# A Deep Regression Framework Toward Laboratory Accuracy in the Shop Floor of Microelectronics

Apostolos Evangelidis , Nikolaos Dimitriou , *Member, IEEE*, Lampros Leontaris , *Member, IEEE*, Dimosthenis Ioannidis , Gregory Tinker, and Dimitrios Tzovaras, *Senior Member, IEEE*

*Abstract*—**Deep learning (DL) has certainly improved industrial inspection, while significant progress has also been achieved in metrology with impressive results reached through their combination. However, it is not easy to deploy metrology sensors in a factory, as they are expensive, and require special acquisition conditions. In this article, we propose a methodology to replace a high-end sensor with a low-cost one introducing a data-driven soft sensor (SS) model. Concretely, a residual architecture (R$^2$esNet) is proposed for quality inspection, along with an error-correction scheme to lessen noise impact. Our method is validated in printed circuit board (PCB) manufacturing, through the identification of defects related to glue dispensing before the attachment of silicon dies. Finally, a detection system is developed to localize PCB regions of interest, thus offering flexibility during data acquisition. Our methodology is evaluated under operational conditions achieving promising results, whereas PCB inspection takes a fraction of the time needed by other methods.**

*Index Terms*—**Defect detection, printed circuit board (PCB), residual network (ResNet), smart manufacturing.**

## I. INTRODUCTION

**I**NDUSTRY 4.0 aims to automate manufacturing processes using smart technologies and has greatly benefited from recent advancements in deep learning (DL). The success of DL has been demonstrated by its numerous applications in a variety of different industrial sectors ranging from telecommunications [1] to electronics [2]. However, the large amounts of accurately annotated data needed to train DL models necessitate the deployment of multiple high-accuracy sensors for industrial

process monitoring. The use of high-end laboratory sensors, which are accurate, stable, and robust, is not always a feasible solution due to the high deployment and maintenance cost they incur. Hence it is necessary to rely on less accurate, yet low-cost sensors.

This article is a direct extension of a previous work by Dimitriou *et al.* [3], in which a printed circuit board (PCB) defect detection system is developed. Prior to the attachment of integrated circuits, it is necessary to dispense conductive glue on the PCB's substrate surface, during which defects related to the dispensed glue volume may occur. Identification of such defects is achieved through the use of a modular scanning system and a 3-D convolutional neural network (3DCNN), which regresses the volume of each glue deposit of a PCB. Interestingly, glue volume is accurately estimated even after die attachment when only part of the glue is visible and partial information about its shape is available.

Despite the promising results obtained, this method relies on the use of a laser profilometer, which is expensive, slow, needs to be calibrated, and requires controlled illumination conditions during measurements. These limitations are the main motivation for our work, where we propose to replace the laser profilometer with a single RGB camera and a DL architecture that provides laboratory level accuracy in the shop floor (*in situ*). The replacement of a profilometer with an industrial camera has numerous advantages, as it significantly reduces the cost of automated inspection, making it a suitable choice for *in situ* deployment, whereas both data acquisition and processing are faster, thus notably reducing inspection time. To automate the inspection process using a low-cost sensor, ground truth is acquired in the laboratory by analytically calculating the volume of each sample from its corresponding 3-D point cloud representation and is subsequently used to train a deep network that estimates glue volume from 2-D image data. Additionally, we develop a segmentation and detection system that localizes glue regions within the image irrespective of the PCB's orientation. Finally, we propose a method to deal with label noise, inherently existent due to the finite resolution of the 3-D scans, using only a few reliable measurements. The main novelties of our work are summarized as follows.

   1) We propose a new methodology to replace expensive sensors with *in situ* ones.

2) We provide increased flexibility during data acquisition, as accurate part placement is not necessary.
3) We extrapolate a 3-D variable (volume) from 2-D data (RGB image).
4) A modified residual network for regression ($R^2$esNet) is introduced, and the impact of different choices of depth is examined.
5) We demonstrate the potential of this approach in a semiconductor manufacturing use case.
6) A new benchmark dataset is generated.

The rest of this article is organized as follows. Section II reviews the related work in soft sensors (SS) and metrology methods for defect detection for industrial applications. Section III outlines the examined defect detection use case as well as the proposed methodology. A detailed description of the available data is given and the various deep architectures used are introduced. In Section IV, our methodology is evaluated and the experimental results are presented. Finally, Section V concludes this article.

## II. RELATED WORK

In this section, we give an overview of the most recent work related to our method. Since we develop a SS model to indirectly estimate glue volume from RGB image data and use the estimated value to identify defects, we further elaborate on the recent advancements in the fields of defect detection via deep learning, soft sensor models, and AI.

### A. Defect Detection via DL

The successful application of DL in the industry can be partially attributed to the introduction of several deep convolutional architectures that have revolutionized image classification and have since been used as backbone networks for various tasks. In [4], the authors introduce VGG, a deep convolutional architecture that makes use of multiple layers, while keeping a small, fixed, $3 \times 3$ size for the convolutional kernels, which has since become a standard process. Proposing the use of residual connections, He *et al.* [5] developed residual network (ResNet), achieving state-of-the-art results in multiple image classification datasets. Extending this idea, DenseNet is proposed in [6], by concatenating the feature maps of each layer with the input of every subsequent layer, thus facilitating the gradient flow to the initial layers during backpropagation. To address the increasing demand for mobile and real-time embedded vision applications, Howard *et al.* and Sandler *et al.* [7] and [8] proposed MobileNet, which is based on depthwise separable convolutions to lower the computational cost both during training and inference without significantly affecting performance.

In recent years, DL approaches have been successfully adopted for various industrial applications, mainly focusing on optimizing the production process through the identification of defective parts. In [3], upon which our work is based, the authors consider a PCB defect detection use case and develop a 3DCNN regression architecture called RNet to estimate the volume of a conductive glue deposit before die attachment. In [9], a data-driven fault diagnosis system is developed using a 2DCNN in combination with a parameter-free conversion method to transform 1-D signals into 2-D images, and its effectiveness is demonstrated on three distinct fault diagnosis use cases. Moreover, in [10] a deep convolutional transfer learning network is developed to address the discrepancy between source and target domain during training and testing and is evaluated on three motor bearing fault diagnosis datasets. Lee *et al.* [11] propose the use of a fault diagnosis classification CNN, which enables the association of the output of the first convolutional layer with the structural meaning of the raw data, making it possible to extract information regarding the cause of defects, and test their method on a chemical vapor dataset. To cope with noisy data, a wavelet-inspired soft thresholding approach is adopted in [12] in which the optimal threshold values are learned, and it is applied for the identification of bearing and gear faults in rotating machines. A joint detection and classification scheme is proposed in [13] for steel plate defect inspection through the fusion of multilevel features. Another defect detection application in the hard metal industry is developed in [14], where data gathered from multiple sensors are used for quality assessment. Chen and Li [15] developed another feature fusion method exploiting sparse auto encoders, whereas in [16], the use of a segmentation-based deep architecture that can generalize based on only few training samples is suggested to identify surface crack defects. Finally, both [17] and [18] take advantage of deep learning models to identify welding defects.

### B. Soft Sensors and AI

Another line of DL applications in the industry, involves the development of SS for the indirect monitoring of hidden variables during production. A data-driven approach to estimate hazardous gas concentrations using principal component analysis to decorrelate the input signals along with a deep belief network is proposed in [19]. Furthermore, a soft sensor application is developed in [20], where support vector machines and other predictive models are used to regress key variables in a refinery isomerization process. In a related application [21], a stacked autoencoder soft sensing model is introduced, where a product concentration prediction use case on an industrial debutanizer column process is examined. In [22], a nonlinear finite impulse response model is used to estimate the deflection of a polymeric mechanical actuator, whereas in [23] a semisupervised approach is introduced to exploit unlabeled data. Another semisupervised method is suggested in [24] to estimate the $CO_2$ concentration in an ammonia synthesis process. A spatiotemporal attention-based long short-term memory network is proposed in [26] and is evaluated on an industrial hydrocracking process use case.

## III. PROPOSED METHODOLOGY

In this section, the proposed methodology is outlined. First, an overview of our method is given and the studied use case is described in detail. Next, the *in situ* sensing module under consideration is introduced along with the high-end sensor that is
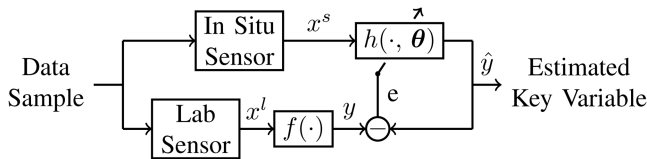
Fig. 1. Schematic overview of the proposed methodology. During training a high-end sensor is used to extract ground truth measurements. Subsequently the low cost sensor along with the trained model are deployed in the shop floor.

replaced. Finally, our R$^2$esNet architecture and the segmentation model for glue detection are described.

### A. Overview

In order to transfer the accuracy and robustness of a laboratory sensor in the shop floor we adopt the following approach: For each data sample, we obtain two distinct representations $x^l$ and $x^s$, corresponding to the measurements acquired using a high-end laboratory sensor and a low-cost *in situ* one, respectively. Subsequently, measurements $x^l$ are used to create annotated samples by analytically estimating the quality variable under consideration through $y = f(x^l)$, where $f(\cdot)$ is considered to be known. Exploiting the existing correspondence between the samples we create a dataset $\{(x^s_i, y_i)\}^N_{i=1}$, which is then used to train a deep network $h(x^s, \boldsymbol{\theta})$ such that

$$h(x^s, \boldsymbol{\theta^*}) \approx f(x^l) \tag{1}$$

where $\boldsymbol{\theta}^*$ are the optimal network parameters estimated by minimizing the empirical cost as

$$\boldsymbol{\theta}^* = \arg \min_{\boldsymbol{\theta}} \sum_{i=1}^{N} L(h(x^s_i, \boldsymbol{\theta}), y_i) \tag{2}$$

and $L(h(x^s_i, \boldsymbol{\theta}), y_i)$ is any valid cost function. A schematic overview of this approach is given in Fig. 1. The deep network predicts the value $h(x^s, \theta) = \hat{y}$, and its parameters are tuned to minimize the prediction error $e = y - \hat{y}$.

### B. Studied Use Case

The growing wave of small electronics products, such as wearables and Internet of Things devices has led to increased demand for small-scale printed circuit boards. A critical stage in the microelectronics manufacturing industry is the dispensing of conductive glue on an liquid crystal polymer (LCP) substrate surface placed by a glue dispensing machine before the attachment of integrated circuits.

The volume of the dispensed glue deposit is a crucial variable that needs to be monitored during production as it directly affects the quality of the produced circuit. Specifically, excessive glue may lead to internal short-circuits, whereas insufficient glue leads to weak die bonding. The current practice for the detection of such faulty conditions is the manual inspection of the PCB, which is both a time-consuming and highly inaccurate process.

The development of a soft sensor model for the monitoring of the dispensed glue volume and the inspection of a part benefits production and is considered crucial toward the automation of
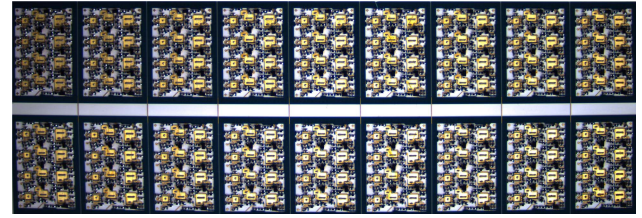


Fig. 2. Example of printed circuit board used for the development of the proposed system. Each PCB consists of two rows and contains 18 circuit modules in total.
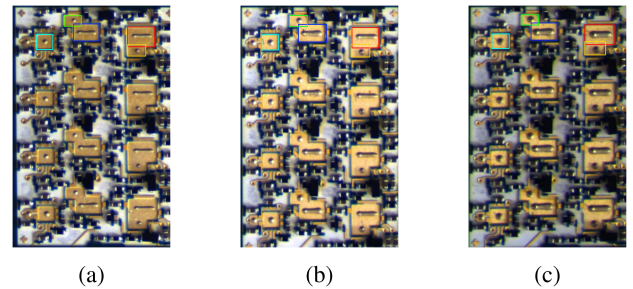


(a) (b) (c)

Fig. 3. Examples of three circuit modules, each corresponding to a different state. The glue dispensed on circuit (a) is insufficient, on circuit (b) normal, whereas on circuit (c) it is excessive.

quality control. The use of a laboratory sensor, such as a laser profilometer [3], even though produces highly accurate and reliable measurements, is expensive and may slow down production. Our proposed methodology aims to transfer the accuracy of a high-end sensor on the shop floor, while simultaneously offering increased flexibility during the acquisition of measurements and highly reducing inspection time and cost. An example of a PCB consisting of 18 identical circuit modules is shown in Fig. 2. Each module contains several glue deposits whose volume needs to be monitored during inspection. There are five different types of glue containers, as can be seen in Fig. 3, which we label A, B, C, D, and E.

### C. Sensing Module

For PCB inspection, the sensing modules under consideration include a modular laser scanning system (high-end sensor) and an industrial RGB camera (low-end sensor). The high-end setting comprises an Optimet Conopoint-10 sensor, a Newport XPS-RL2 motion controller, two linear stages, and the support breadboards. The same zig-zag scanning strategy, as in [3], has been employed to reduce inspection time. Each glue deposit is scanned twice, once with a resolution of 50 $\mu$m and once with 20 $\mu$m. Indicative point cloud representations of the acquired scans can be seen in Fig. 4, from where it is evident that the 3-D geometric structure of the glues is successfully captured. The low-cost setup consists of a Baumer VCXG-201 C.R industrial camera and a Fujinon machine vision CF25ZA-1S lens. The lens' focal length is 25 mm, and, thus, when placed in the right height, the PCB captures a large part of the image compared to the background, thus fully exploiting the setup's potential. All inference computations are performed using a Jetson AGX Xavier, and, thus, the whole system can easily be deployed in
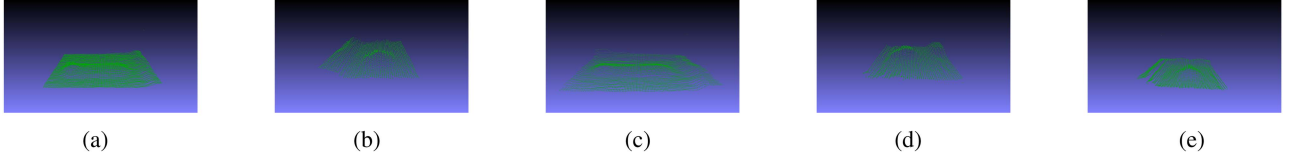
Fig. 4. 3-D scans of glue types A, B, C, D, and E taken with a resolution of 20 $\mu$m. (a) Type A. (b) Type B. (c) Type C. (d) Type D. (e) Type E.



Fig. 5. Proposed setup consisting of a Jetson AGX Xavier and a machine vision camera.



Fig. 6. Illustration of the analytical volume estimation process. Every point $g_i \in G$ is projected onto the substrate plane $p$, and the elementary volume $\Delta v_i$ is calculated. The sum of all $\Delta v_i$ approximates the total glue volume.

the shop floor following the edge computing paradigm. The proposed setup is shown in Fig. 5.

### D. Generating a Dataset

Obtaining an annotated dataset requires expert knowledge and is usually a challenging and time-consuming process, especially in industrial applications, where the sample parts, particularly defective ones, are scarce. Automating the annotation process is highly beneficial as it largely reduces the time and workload needed. To address this common issue, we propose two different glue volume estimation processes, one using the 50 $\mu$m to automatically generate a noisy dataset for training, and another using the 20 $\mu$m to manually generate a reliable and accurate set of measurements for validation purposes.

In both cases, for each regional scan RANSAC [25] is applied to estimate the substrate LCP surface plane. Subsequently, for the low resolution scans the points belonging to the glue deposits are identified as the outliers of the plane's model, whereas for the high resolution scans they are manually cropped. Each 3-D point of the glue point cloud is then projected onto the estimated substrate plane to form a closed surface. Following this process, the resulting 3-D representation is $x^l = G \cup S$, where $G = \{g_i\}_{i=1}^M$ is the glue cloud and $S = \{g_i'\}_{i=1}^M$ the corresponding projected substrate surface cloud. Glue volume is approximated using the formula

$$V(x^l) = \sum_{i=1}^M \|g_i - g_i'\|_2 \, \Delta s_i \qquad (3)$$

where $\Delta s_i$ is the rectangular surface area element and $\|\cdot\|_2$ the standard Euclidean 2-norm of a vector. $\Delta s_i$ depends on the resolution $r$ of the point cloud through the relation $\Delta s_i = r^2$. Equation (3) is geometrically interpreted in Fig. 6. Essentially, the above formula is a discrete approximation of its continuous
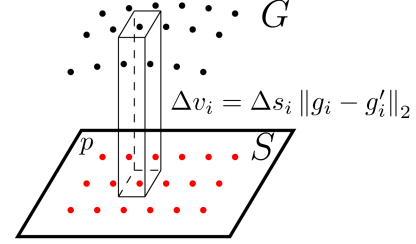
counterpart

$$V = \iint_D g(x, y) \, dx \, dy \qquad (4)$$

where $D$ is the domain of integration, and $g(x, y)$ the height of the glue deposit at any point. The coordinate system chosen is such that the substrate surface plane corresponds to $z = 0$.

Due to the existence of sensor noise during scanning, the imperfect plane fitting and glue-substrate segmentation of the point cloud, and the discretization error introduced in (3), the estimated glue volume is unavoidably corrupted by noise. Let $v_i$ be the true volume value for sample $i$. By construction, we obtain a set of measurements $\{y_i^d\}_{i=1}^N$, estimated from the 20-$\mu$m resolution point clouds and a corresponding set $\{y_i^s\}_{i=1}^N$, estimated from the 50-$\mu$m glue clouds. Both sets of measurements are corrupted by noise such that

$$y_i^d = v_i + n_i^d \qquad (5)$$

$$y_i^s = v_i + n_i^s \qquad (6)$$

where $n_i^s \sim WN(\mu_i^s, \sigma^s)$ and $n_i^d \sim WN(\mu_i^d, \sigma^d)$ are modeled as white noise random processes, where it is assumed that $\sigma^d \ll \sigma^s$ and that $\mu_i^d \approx 0$. After acquiring the PCB images, all glue regions are manually cropped thus obtaining representations $x_i^s$, which are used to form a noisy training set $\mathcal{D}_{\text{Train}} = \{(x_i^s, y_i^s)\}_{i=1}^N$ and a testing set $\mathcal{D}_{\text{Test}} = \{(x_i^s, y_i^d)\}_{i=1}^M$ using the more reliable, dense resolution measurements for evaluation purposes.

### E. Deep Architectures

The developed system consists of a two step process. First, the acquired image is fed into an instance segmentation network that predicts the pixel coordinates of each glue deposit and classifies
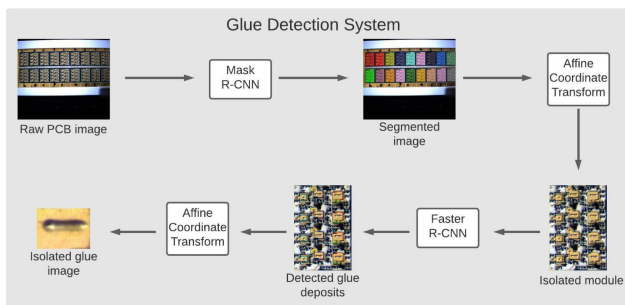
Fig. 7. Schematic representation of the proposed segmentation and detection system. The raw PCB image is fed into the segmentation network, and subsequently the extracted modules are fed into the detection network which regresses a bounding box for each glue deposit.



Fig. 8. Schematic overview of the proposed $R^2$esNet10 architecture. The model consists of four residual blocks, an output, and an input block. Each block sequentially performs convolution, layer normalization, rectification, convolution, and layer normalization.

its type. Subsequently, each detected glue deposit is fed into a regression network that estimates its volume.

*Segmentation and Detection:* During the last few years, computer vision has immensely advanced, especially regarding object detection and instance segmentation tasks. Extensive research in the field has led to the development of two-stage detection methods, like the seminal Faster R-CNN [27] method that is capable of fast and accurate object detection through the utilization of a region proposal network. A variant of faster R-CNN, called Mask R-CNN, is introduced in [28] to perform semantic segmentation on the detected objects.

Since it is necessary to identify all glue deposits within an image, we use Mask R-CNN to first segment the PCB into its constituent modules and Faster R-CNN to detect and classify the glue deposits within each module. For each PCB, the coordinates of its modules are manually annotated. Five of the PCBs are kept for training and one for testing. Due to the small-scale dataset available, we perform heavy data augmentation by applying random rotations, crops, random scaling, and random horizontal and vertical flips during training. Besides increasing the size of the dataset, augmentation also contributes to the development of the desired invariances. Subsequently, for every detected module, we fit the minimum area rectangle to its binary mask and warp the image using the rectangle's four corners. The glue detection faster R-CNN is trained to localize all glue deposits within a module and classify their type. To this end, the pixel coordinates of all glue deposits are manually annotated to create a dataset. The same kind of augmentation is used as before to develop translational, rotational, and scaling invariance. Both networks use a ResNet50 backbone and are pretrained on the ImageNet dataset [29]. Finally, by performing another affine coordinate transform using the detected bounding boxes, the glue deposit images are isolated. A schematic representation of the instance segmentation system is shown in Fig. 7.

*Volume Regression:* To automate the glue volume estimation process, we develop $R^2$esNet, a deep regression network based on the well-known ResNet architecture [5]. ResNets exploit the use of shortcut connections between intermediate layers, and by doing so can define very deep architectures that avoid overfitting and increase the convergence rate during training. As the authors advocate, it is easier to learn residual rather
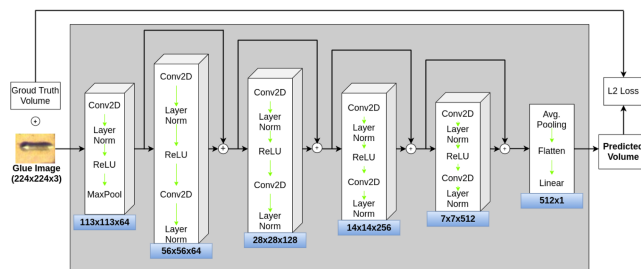
than direct mappings. The final layers of the network consist of a global average pooling layer, along with a 1000-way fully connected layer with softmax for classification.

To perform regression, we modify the existing ResNet architecture by replacing the last classification layer with a single-output fully connected layer and omit the use of the softmax activation function. We experiment with three choices of depth by stacking 4, 8, and 16 residual blocks, which correspond to 10, 18, and 34 layers, respectively. Each network is trained for 90 epochs using Adam with an initial learning rate of $10^{-4}$, which is divided by 10 every 30 epochs. Weights are initialized as in [30] and trained from scratch. All training samples are normalized by subtracting the per-channel mean and dividing by the standard deviation over the whole dataset, whereas the labels are scaled to the interval 0-1. Data augmentation is applied in the form of random rotations, translations, horizontal and vertical flips, as well as brightness, contrast, saturation, and hue jitter. Moreover, due to the small dataset available, we use a mini-batch size of one and, therefore, replace all batch normalization layers with layer normalization. The cost function used is the typical for regression mean square loss (MSE) loss. The schematic overview of the proposed architecture is shown in Fig. 8.

Compared to other deep architectures, $R^2$esNet achieves faster inference due to its reduced size, while the use of residual connections lowers the risk of overfitting. As in most industrial applications, annotated data are scarce, and computational resources are limited, we conclude that $R^2$esNet is well suited for edge deployment. In general, the few hundred samples available would be inadequate to train a deep network and would likely result to the model overfitting the training data. However, due to the controlled image acquisition environment and the relatively small variability of the glue samples, even a small dataset is enough to successfully capture the problem's statistics.

*Error Correction:* As has been described in Section III-D, the available measurements for training are corrupted by white noise, and, thus, it is necessary to employ some regularization strategy to cope with this issue and develop robustness for the trained model. The two sets of measurements for type A glues are illustrated in Fig. 9. It is made evident that besides the random fluctuations present in the sparse resolution measurements, there also exists a deterministic bias directly correlated with the amount of the dispensed glue. The existence of this bias is
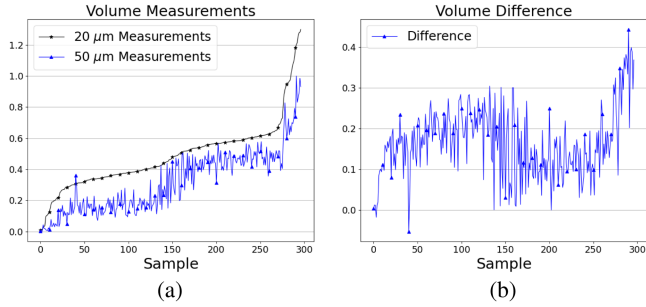
Fig. 9. Plot of $y_i^s$ and $y_i^c$ on the left (a) figure, and their difference $y_i^c - y_i^s$ on the right (b) for type A glues. For better illustration purposes measurements $y_i^c$ are sorted in ascending order.
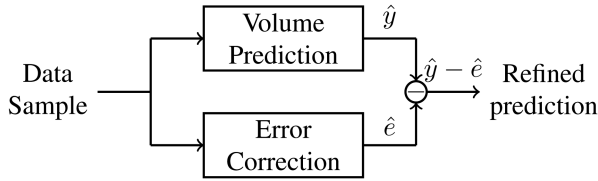


Fig. 10. Schematic representation of the proposed noise compensation methodology.

attributed to the overestimation of the substrate plane's height, as well as the inability of RANSAC to correctly classify the points near the boundaries of the glue. Inspired by this observation, we hypothesize that the existent error is related to several other characteristics, such as glue shape, and can thus be partially predicted from the 2-D input images. Formally, we assume that there exists a deterministic error term $e_i^s$ such that

$$y_i^s = v_i + e_i^s + n_i^s. \qquad (7)$$

In order to refine the network's predictions, we employ a parallel error-correction network that functions as a regularizer and whose aim is to regress the deterministic error from the input image $x_i$. The structure of the proposed model is illustrated in Fig. 10.

To minimize the computational load added through the error-correction network, we choose to employ a R$^2$esNet architecture of depth 10 in every case. A small number of reliable dense resolution measurements acquired from the 20 $\mu$m scans is used for training, during which the parameters of the volume prediction network remain frozen. The same optimization and augmentation strategy is used as before. The small number of available training samples further justifies the choice of a shallower architecture.

## IV. EXPERIMENTAL EVALUATION

In this section, our proposed methodology is evaluated. First, our instance segmentation system is visually evaluated, and its robustness to translations and rotations of the input image is demonstrated. Subsequently, we evaluate the developed regression model R$^2$esNet. The three different choices of depth are compared through the use of tenfold cross-validation, whereas testing set predictions are shown. Finally, inspection times for

**Algorithm 1:** Printed Circuit Board Inspection.

> **Input:** Raw RGB image $x \in \mathbb{R}^{3 \times H \times W}$
> **Output:** Estimated glue volumes $\mathcal{V} = \{v_{ij}\}_{i=1,j=1}^{K,M}$
> $K$ : Number of Modules
> $M$ : Number of glue deposits per module
> 1: PCB Segmentation $(m_i)_{i=1}^{K} \triangleright m_i$: Circuit module $i$
> 2: **for** $i = 1 : K$ **do**
> 3: Glue Localization $(g_{ij})_{j=1}^{M} \triangleright g_{ij}$ Glue deposit $j$
> 4: **for** $j = 1 : M$ **do**
> 5: Volume Estimation $v_{ij} = \text{R}^2\text{esNet}(g_{ij})$
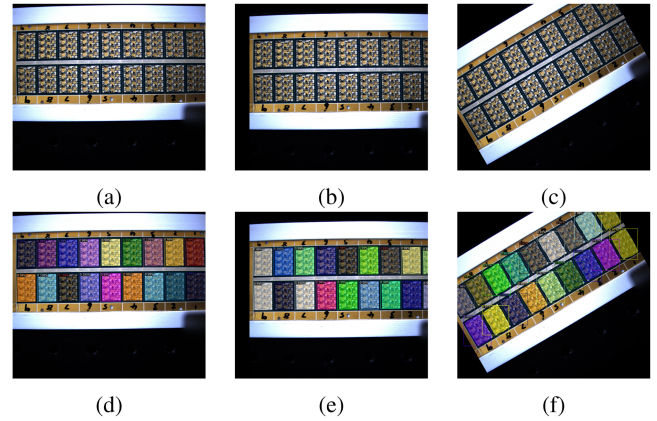> **end for**
> **end for**



Fig. 11. Testing PCB segmentation results. In the first column the original image (a) is segmented (d), in the second the image is translated (b) before segmentation (e), whereas in the third it is rotated (c) and segmented (f).

the six defective and nondefective parts are calculated and presented.

### A. Module Segmentation and Glue Detection

To evaluate the robustness of our segmentation model, we artificially apply random rotations and translations to the acquired input images to simulate the effect of part misplacement. As can be seen in Fig. 11, the developed system successfully segments the PCB into its modules in all three cases. Even in the more challenging cases, where the image is either rotated or translated, all modules are detected despite the partial occlusion near the boundaries of the PCB, which indicates that the system has learned to generalize. Qualitative results of our glue detection system are shown in Fig. 12. Notice that despite the large difference in both illumination and the amount of the dispensed glue deposits, all regions are accurately localized and their type is correctly predicted.

### B. Glue Volume Estimation

The available data for the development of the proposed system consists of six PCBs, each containing 18 circuit modules, as shown in Fig. 2. There are five different types of glue containers annotated A, B, C, D, and E, each corresponding to a different
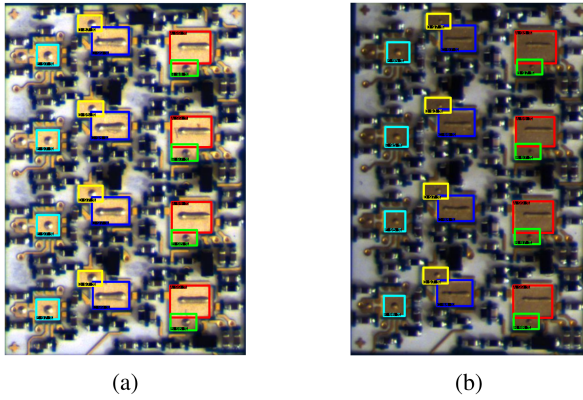
**Fig. 12.** Examples of the detected glue regions within two PCB circuit modules under: (a) regular illumination conditions and (b) low illumination conditions.

shape and amount of glue that needs to be placed on the LCP substrate surface, as is illustrated in Fig 3. Within each circuit module, there are four placeholders for each glue type thus resulting in a total number of 20 placeholders per module. On the top row of one of the PCBs dies have been attached, and hence the total number of available samples per glue type is $5 \times 4 \times 18 + 4 \times 9 = 396$.

To obtain reliable testing results we adopt the following strategy to split the dataset into training, validation, and testing sets: We keep the glues located in the third row of every module for testing, whereas the remaining samples are randomly divided into training and validation samples. This results in a 75%–25% training\validation-testing split, in which the testing set contains glues from all PCBs, and whose volume varies from very little to excessive, that way successfully capturing the whole dataset's statistics. For the same reasons, we use the ground truth volumes estimated using the 20-$\mu$m resolution scans from the first row of each PCB to obtain a clean dataset, which is used to train the error-correction network. That way, the number of the resulting reliable training samples is 99. The average validation error over all tenfolds is shown in Fig. 13. We observe that the deepest model trained consisting of 34 layers performs the worst, whereas $R^2$esNet10 and $R^2$esNet18 perform comparably with $R^2$esNet10 converging faster in most cases.

In order to quantitatively evaluate and compare the developed models, we use the normalized root mean square error (NRMSE) metric as

$$\text{NRMSE} = 100 \times \sqrt{\frac{\sum_{i=1}^{M}(\hat{y}_i - y_i^d)^2}{\sum_{i=1}^{M}(y_i^d)^2}}. \quad (8)$$

To evaluate the effectiveness of $R^2$esNet compared to other popular backbone architectures we perform the same experiments using VGG, DenseNet, and MobileNetV2. For training, all hyperparameters are kept the same as in $R^2$esNet, whereas batch normalization layers are replaced with layer normalization. As in $R^2$esNet, we decrease the number of layers used by stacking eight densely connected blocks for DenseNet, which results to a 20-layer network. Due to its efficient and lightweight

#### TABLE I
##### AVERAGE VALIDATION NRMSE $\times$ 100 %, 50 $\mu m$

| Type | A | B | C | D | E |
|---|---|---|---|---|---|
| $R^2$esNet10 | 19.16 | **34.85** | **26.27** | 56.69 | 29.9 |
| $R^2$esNet18 | 19.57 | 36.97 | 30.94 | **51.72** | **27.98** |
| $R^2$esNet34 | 30.07 | 51.95 | 44.75 | 43.15 | 52.51 |
| DenseNet | **18.98** | 42.75 | 34.11 | 52.38 | 28.26 |
| MobileNetV2 | 20.19 | 44.93 | 31.29 | 55.26 | 33.54 |
| VGG | 48.31 | 59.92 | 50.44 | 65.92 | 45.87 |

#### TABLE II
##### TESTING NRMSE $\times$ 100 % 20 $\mu m$

| Type | A | B | C | D | E |
|---|---|---|---|---|---|
| $R^2$esNet10 | 35.71 | 48.16 | 47.64 | 43.11 | 29.31 |
| ec-$R^2$esNet10 | 18.53 | 30.69 | 28.05 | **28.13** | 26.35 |
| $R^2$esNet18 | 34.60 | 47.74 | 48.04 | 43.13 | 26.20 |
| ec-$R^2$esNet18 | **14.51** | 33.46 | 27.13 | 30.16 | 26.47 |
| $R^2$esNet34 | 37.02 | 46.48 | 50.81 | 46.81 | 25.11 |
| ec-$R^2$esNet34 | 19.94 | **28.39** | 29.45 | 30.86 | **22.62** |
| DenseNet | 35.61 | 52.77 | 38.79 | 58.89 | 46.57 |
| ec-DenseNet | 15.72 | 41.74 | **25.32** | 45.04 | 32.47 |
| MobileNetV2 | 36.68 | 51.76 | 38.11 | 57.30 | 41.27 |
| ec-MobileNet | 17.78 | 39.03 | 27.02 | 43.01 | 32.88 |
| VGG | 39.37 | 36.62 | 46.64 | 40.83 | 39.55 |
| ec-VGG | 24.84 | 32.63 | 35.00 | 31.83 | 33.33 |

#### TABLE III
##### PART INSPECTION TIMES IN SECONDS

| | PCB 1 | PCB 2 | PCB 3 | PCB 4 | PCB 5 | PCB 6 |
|---|---|---|---|---|---|---|
| $R^2$esNet10 | **45.04** | **44.38** | **44.45** | **44.51** | **44.48** | **44.68** |
| ec-$R^2$esNet10 | 55.60 | 55.40 | 55.58 | 55.41 | 55.24 | 55.32 |
| $R^2$esNet18 | 55.57 | 54.17 | 54.21 | 55.22 | 54.23 | 54.39 |
| ec-$R^2$esNet18 | 64.03 | 63.58 | 63.75 | 63.64 | 63.73 | 63.60 |
| $R^2$esNet34 | 75.41 | 72.96 | 72.26 | 71.46 | 71.93 | 72.28 |
| ec-$R^2$esNet34 | 80.63 | 81.28 | 80.81 | 80.82 | 81.07 | 81.09 |
| DenseNet | 51.75 | 51.94 | 51.44 | 51.47 | 50.90 | 50.77 |
| ec-DenseNet | 61.21 | 60.28 | 60.35 | 60.07 | 59.36 | 59.16 |
| MobileNetV2 | 63.63 | 63.49 | 63.30 | 63.38 | 63.28 | 63.08 |
| ec-MobileNetV2 | 90.23 | 89.33 | 88.53 | 88.20 | 88.26 | 88.09 |
| VGG | 110.54 | 110.26 | 11.60 | 111.98 | 111.61 | 111.83 |
| ec-VGG | 188.44 | 190.70 | 188.29 | 190.51 | 190.85 | 190.37 |

implementation MobileNetV2 is kept intact, whereas a VGG network consisting of 11 layers is used. The average validation NRMSE and testing NRMSE are reported in Tables I and II, respectively. Contrary to the usual case, where deeper architectures yield improved results, we observe the opposite, even though ResNets are known to tackle the degradation problem associated with increased depth. During cross-validation, in most cases $R^2$esNet10 produces the best results, whereas $R^2$esNet34 performs worse by a significant margin. In the testing set, irrespective of the choice of depth, all networks perform approximately the same, whereas we observe a significant improvement when using a parallel error-correction network. Compared to other backbone architectures, $R^2$esNet performs better by a small margin in most cases, whereas the classical convolutional VGG network performs significantly worse.

Inspection times for all six PCBs are shown in Table III. We observe that for the most shallow models $R^2$esNet10 and ec-$R^2$esNet10 it takes less than a minute for the inspection of a part, which is a drastic reduction compared to the 20–30 min needed in [3] only for the scanning module to operate, as
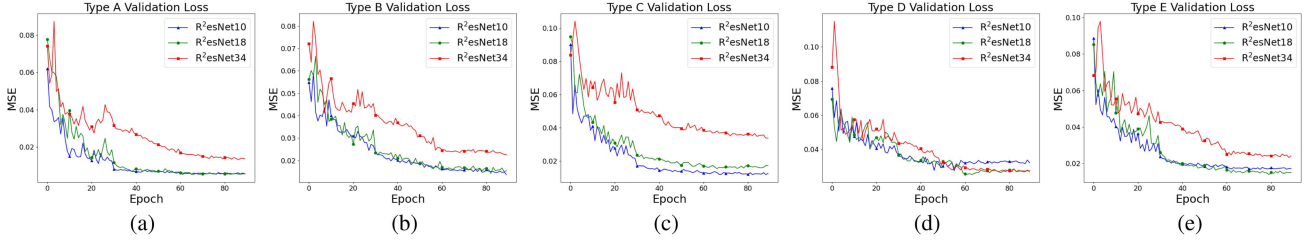
Fig. 13. Average cross-validation loss of the three trained models for all glue types. (a) Type A validation loss. (b) Type B validation loss. (c) Type C validation loss. (d) Type D validation loss. (e) Type E validation loss.
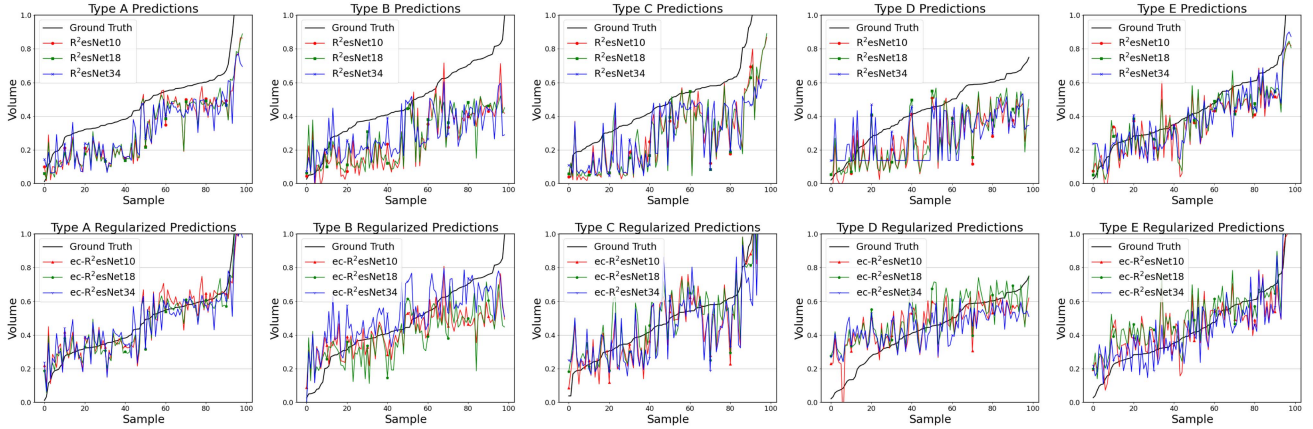


Fig. 14. Testing set predictions for the trained models. On the top row predictions without the use of error-correction are shown, whereas on the bottom the predictions are regularized.

TABLE IV
SCANNING AND INSPECTION TIME COMPARISON IN SECONDS BETWEEN [3] AND OUR METHOD

| Model | Scanning time | Inference time | Total time |
|---|---|---|---|
| RNet, 20 $\mu$m | 2620 | 470 | 3090 |
| RNet, 50 $\mu$m | 1181 | 470 | 1651 |
| ec-$R^2$esNet10 | $\approx 0.17$ | **55.42** | **55.59** |
| ec-$R^2$esNet18 | $\approx 0.17$ | 63.72 | 63.89 |
| ec-$R^2$esNet34 | $\approx 0.17$ | 80.94 | 81.11 |
| ec-DenseNet | $\approx 0.17$ | 60.07 | 60.24 |
| ec-Mobilenet | $\approx 0.17$ | 88.77 | 88.94 |
| ec-VGG | $\approx 0.17$ | 189.86 | 190.03 |

shown in Table IV. It is also noted that we have not optimized our implementation for the Jetson so there is further room for improvement in terms of execution time.

Testing set predictions for all the developed models are shown in Fig. 14. We observe that the predictions follow the increasing ground truth trend, and, thus, offer insight into the amount of the dispensed glue. Without error correction to regularize the outputs, the predicted volume is usually underestimated, which is attributable to the bias existent in the training samples. On the other hand, after regularization, predictions accurately follow the mean. Furthermore, we notice that the performance of the models is directly related to the variance of label noise. Specifically, the MSE is lowest for types A and E, which are the least affected by noise, whereas it is maximal for types B and D.

## V. CONCLUSION

In this article, we propose a methodology to replace expensive laboratory sensors with *in situ* ones, and demonstrate its potential by applying it for the development of a PCB inspection system that only relies on the use of an industrial camera and a Jetson unit. The developed system makes use of our $R^2$esNet architecture to perform glue volume regression along with an error-correction network to cope with label noise. Moreover, a segmentation and detection system that significantly simplifies the data acquisition process was developed. An important contribution of this work is the ability of $R^2$esNet to regress a 3-D geometric quantity from 2-D data.

The effectiveness of our system was demonstrated by the various experiments performed. Despite the challenging nature of the problem addressed, we obtained satisfactory results. Specifically, predictions were accurate enough to facilitate the quality inspection process and hence reduce inspection time, which is a crucial factor toward the optimization of the production process. Another interesting finding is the performance degradation observed for deeper architectures, even though ResNets are specifically build to address this issue.

To further benchmark and evaluate the limitations of the proposed methodology, a potential extension is its deployment in other industrial use cases, where we can explore how well it can generalize in terms of domain adaptation and inferring values outside the nominal ones used during training. Another interesting extension that can further improve the inspection

process is the deployment of the proposed system on augmented reality gear.

## REFERENCES

[1] R. Prashanth, K. Deepak, and A. K. Meher, "High accuracy predictive modelling for customer churn prediction in telecom industry," in *Proc. Int. Conf. Mach. Learn. Data Mining Pattern Recognit.*, 2017, pp. 391–402.

[2] Z. Liu, Z. Jia, C. Vong, S. Bu, J. Han, and X. Tang, "Capturing high-discriminative fault features for electronics-rich analog system via deep learning," *IEEE Trans. Ind. Inform.*, vol. 13, no. 3, pp. 1213–1226, Jun. 2017.

[3] N. Dimitriou *et al.*, "Fault diagnosis in microelectronics attachment via deep learning analysis of 3-D laser scans," *IEEE Trans. Ind. Electron.*, vol. 67, no. 7, pp. 5748–5757, Jul. 2020.

[4] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Representations*, San Diego, CA, USA, 2015.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*. 2016, pp. 770–778.

[6] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2261–2269.

[7] A. G. Howard *et al.*, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[8] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4510–4520.

[9] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5990–5998, Jul. 2018.

[10] L. Guo, Y. Lei, S. Xing, T. Yan, and N. Li, "Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data," *IEEE Trans. Ind. Electron.*, vol. 66, no. 9, pp. 7316–7325, Sep. 2019.

[11] K. B. Lee, S. Cheon, and C. O. Kim, "A convolutional neural network for fault classification and diagnosis in semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 30, no. 2, pp. 135–142, May 2017.

[12] M. Zhao, S. Zhong, X. Fu, B. Tang, and M. Pecht, "Deep residual shrinkage networks for fault diagnosis," *IEEE Trans. Ind. Inform.*, vol. 16, no. 7, pp. 4681–4690, Jul. 2020.

[13] Y. He, K. Song, Q. Meng, and Y. Yan, "An end-to-end steel surface defect detection approach via fusing multiple hierarchical features," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1493–1504, Apr. 2020.

[14] T. Kotsiopoulos *et al.*, "Deep multi-sensorial data analysis for production monitoring in hard metal industry," *Int. J. Adv. Manuf. Technol.*, vol. 115, pp. 823–836, 2021.

[15] Z. Chen and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1693–1702, Jul. 2017.

[16] D. Tabernik *et al.*, "Segmentation-based deep-learning approach for surface-defect detection," *J. Intell. Manuf.*, vol. 31, pp. 759–776, 2020.

[17] P. Sassi, P. Tripicchio, and C. A. Avizzano, "A smart monitoring system for automatic welding defect detection," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9641–9650, Dec. 2019.

[18] Y. Feng, Z. Chen, D. Wang, J. Chen, and Z. Feng, "DeepWelding: A. deep learning enhanced approach to GTAW using multisource sensing images," *IEEE Trans. Ind. Inform.*, vol. 16, no. 1, pp. 465–474, Jan. 2020.

[19] M. G. Xibilia, M. Latino, Z. Marinković, A. Atanasković, and N. Donato, "Soft sensors based on deep neural networks for applications in security and safety," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 10, pp. 7869–7876, Oct. 2020.

[20] S. Herceg, Ž. U. Andrijić, and N. Bolf, "Development of soft sensors for isomerization process based on support vector machine regression and dynamic polynomial models," *Chem. Eng. Res. Des.*, vol. 149, pp. 95–103, 2019.

[21] X. Yuan, B. Huang, Y. Wang, C. Yang, and W. Gui, "Deep learning-based feature representation and its application for soft sensor modeling with variable-wise weighted SAE," *IEEE Trans. Ind. Inform.*, vol. 14, no. 7, pp. 3235–3243, Jul. 2018.

[22] B. Andó, S. Graziani, and M. G. Xibilia, "Low-order nonlinear finite-impulse response soft sensors for ionic electroactive actuators based on deep learning," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 5, pp. 1637–1646, May 2019.

[23] L. Yao and Z. Ge, "Deep learning of semisupervised process data with hierarchical extreme learning machine and soft sensor application," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1490–1498, Feb. 2018.

[24] Q. Sun and Z. Ge, "Probabilistic sequential network for deep learning of complex process data and soft sensor application," *IEEE Trans. Ind. Inform.*, vol. 15, no. 5, pp. 2700–2709, May 2019.

[25] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[26] X. Yuan, L. Li, Y. A. W. Shardt, Y. Wang, and C. Yang, "Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development," *IEEE Trans. Ind. Electron.*, vol. 68, no. 5, pp. 4404–4414, May 2021.

[27] S. Ren *et al.*, "Faster R-CNN: Towards real-time object detection with region proposal networks," *Adv. Neural Inf. Process. Syst.*, vol. 28, pp. 91–99, 2015.

[28] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.

[29] J. Deng, W. Dong, R. Socher, L. -J. Li, K. Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 248–255.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1026–1034.

**Apostolos Evangelidis** received the diploma in electrical and computer engineering from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2020.

He has been a Research Assistant with the Information Technologies Institute, Centre for Research and Technology Hellas, Thessaloniki, since July 2021. His research interests include sparse representations, wavelet analysis, and machine learning.



**Nikolaos Dimitriou** (Member, IEEE) received the diploma and Ph.D. degree in electrical and computer engineering from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2007 and 2014, respectively.

Since July 2014, he has been a Postdoctoral Research Associate with the Informatics and Telematics Institute, Centre for Research and Technology Hellas, Thessaloniki. His research interests include machine vision and deep learning with a focus on industrial applications.



**Lampros Leontaris** (Member, IEEE) received the diploma in electrical and computer engineering from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2016.

He has been a Research Assistant with the Information Technologies Institute, Centre for Research and Technology Hellas, Thessaloniki, since March 2017. His research interests include computational intelligence methods, machine learning, and fuzzy systems.

**Dimosthenis Ioannidis** received the diploma in electrical and computing engineering and the master's degree in advanced communication systems and engineering from the Aristotle University of Thessaloniki (AUTh), Thessaloniki, Greece, in 2000 and 2005, respectively.

Since 2006, he has been a Research Associate with the Informatics and Telematics Institute, Centre for Research and Technology Hellas, Thessaloniki. His research interests include biometrics (gait and activity-related recognition), 3-D data processing, and web semantics.

**Dimitrios Tzovaras** (Senior Member, IEEE) received the diploma and Ph.D. degrees in electrical and computer engineering from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 1992 and 1997, respectively.

He is currently a Senior Researcher Grade A' and the President of the Board with the Centre for Research and Technology Hellas, Thessaloniki. His research interests include visual analytics, 3-D object recognition, search and retrieval, behavioral biometrics, assistive technologies, information and knowledge management, computer graphics, and virtual reality.
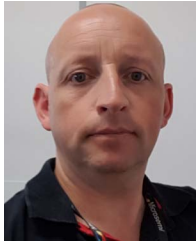
**Gregory Tinker** received the higher national diploma in electronics and communications from Plymouth University, Plymouth, U.K., in 1993.

He was with Meggitt Electronics, Holsworthy, U.K., as a Production Engineer, from 1993 to 1998, and subsequently moved to Mitel, South Wales, U.K, as an Equipment Engineer. The company has changed name several times since and so have the roles he has fulfilled in 20 years of service including SMT Process Engineer and Equipment and Facilities Senior Engineer. He is currently a Senior Process Engineer.