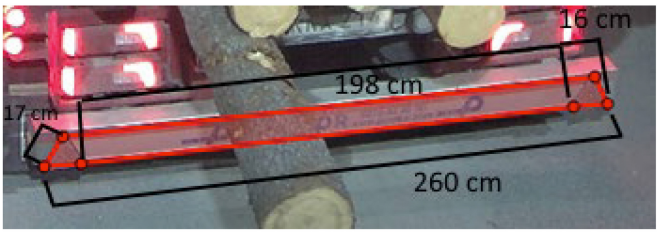
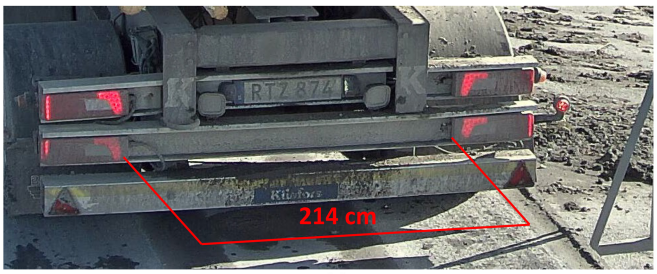


FIGURE 9. Target selection.



(a)

(b)

FIGURE 10. Illustration of the two steps required for log measurement: (a) DLT validation and (b) log measurement.

truck backplate) to assess the reprojection error. As an evaluation of the calibration goodness, one of the measurements made on the length of the upper metal bar in the lower back part of the trailer has been reported in Fig. 10(a).

The upper metal bar had a value defined as 214 cm by legislation; the result of the DLT algorithm was instead equal to 214.4 cm, with a standard deviation equal to 1.5 cm. Therefore, the results have been compatible with the reference value. Measurement results were then compared with those obtained in the field using a special gauge by Biometria operators, as shown in Fig. 10(b). The entire measurement process can therefore be summarized as presented in Fig. 11.

IV. RESULTS COMPARISON

The measurement results obtained using the deep learning neural network for object detection YOLO have been compared with those obtained using classical computer vision

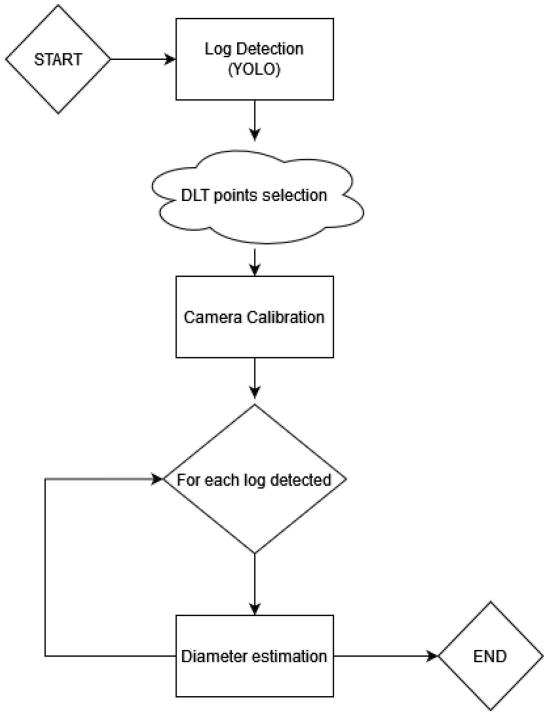


FIGURE 11. Measurement process.

algorithms and other deep learning-based object detection methodologies. The classical computer vision consisted of analyses with edge detection algorithms (e.g., HOG and Canny) combined with a Hough transform and the generalized framework cell segmentation method, as reported in the state of art section. On the other hand, the results obtained with the YOLO network have been compared with the RetinaNet Object Detection network [39], [40]. The choice of RetinaNet was dictated by the fact that it is a single-stage-based Object Detection network, just like YOLO. Therefore, it is characterized by comparable computation times with YOLO.

The comparison results can be seen in Table 3, where a set of 20 images containing 778 individual logs manually counted by a Biometria operator has been used to evaluate the analyzed methods' overall performances in detection capabilities, and in Fig. 12, where the diameter of each log has been manually measured for the first six images of Table 3. In particular, the images were chosen to cover as much as possible all the variability factors in the measurement environment, such as different times of day, different shooting angles, and the presence of rain or snow.

As a first result, Table 3 reports the detection capability of the proposed solution considering both the correct detection index and mean Dice score [41]. Specifically, the Dice score was calculated using (2), where A and B are the baseline and predicted sets of observations and the predicted bounding box. In this case, the intersection between A and B represented the intersection between the areas of the baseline and predicted bounding boxes. In contrast, the union

TABLE 3. Comparison of log detection performances between deep learning and classical computer vision algorithm.

Information		*Circle finder algorithm (Hough transform)*		*Generalized framework cell segmentation**		*Object detection algorithm (ResNet network)*		*Object detection algorithm (YOLO network)*	
Image	**Baseline log number**	Correct detection	Mean Dice	Correct detection	Mean Dice	**Correct detection**	**Mean Dice**	**Correct detection**	**Mean Dice**
#1	**38**	2	0.04	15	0.71	16	0.75	**29**	**0.87**
#2	**34**	9	0.19	11	0.51	13	0.62	**23**	**0.65**
#3	**33**	4	0.01	20	0.69	18	0.71	**24**	**0.72**
#4	**36**	10	0.16	9	0.35	11	0.21	**26**	**0.64**
#5	**32**	4	0.10	16	0.78	13	0.15	**24**	**0.82**
#6	**34**	0	0.00	20	0.38	18	0.35	**30**	**0.91**
#7	**42**	0	0.00	19	0.84	17	0.79	**41**	**0.92**
#8	**41**	10	0.01	20	0.81	12	0.14	**38**	**0.85**
#9	**37**	4	0.03	14	0.74	30	0.68	**35**	**0.74**
#10	**39**	3	0.02	10	0.61	34	0.81	**34**	**0.86**
#11	**45**	12	0.17	21	0.82	25	0.85	**42**	**0.75**
#12	**48**	28	0.18	38	0.68	35	0.68	**46**	**0.92**
#13	**41**	24	0.21	31	0.65	35	0.69	**40**	**0.77**
#14	**47**	22	0.28	26	0.78	28	0.79	**45**	**0.81**
#15	**37**	26	0.17	27	0.74	32	0.87	**32**	**0.85**
#16	**39**	20	0.15	22	0.62	25	0.65	**37**	**0.87**
#17	**41**	5	0.01	31	0.75	28	0.72	**39**	**0.95**
#18	**40**	0	0.00	27	0.79	34	0.78	**35**	**0.74**
#19	**37**	19	0.08	30	0.80	18	0.68	**36**	**0.77**
#20	**37**	22	0.13	28	0.76	25	0.76	**36**	**0.88**
TOTAL	**778**	**224**	**0.10**	**435**	**0.69**	**467**	**0.63**	**692**	**0.82**

between A and B represents the sum of these areas. The average Dice score is obtained by calculating the average of all Dice scores of individual log bundle images

$$\text{Dice} = \frac{2 * |A \cap B|}{|A| + |B|}. \tag{2}$$

The YOLO network outperformed the circle finder algorithm regarding the number of logs correctly detected and the Dice score.

Indeed, the overall performances show that, out of 778 logs, the YOLO network correctly detected 692 logs with an average Dice of 0.82. In comparison, the Circle finder algorithm correctly detected 224 logs with an average Dice of 0.10.

The motivation behind this result can be found in the nonperfect circularity of the logs analyzed by the algorithm. In fact, the circle finder algorithm has as its ideal working conditions those in which the object to be detected in the image is circular and undistorted, such as an ellipse may be. In the present case, the logs have various deformations and jagged irregularities, which lead the algorithm to make errors. On the other hand, the generalized framework cell segmentation performed better for log counting but achieved a lower segmentation performance than YOLO; it detected 435 logs with an average Dice score of 0.69. However, the latter technique has the defect of requiring images already cropped on the ROI, in this case, the back of the truck; this is necessary because the landscape in the background and the other parts of the truck trailer are sources of false positives for the algorithm: in fact, the cell segmentation algorithm does not work well on inhomogeneous backgrounds that are easily confused with the foreground. For this reason, its performance dropped significantly when whole images were used, i.e., not manually cropped on the load. Nevertheless, the automatic cropping of the log bundle is another issue that has challenged traditional computer vision algorithms the most, which is the reason why it was decided to discard this algorithm for measuring log diameter.

For the comparison with RetinaNet Object Detection neural network, training with the same hyperparameters and dataset used for YOLO has been performed. The model was then executed on the same images used for the previous comparisons. The results of this comparison are shown in Table 3. As can be seen from the table, although the results obtained were slightly higher than those obtained with classical image processing algorithms, particularly generalized framework cell segmentation, the performance was still significantly lower than that of YOLO. As already proven by the state-of-the-art, this test demonstrated that the YOLO network is the most suitable among those designed for object detection.

As a final result, Fig. 12 reports the error of logs diameter measurement obtained with the proposed approach involving the DLT process, as presented previously. In addition, the

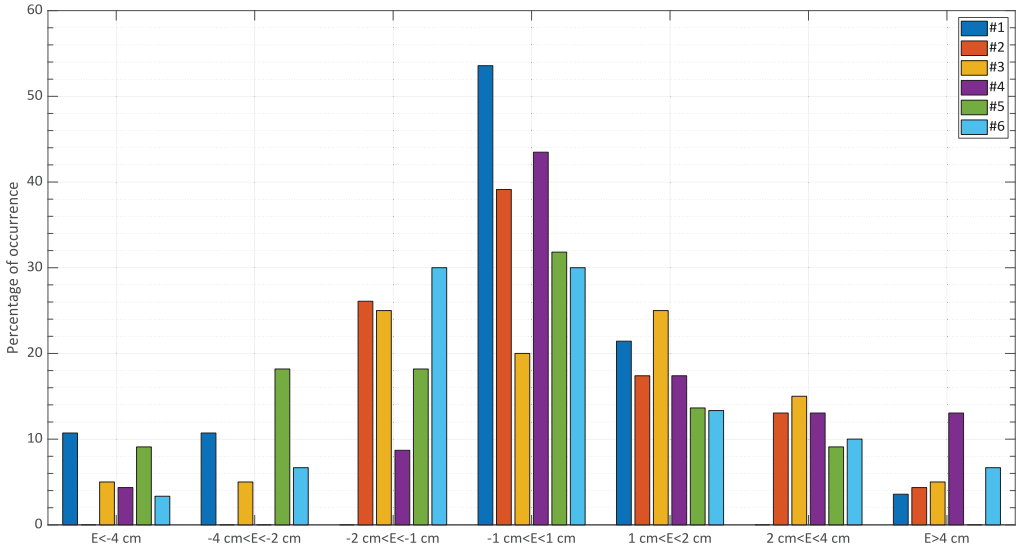


FIGURE 12. Log diameter measurement error using the proposed approach for the first six images of Table 3. The *x*-axis shows the intervals in terms of measurement error, while the *y*-axis shows the percentage of occurrences of each interval for each analyzed image.

performances have been evaluated using as reference the true measurements of the logs themselves measured with a classical gauge by the operators on site.

As shown in Fig. 12, the error obtained is in the range [−4;4] cm for 80% of cases and most of the six images considered. However, considering a mean log diameter of 50 cm, this error represents only 8% of the measurement. Moreover, it should be noted that at least in 50% of cases, the error is even comprised in the range [−2;2] cm. These results are in line with the expectation of the forest industry and, in particular, outperform human-made measurements done through remote visual inspection. Additionally, these results underline an excellent performance of the proposed approach in the correct spatial identification of the bounding box containing the logs.

V. CONCLUSION

In this article, CNN networks and DLT have demonstrated to be strong points of the innovative approach to measurements essential to set the use destination and selling price of timber bundles. Logs' diameters are extracted directly from a bundle without human operators on-site while respecting the Swedish Council's requirements for Measurement and Reporting in the class quality evaluation. The automatic procedure has reduced measurement costs and increased the process's repeatability. The metrological performances resulting from the designed measurement system have been entirely satisfying and beyond the expectations of Biometria. In particular, the proposed technique was the best in log detection, which determined that the measurements were accurate and repeatable.

Future efforts will be devoted to the real-time implementation of suitable hardware for the measurement technique so that the instrument can promptly provide the measurements of the log diameters to the operators on-site to carry out the relevant checks on the passing truck immediately. Challenges with this approach include automating target detection for the DLT procedure and ensuring the timeliness of the entire system in providing measurements. Other efforts will also be dedicated to the refinement of the dataset and the training procedure in order to increase the number of logs detected correctly within the image.

REFERENCES

[1] "TIMBER MEASUREMENT MANUAL, Standard procedures for the measurement of round timber for sale purposes in Ireland." Apr. 2022. [Online]. Available: http://www.coford.ie/media/coford/content/publications/projectreports/TimberMeasurementManual.pdf
[2] "Biometria." Dec. 2021. [Online]. Available: https://www.biometria.se
[3] M. Carratù, C. Liguori, A. Pietrosanto, M. O'Nils, and J. Lundgren, "Data fusion for timber bundle volume measurement," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, 2019, pp. 1–6, doi: 10.1109/I2MTC.2019.8826961.
[4] M. Carratú, V. Gallo, C. Liguori, A. Pietrosanto, M. O'Nils, and J. Lundgren, "A CNN-based approach to measure wood quality in timber bundle images," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, 2021, pp. 1–6, doi: 10.1109/I2MTC50364.2021.9459906.
[5] M. Carlos, *Woodech Forestry 4.0*, Expocorma 2020, Santiago, Chile, Mar. 2020.
[6] B.-S. Park, W. Kim, J.-K. Kim, D.-W. Kim, and Y.-H. Seo, "Iterative extrinsic calibration using virtual viewpoint for 3D reconstruction," *Signal Process.*, vol. 197, 2022, Art. no. 108535. [Online]. Available: https://doi.org/10.1016/j.sigpro.2022.108535
[7] L. Xie, X. Zhang, and D. Tu, "Underwater large field of view 3D imaging based on fisheye lens," *Opt. Commun.*, vol. 511, May 2022, Art. no. 127975. [Online]. Available: https://doi.org/10.1016/j.optcom.2022.127975
[8] Z. Zimiao, X. Kai, W. Yanan, Z. Shihai, and Q. Yang, "A simple and precise calibration method for binocular vision," *Meas. Sci. Technol.*, vol. 33, no. 6, 2022, Art. no. 65016, doi: 10.1088/1361-6501/ac4ce5.
[9] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000, doi: 10.1109/34.888718.
[10] D. L. Giuseppe, L. Consolatina, P. Alfredo, and P. Antonio, "Machine vision systems for on line quality monitoring in industrial applications," *Acta IMEKO*, vol. 4, no. 1, pp. 121–127, 2015, doi: 10.21014/acta_imeko.v4i1.176.

[11] G. Di Leo, C. Liguori, A. Pietrosanto, and R. Lengu, "Uncertainty of line camera image based measurements," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, 2017, pp. 1–6, doi: 10.1109/I2MTC.2017.7969698.
[12] R. Anchini, G. di Leo, C. Liguori, and A. Paolillo, "Metrological characterization of a vision-based measurement system for the online inspection of automotive rubber profile," in *Proc. IEEE Int. Workshop Adv. Methods Uncertainty Estimation Meas.*, 2007, pp. 121–126, doi: 10.1109/AMUEM.2007.4362583.
[13] R. K. Megalingam, V. P. Darla, C. S. K. Nimmala, and K. S. Sankardas, "Computer vision-based measuring method to estimate the diameter of the coconut tree trunk," in *Proc. Int. Conf. Adv. Technol. (ICONAT)*, 2022, pp. 1–6, doi: 10.1109/ICONAT53423.2022.9725999.
[14] B. T. W. Putra, N. J. Ramadhani, D. W. Soedibyo, B. Marhaenanto, I. Indarto, and Y. Yualianto, "The use of computer vision to estimate tree diameter and circumference in homogeneous and production forests using a non-contact method," *Forest Sci. Technol.*, vol. 17, no. 1, pp. 32–38, 2021, doi: 10.1080/21580103.2021.1873866.
[15] C. Herbon, K. Tönnies, and B. Stock, "Detection and segmentation of clustered objects by using iterative classification, segmentation, and Gaussian mixture models and application to wood log detection," in *Proc. 36th German Conf. (GCPR)*, Münster, Germany, Sep. 2014, pp. 354–364, doi: 10.1007/978-3-319-11752-2_28.
[16] Y. V. Chiryshev, A. V. Kruglov, and A. S. Atamanova, "Automatic detection of round timber in digital images using random decision forests algorithm," in *Proc. Int. Conf. Control Comput. Vis.*, 2018, pp. 39–44, doi: 10.1145/3232651.3232667.
[17] D. J. V. Lopes, G. W. Burgreen, and E. D. Entsminger, "North American hardwoods identification using machine-learning," *Forests*, vol. 11, no. 3, p. 298, Mar. 2020, doi: 10.3390/f11030298.
[18] Z. Wang, "A new approach for segmentation and quantification of cells or nanoparticles," *IEEE Trans. Ind. Informat.*, vol. 12, no. 3, pp. 962–971, Jun. 2016, doi: 10.1109/TII.2016.2542043.
[19] M. Kryl, L. Danys, R. Jaros, R. Martinek, P. Kodytek, and P. Bilik, "Wood recognition and quality imaging inspection systems," *J. Sens.*, vol. 2020, Sep. 2020, Art. no. 3217126. [Online]. Available: https://doi.org/10.1155/2020/3217126
[20] R. Anchini, G. Di Leo, C. Liguori, and A. Paolillo, "Metrological characterization of a vision-based measurement system for the online inspection of automotive rubber profile," *IEEE Trans. Instrum. Meas.*, vol. 58, pp. 4–13, 2009, doi: 10.1109/TIM.2008.2004979.
[21] Q. Luo, X. Fang, L. Liu, C. Yang, and Y. Sun, "Automated visual defect detection for flat steel surface: A survey," *IEEE Trans. Instrum. Meas.*, vol. 69, pp. 626–644, 2020, doi: 10.1109/TIM.2019.2963555.
[22] J. Chen, Z. Liu, H. Wang, A. Núñez, and Z. Han, "Automatic defect detection of fasteners on the catenary support device using deep convolutional neural network," *IEEE Trans. Instrum. Meas.*, vol. 67, pp. 257–269, 2018, doi: 10.1109/TIM.2017.2775345.
[23] S. Liu, W. Jiang, L. Wu, H. Wen, M. Liu, and Y. Wang, "Real-time classification of rubber wood boards using an SSR-based CNN," *IEEE Trans. Instrum. Meas.*, vol. 69, pp. 8725–8734, 2020, doi: 10.1109/TIM.2020.3001370.
[24] S. Khan, H. Rahmani, S. Shah, M. Bennamoun, G. Medioni, and S. Dickinson, "A guide to convolutional neural networks for computer vision," *J. Circuits Syst. Comput.*, vol. 18, pp. 629–645, Jan. 2018.
[25] M. Jogin, Mohana, M. S. Madhulika, G. D. Divya, R. K. Meghana, and S. Apoorva, "Feature extraction using convolution neural networks (CNN) and deep learning," in *Proc. 3rd IEEE Int. Conf. Recent Trends Electron. Inf. Commun. Technol. (RTEICT)*, 2018, pp. 2319–2323, doi: 10.1109/RTEICT42901.2018.9012507.
[26] "Basics of the classic CNN." Dec. 2021. [Online]. Available: https://towardsdatascience.com/basics-of-the-classic-cnn-a3dce1225add
[27] J. Son and S. Lee, "Hidden enemy visualization using fast panoptic segmentation on battlefields," in *Proc. IEEE Int. Conf. Big Data Smart Comput. (BigComp)*, 2021, pp. 291–294, doi: 10.1109/BigComp51126.2021.00061.
[28] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," Apr. 2018, *arXiv:1804.02767*.
[29] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
[30] I. Shallari, V. Gallo, M. Carratu, M. O'Nils, C. Liguori, and M. Hussain, "Image scaling effects on deep learning based applications," in *Proc. IEEE Int. Symp. Meas. Netw. (M N)*, 2022, pp. 1–6, doi: 10.1109/MN55117.2022.9887705.
[31] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," in *Computer Vision (ECCV)*. Cham, Switzerland: Springer Int., 2014, pp. 346–361, doi: 10.1007/978-3-319-10578-9_23.
[32] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, p. 60, Jul. 2019, doi: 10.1186/s40537-019-0197-0.
[33] Z. Huang, J. Wang, X. Fu, T. Yu, Y. Guo, and R. Wang, "DC-SPP-YOLO: Dense connection and spatial pyramid pooling based YOLO for object detection," 2020, *arXiv:1903.08589*.
[34] A. Kumar, A. Kalia, A. Sharma, and M. Kaushal, "A hybrid Tiny YOLO v4-SPP module based improved face mask detection vision system," *J. Ambient Intell. Humanized Comput.*, to be published, doi: 10.1007/s12652-021-03541-x.
[35] C. D'Argenio, G. Di Leo, A. Paolillo, and A. Pietrosanto, "Metrological performances of camera self-calibration techniques," in *Proc. IEEE I2MTC*, 2008, pp. 1750–1755, doi: 10.1109/IMTC.2008.4547327.
[36] K. L. A. El-Ashmawy, "Using direct linear transformation (DLT) method for aerial photogrammetry applications," *Geodesy Cartogr.*, vol. 44, no. 3, pp. 71–79, Oct. 2018.
[37] G.-Q. Wei and S. De Ma, "Implicit and explicit camera calibration: Theory and experiments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 5, pp. 469–480, May 1994, doi: 10.1109/34.291450.
[38] A. Agarwal, C. Jawahar, and P. Narayanan, "A survey of planar homography estimation techniques," Centre Vis. Inf. Technol., Hyderabad, India, Rep. IIIT/TR/2005/12, 2005.
[39] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, 2017, pp. 936–944, doi: 10.1109/CVPR.2017.106.
[40] M. Ahmed, Y. Wang, A. Maher, and X. Bai, "Fused RetinaNet for small target detection in aerial images," *Int. J. Remote Sens.*, vol. 43, no. 8, pp. 2813–2836, 2022, doi: 10.1080/01431161.2022.2071115.
[41] K. H. Zou et al., "Statistical validation of image segmentation quality based on a spatial overlap index1: Scientific reports," *Acad. Radiol.*, vol. 11, no. 2, pp. 178–189, 2004, doi: 10.1016/s1076-6332(03)00671-8.

MARCO CARRATÙ (Member, IEEE) received the M.S. degree in electronic engineering and the Ph.D. degree from the University of Salerno, Fisciano, Italy, in 2015 and 2019, respectively.

He is currently an Assistant Professor of Electronic Measurements with the Department of Industrial Engineering, University of Salerno. His current research interests include instrument fault detection and isolation, sensor data fusion, digital signal processing for advanced instrumentation, artificial intelligence for instrumentation and measurement, and real-time embedded systems.

VINCENZO GALLO (Graduate Student Member, IEEE) was born in 1997. He received the B.S. and M.S. degrees (Hons.) in electronic engineering from the University of Salerno, Fisciano, Italy, in 2018 and 2020, respectively, where he is currently pursuing the Ph.D. degree in industrial engineering.

His research interests include the development of deep learning techniques for image-based measurements and the development of novel measurement techniques based on innovative machine learning approaches.

He is currently a member of the IEEE Instrumentation and Measurement Society and the Italian Group of Electric and Electronic Measurement "GMEE," Italy.

CONSOLATINA LIGUORI (Senior Member, IEEE) received the Ph.D. degree in industrial engineering from the University of Cassino, Cassino, Italy, in 1997.

She has been a Full Professor of Electrical and Electronic Measurements with the University of Salerno, Fisciano, Italy, since 2012. From 2016 to 2021, she was a Guest Professor with the STC Research Centre, Mid Sweden University, Sundsvall, Sweden. She is the Co-Founder of three spin-off companies of the UNISA: "SPRING OFF srl," "Metering Research srl," and "Hippocratica Imaging." Her current research interests include smart metering, image-based measurement systems and digital signal processing, measurement characterization, and instrument fault detection and isolation.

JAN LUNDGREN received the Ph.D. degree from Mid Sweden University, Sundsvall, Sweden, in 2007.

After receiving the Ph.D. degree, he worked as an Assistant Professor. Since 2013, he has been an Associate Professor with Mid Sweden University, where he leads a research group focusing on AI-supported sensor systems, including industrial sound and imaging measurements and also medical sound and imaging measurements. His research interests include industrial sound measurements, medical sound and imaging measurements, and industrial soft sensors.

MATTIAS O'NILS received the B.S. degree in electrical engineering from Mid Sweden University, Sundsvall, Sweden, in 1993, and the Licentiate and Ph.D. degrees in electronic systems design from the Royal Institute of Technology, Stockholm, Sweden, in 1996 and 1999, respectively.

He is currently a Professor with the Department of Computer and Electrical Engineering and leads the Research Group in Embedded IoT Systems, Mid Sweden University. His current research interests include design methods and implementation of embedded DNN-based systems, especially in the implementation of real-time video processing and time-series processing systems.

ANTONIO PIETROSANTO (Senior Member, IEEE) was born in 1961. He received the Ph.D. degree in electrical engineering from the University of Naples, Naples, Italy, in 1990.

He was an Assistant Professor of Electrical and Electronic Measurement with the University of Salerno (UniSA), Fisciano, Italy, from 1991 to 1999, where he has been a Full Professor of Instrumentation and Measurement since 2001. He has been the Founder of three spin-off companies with UniSA: SPRING OFF, Metering Research, and Hippocratica Imaging. He has coauthored more than 150 articles in international journals and conference proceedings. His main research activities include the fields of instrument fault detection and isolation, sensors, wireless sensor networks, real-time measurements, embedded systems, metrological characterization of measurement software, advanced system for food quality inspection, and image-based measurements.

Open Access funding provided by 'Università degli Studi di Salerno' within the CRUI CARE Agreement