Received November 26, 2021, accepted December 13, 2021, date of publication December 15, 2021, date of current version December 28, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3135973

A Deep-Learning Based Solution to Automatically Control Closure and Seal of Pizza Packages

NÚRIA BANÚS[1,2], IMMA BOADA[1], ANTON BARDERA[1], AND POL TOLDRÀ[2]
[1]Graphics and Imaging Laboratory, University of Girona, 17004 Girona, Catalonia
[2]Vision Department (Research and Development), TAVIL IND, S.A.U., 17854 Girona, Catalonia

Corresponding author: Núria Banús (nuria.banus@udg.edu)

This work was supported in part by the Doctorate Industrial Program through the Catalan Government (2018–2021) under Grant 2018DI026, in part by the Catalan Government under Grant 2017-SGR-1101, in part by the Spanish Government under Grant MICIUPID2019-106426RB-C31, and in part by the Chinese Academy of Sciences President's International Fellowship Initiative under Grant 2021VTB0004.

ABSTRACT Closure and seal inspection is one of the key steps in quality control of pizza packages. This is generally carried out by human operators that are not able to inspect all the packages due to cadence restrictions. To overcome this limitation, a computer vision system that automatically performs 100% inline seal and closure inspection is proposed. In this paper, after evaluating pizza package features, the manual quality control procedure, and the packaging machines of a real industrial scenario, a detailed description of hardware and software components of the proposed system as well as the main design decisions are presented. Focusing on the hardware, line-scan technology and hyperspectral imaging has been considered to ensure that all relevant information can be acquired independently of the pizza brand, topping, and film features. Focusing on the software, this applies a three-phases strategy that, first, applies a set of basic rejection controls; second, identifies the sealing region; and third, prepares the data for prediction through the classification of the pizzas using a deep learning network. This network is one of the software key elements and has been selected after comparing the commercial off-the-shelf (pretrained-dl-classifier-resnet50 from MVTec Halcon) and the custom-developed (ResNet18) architectures designed to automate the accept/reject classification of pizza packages. To train the networks, a classification of pizza package defects, focusing on sealing and closure, and an image-based method able to automatically detect them have been proposed. The system has been tested in laboratory and in real industrial conditions comparing it with the manual scenario and considering three pizza brands with two toppings per brand. From the evaluation, it has been seen that ResNet18 achieves the best results with mean, maximum, and minimum precision values of 99.87%, 99.95%, and 99.74%, respectively. Moreover, our system achieves twice the throughput rate with respect to the manual scenario, with the guarantee that all pizzas are evaluated, which is not possible in the manual scenario due to operator fatigue. The proposed solution can be easily adapted to similar contexts, even considering packages with other shapes.

INDEX TERMS Convolutional neural networks, artificial intelligence classifier, machine vision, food processing industry, image processing, industry automation, pizza packaging, seal inspection.

I. INTRODUCTION

Computer vision systems and image processing techniques have become powerful tools in the food industry. They have been successfully adopted for the quality analysis of meat, fish, pizza, cheese, and bread and for the inspection and grading of fruits and vegetables, among others [1]–[8]. Their capability to replace tedious and subjective human procedures with automatic ones has provided many economic and safety benefits [9], [10]. On the one hand, human presence and human errors have been reduced; on the other hand, efficiency and stability of the processes have increased by promoting faster and more economical inspection. Recently, these systems have been improved with the integration of methods from the field of artificial intelligence, such as deep-learning techniques, that enhance the accuracy and precision of the processes [11]–[15]. In quality control, all these technologies have been combined to automatically classify or detect and discard defective products [16]–[21].

The associate editor coordinating the review of this manuscript and approving it for publication was Hiu Yung Wong.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

Focusing on applied industrial computer vision systems, two main components have to be considered, the hardware which includes illumination system, image acquisition devices or computing elements, and the software which is generally composed of modules that integrate different image processing algorithms as well as deep-learning techniques to determine whether the processed product has to be accepted or rejected [22], [23]. Obviously, all these components are specifically designed to meet the restrictions imposed by the features of the product and the quality requirements [24]–[26]. Moreover, in real scenarios, an extra effort is required to properly synchronize all software components with the industrial hardware to accept or reject products in real time [27]–[29]. In this paper, we have focused on the quality control of pizza packages.

Computer vision techniques have been previously used in pizza production to inspect different pizza components such as the base, the sauce, or the topping [30]–[32]. Du and Sun [33] developed a pizza base shape inspection system that uses image processing techniques to extract the edge of the pizza base through the Fourier analysis of the radius function of the segmented base. This analysis was then used to classify the shapes as acceptable or not by using Support Vector Machines (SVM). Based on color features and SVM, the same authors [34] proposed a method to classify pizzas into five different classes (even spread, acceptable overwipe, reject overwipe, acceptable underwipe, and reject underwipe) with an accuracy of more than 95%. A similar strategy was used in [35] to classify the pizza topping by analyzing different color spaces and using SVM. Many studies have shown that convolutional neural networks methods perform better than SVM-based methods in similar applications [36], [37]. A convolutional neural network (CNN) is a type of deep neural network designed to process grid pattern data, such as images, and automatically and adaptively learn spatial hierarchies of features. A CNN has a multilayer structure where the first two type of layers, the convolution and pooling layers, perform feature extraction, and the third one, the fully connected layer, maps the extracted features into a final output. Since one layer feeds the next with its output, the extracted features can hierarchically and progressively become more complex. Although different CNN solutions are available for industrial applications, its use becomes complex since a training process with specific training datasets is required [38]–[43].

In our case, we focus on the quality control of pizza packages which requires: a serigraphy test to check the correct position of the printings; a traceability test to check the good visibility of date and lot number; and a closure and seal test to ensure the product preservation. This last test is the more complex and the one we have centered [44]. Our aim has been the development of a CNN-based approach to automatically control closure and seal of pizza packages. To reach this objective three main issues that need to be faced.

- First, focusing on seal and closure, there is no classification of pizza package features that determines when a package should be accepted or rejected. Generally, the quality control is performed by human inspection and an analysis of this human procedure has to be carried out to automatically reproduce it and define a suitable training dataset.
- Second, pizza packages have different features depending on the brands and a proper solution should support all of them. Since all information is obtained from product images, the proposed image acquisition system has to ensure that features can be detected regardless of the brand.
- Third, the desired solution has to support real-time inspection of all pizzas on the manufacturing line to guarantee that only valid pizzas reach the end of the supply chain. Therefore, all synchronization issues must be solved to obtain an efficient performance in real scenarios while satisfying time restrictions.

The aim of this paper is to present the solutions that have been proposed to solve each of the aforementioned issues, leading to a computer vision system that efficiently replaces the current manual quality control procedure of a real company. The main contributions of the paper and the related technical challenges can be summarized as follows:

1) A classification of pizza package defects, focusing on seal and closure, with an image-based method able to automatically detect them. To tackle this problem it has been necessary to identify and classify the features that can be considered defects and evaluate the capabilities of the different camera systems to detect these features independently of the package characteristics;
2) A computer vision system able to automatically perform 100% inline inspection of pizza packages while satisfying production cadence. Different configurations have been considered to obtain the one that satisfies the client demands ensuring that all the pizzas are processed by the system. Such a process has been done, first, experimentally in the laboratory and then, in the real scenario;
3) The validation of the proposed solution in a real scenario taking into account different solutions for the CNN implementation as well as a comparison of the proposal with the traditional quality control method considering different datasets. From a technical point of view, it has been necessary to implement a software solution capable to fit different CNN architectures to properly evaluate them. The test in a real scenario ensures a complete test coverage and also an accurate solution for each use case, but makes the implementation more complex.

Note that the issues presented are common to products with similar packaging. Therefore, although our study will be centered on pizzas, the proposed solution can be easily extended to other products.

In Table 1, our proposal is compared with some of the systems previously described considering: the target where the quality control approach is applied; the classification

TABLE 1. **Comparison of some state-of-the-art methods, including our proposal in the last column. The abbreviations correspondence is *Clus* for clustering; $N-N$ for nearest-neighbor; *SVM* for support vector machine; *RBF* for radial basis function; *MLP* for multi-layer perception; *ANN* for artificial neural network; *CNN* for convolutional neural network; *Pre* for pre-trained with ImageNet; *DA* for data augmentation; $L-1-O$ for leave-1-out; *CV* for cross-validation; *F* for fold; *B* for images with constant background; ∗ for candidate defects by a traditional image processing method based on candidate defective regions; *Mono* for monochrome; and *IRAS* for infrared image acquisition system selected by hyperspectral imaging system study.**

	Bairampalli et al. [21]	Du and Sun [33]	Du and Sun [34]	Du and Sun [35]	Darvishi et al. [15]	Banan et al. [18]	Nasiri et al. [20]	Fan et al. [37]	Our method
Quality control	Sensors in moka coffe pot	Pizza bases	Pizza sauce spread	Pizza sauce spread	Sensors	Asian craps species	Eggs (un-washed)	'Fuji' Apples	Pizza packages
Method (Accuracy)									
K-Means Clus.	62%								
Hierarchical Clus.	40%								
N-N Clus.	83%								
SVM Lineal		86.7%		Neglected					
SVM Polynomial		95.0%	96.7%	96.7%					
SVM RBF	87%	98.3%	95.0%	90.0%				87.1%	
MLP ANN					91.3%				
CNN-VGG16(Pre)						100%	96.6%		
CNN-Proposed								96.5%	
CNN-ResNet18(Pre)									**99.1%**
CNN-Halcon									98.7%
Feature extraction	5 specific features	Fourier Transform	Color vision	Color vision	Raw data	Raw image	Raw image	Candidate defects*	Candidate defects*
Input data	Sensors's data	RGB images	RGB images	RGB images	Sensors's data	RGB images B	RGB images	RGB images	**Mono. images IRAS**
Datasets					3-real-world	Only carp species			
Total		120	120	120		409	315	792000 DA	**4558**
Training		60	60	60	70%	347	253	63360 DA	3647
Test		60	60	60	15%	62	62	15840 DA	455
Validation	L-1-O CV				15%	5-F CV	5-F CV	10-F CV	456
Data augm. (DA)	No	No	No	No	No	Yes	Yes	Yes	Yes
Testing scenario									
Laboratory	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Real	No	No	No	No	No	No	No	Yes	**Yes**

accuracy obtained; the method applied; the extracted features; the input data required by the method; the datasets used for training, testing, and validation, and whether or not data augmentation has been applied; and the scenario in which the method has been tested.

Besides this introduction, the paper has been structured as follows. In Section 2, the key elements of the real industrial scenario related to the proposed computer vision system are described. In Section 3, the hardware components of the proposed computer vision system are presented with special attention to the image acquisition system. Then, in Section 4, a detailed description of the software components of the system is given focusing on the considered CNN architectures and the used image processing techniques. Results and Discussion are given in Section 5, where the experiments designed to compare the two CNN architecture alternatives, the manual and automatic performance, and the correctness of the proposed solution are presented. Finally, Conclusions and Future work are given in Section 6.

II. THE REAL INDUSTRIAL SCENARIO

To design a computer vision system to automatically perform the quality control of seal and closure of pizza packages, it is necessary to evaluate the real industrial scenario focusing on: the pizza packages considering different brands and toppings in order to identify the key features; how operators proceed to accept or reject a pizza package in order to reproduce and enhance this procedure; and the machines involved in the packaging process in order to determine the computer vision system location.

A. FEATURES OF PIZZA PACKAGES

The visual attributes of pizza packages has been done considering samples of three brands and six toppings (margherita, ham and cheese, mediterranean, kebab, chicken, and capricciosa) provided by our industrial partner. As it can be observed in Figure 1(a-c), the size, shape, and distribution of opaque and transparent areas of the package depend on the brand, as well as the color and density of the film. Note that $Brand_1$ has a uniform sealing with constant film density and transparency, $Brand_2$ has a dark zone with a dark color printed, and $Brand_3$ has a film with non-constant density and non-transparent sealing. It can also be noted that all packages have common features such as the diameter, which is approximately $30cm$; the shape, which is never completely round; and the width of the sealing, which is not constant for all parts of the pizza. Focusing on the sealing, it can be seen that some parts are common to all brands. These has been labeled as follows (see Figure 1(d)): the single sealing (SS) is the region that covers the main part of the pizza sealing; the opening sealing (OS) is the part that allows the consumer to easily open the package; the double sealing (DS) is the part that connects the SS to the OS. There is also the easy-opening region (EOR), composed of the OS and DS regions, and

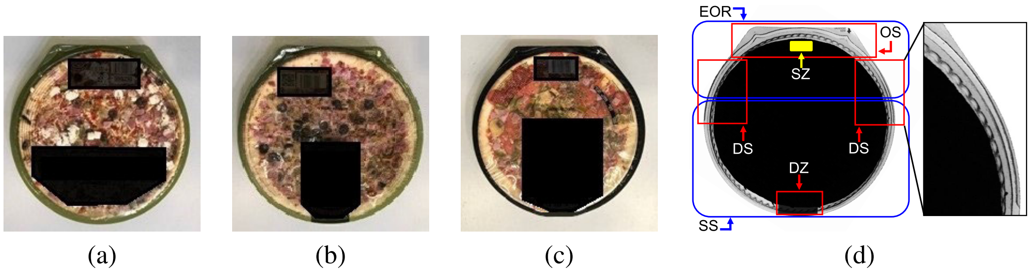


FIGURE 1. (a-c) Packages from $Brand_1$, $Brand_2$, and $Brand_3$, respectively, and (d) parts of pizza package, where SZ is the search zone, OS is the opening sealing, EOR is the easy-opening region, DS is the double sealing, SS is the single sealing, and DZ is the dark zone (only for $Brand_2$).

which is different for all brands analyzed, and the dark zone (DZ), that is the part of the package with a dark color printed, and which only appears in some brands such as $Brand_2$ (see Figure 1(b)).

B. PIZZA PACKAGING QUALITY CONTROL

The quality control in pizza packaging considers three tests: a *serigraphy test* that checks the correct position of the printings; a *traceability test* that checks the good visibility of the expiration date and the lot number; and a *closing and sealing test* that checks the package closure and seal. The package is accepted when all the tests are passed. Our interest has been focused on the most challenging one, the closing and sealing test [44].

After examining the manual procedure and images of the products (acquired with the system described in Section III-A), it has been found that: (i) the area to be inspected is similar for all pizza brands and corresponds to an area between 3 and 8*mm* wide placed about 3*mm* from the package boundary; and (ii) besides the failure in the serigraphy or traceability tests, a pizza is rejected if at least one of the following conditions is given:

- (i) empty package, without pizza (see Figure 2(a));
- (ii) deformed pizza with bumps or scratches (see Figure 2(b,c));
- (iii) overlapping pizzas, i.e., when two whole pizzas appear in a single image and cannot be analyzed (see Figure 2(d)). This case is denoted as *double*;
- (iv) pizza without attached film (see Figure 2(e,f)); and
- (v) large bubbles or large food ingredients in the pizza sealing that pass through it (see Figure 2(g-r)).

On the other hand, there are other conditions that are not reason of rejection such as:

- (i) the marks of the machine that joins the film with the product (see Figure 3(a-e));
- (ii) the humidity generated by the temperature of the pizza at sealing time (see Figure 3(f));
- (iii) the fingerprints that can appear when the product is touched with gloves by the staff (see Figure 3(g,h));
- (iv) some flour over the package (see Figure 3(i)); and
- (v) small bubbles or small food ingredients in the pizza sealing that do not pass through it (see Figure 3(j-o)).

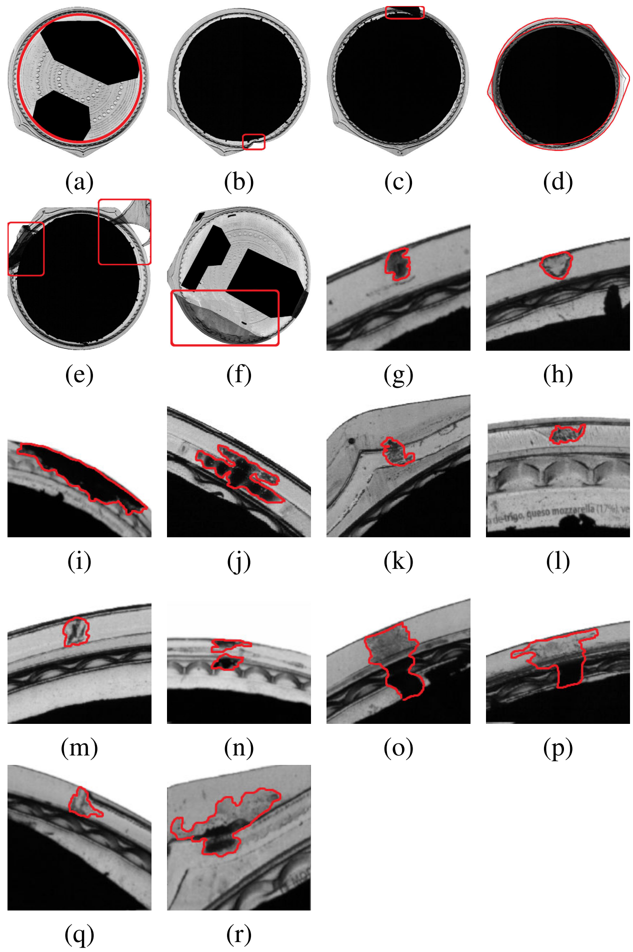


FIGURE 2. Examples of images (or parts of images) corresponding to cases that have to be rejected: (a) empty pizza; (b-c) deformed pizza packages; (d) double, when two whole overlapping pizzas appear in a single image; (e-f) no film attached; (g-r) large bubbles or large food ingredients in the pizza sealing that pass through the sealing.

To better illustrate these cases, parts of scanned images and the real product are shown in Figure 4 for several situations. Some of the cases, such as Figure 3(a-i) and Figure 4 (VI.a / VI.b) and (VII.a / VII.b), are better perceived by a computer than by a human operator. In addition, two new cases, denoted *cut* and *double cut*, and which appear when the system acquires only a part of one or two pizzas, are identified (see Figure 5(a-c)).

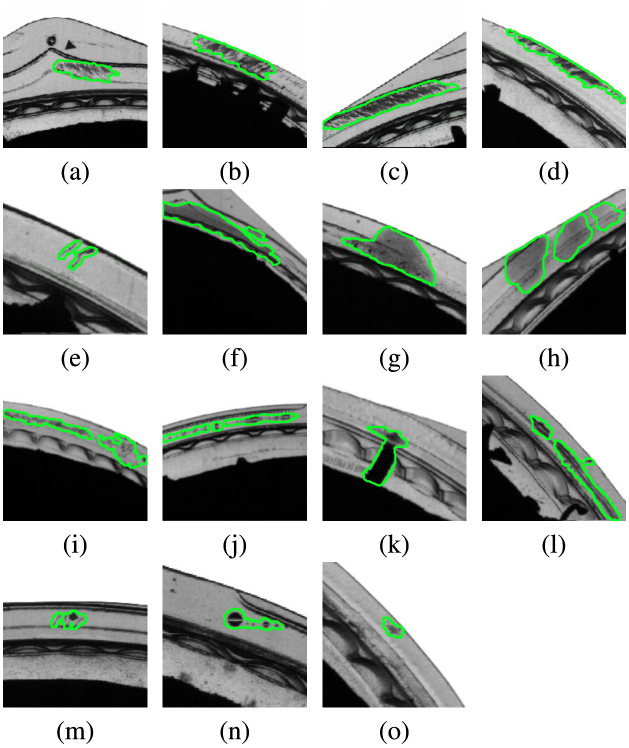


FIGURE 3. Parts of images corresponding to some of the situations that do not imply pizza rejection: (a-e) marks of the machine that joins the film with the product; (f) humidity generated by the temperature of the pizza; (g-h) bold traces; (i) flour on the package; (j-o) small bubbles or small food ingredients in the pizza sealing that do not pass through the sealing.

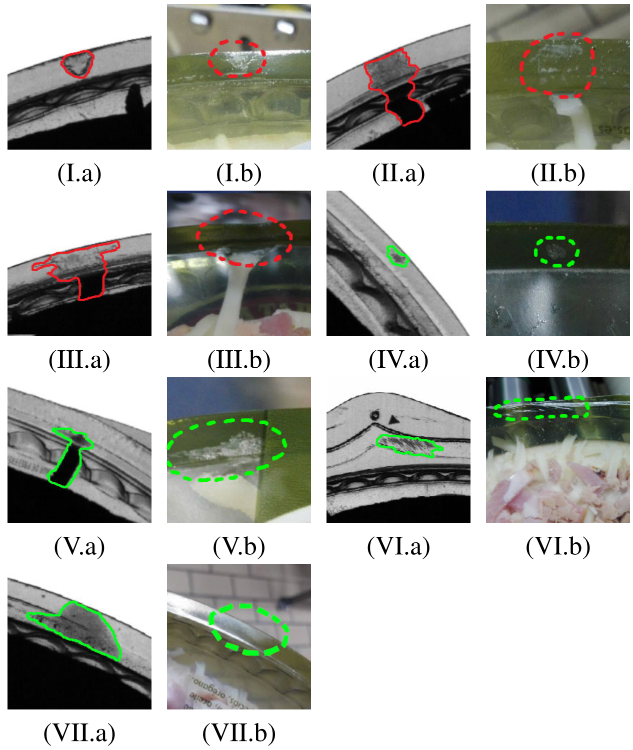


FIGURE 4. Pairs of images where the first (a) corresponds to the processed image and the second (b) to the real image.

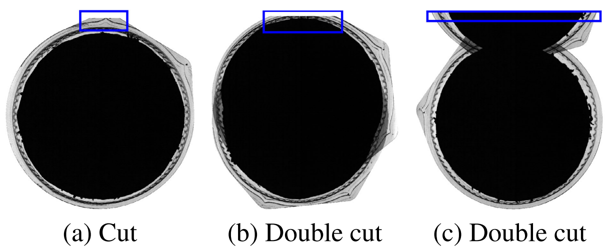


FIGURE 5. Examples of image acquisition inconsistencies: (a) single pizza with a part of the package cropped; (b-c) two overlapping pizzas with a part of the package cropped.

C. PACKAGING MACHINES

Since images of the products are the key of the quality control process it is necessary to determine the best position and composition of its image acquisition system. To define the location, it is necessary to identify the packaging phase machines. These are the thermoforming packaging machine, which provides the lower part of the packages; the tray sealer machine, which seals the upper film; and the cutting machine, which cuts the packages according to the needs of the brand [45]. To meet the requirements of the industrial partner, the image acquisition system has to be installed in the clean room, just after the tray sealer and the cutting machines, and before other control processes. To determine the components of the machine vision system, the following considerations have been taken into account:

- Pizzas are transported on a conveyor belt at $1.0m/s$. No restriction on position or orientation are imposed which may lead to overlapping between products.
- Images must be acquired and processed in less than $300ms$ to satisfy production cadence requirements.
- The inline inspection system has to process the whole upper surface of the pizza package and hence a high-speed system that provides high-resolution images is required [27]–[29].

With all these information in mind, we propose a computer vision system composed of hardware and software components presented in next sections.

III. COMPUTER VISION SYSTEM HARDWARE

To describe the hardware of the proposed computer vision system, first we will present the image acquisition system and then a global view of all the system configuration.

A. IMAGE ACQUISITION SYSTEM

The analysis of the real scenario imposed a set of restriction that can be satisfied with line scan technology [46]. This acquires two-dimensional images line by line while, in our case, the pizza package moves perpendicular to the fixed camera. Moreover, to ensure that all relevant information is always extracted from the acquired images, regardless of brands and toppings, hyperspectral imaging has also been considered. Hyperspectral imaging integrates conventional imaging and spectroscopy to simultaneously obtain spatial and spectral information about an object. The data obtained is generally arranged as a three-dimensional cube, called a hypercube, with two spatial dimensions and one

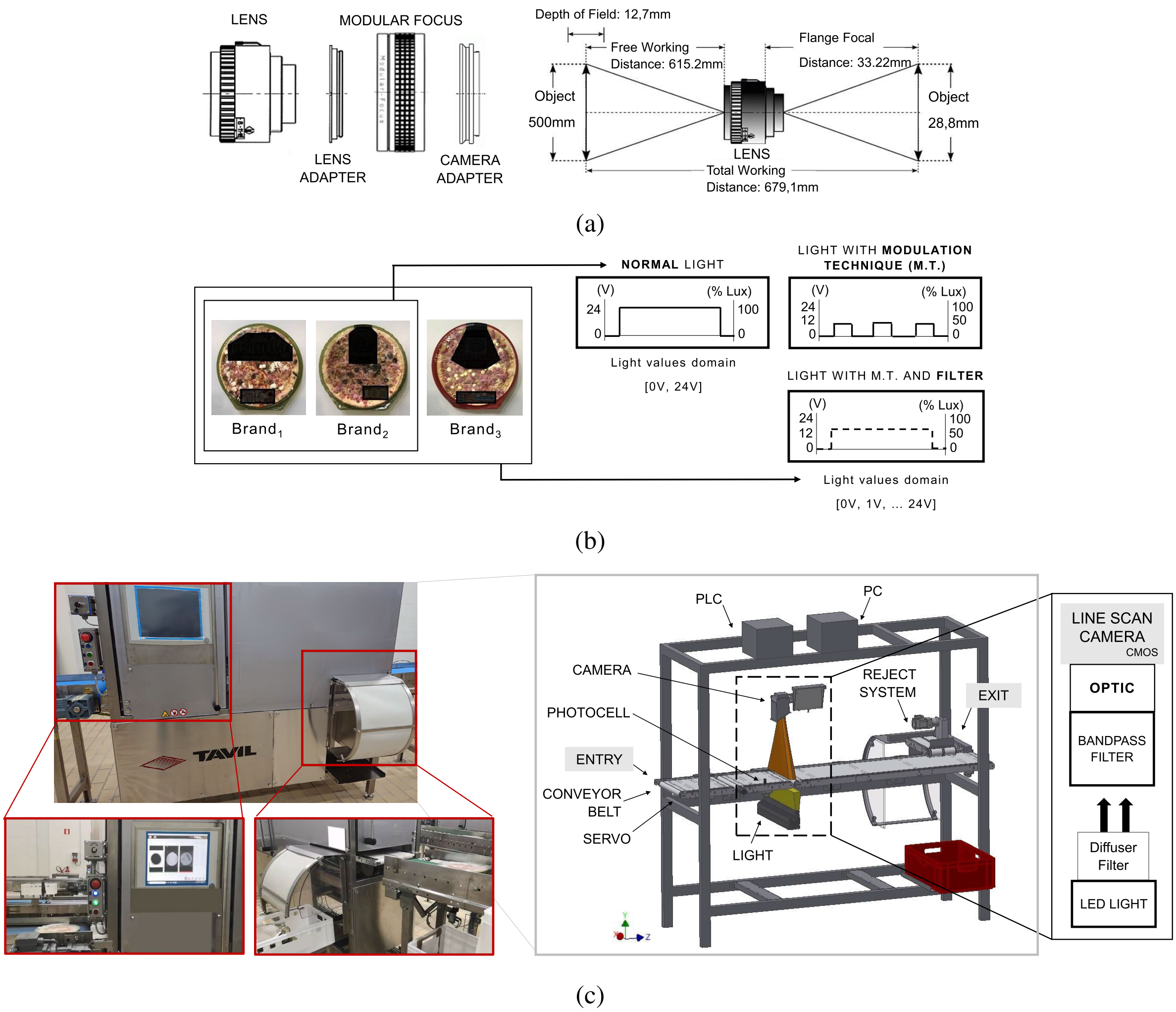


FIGURE 6. **(a) Lens details of the image acquisition system; (b) modulation technique with the filter applied to support the processing of all brands; and (c) main hardware components of the proposed system to control pizza packaging, and some views of the real scenario.**

spectral dimension. This hypercube corresponds to a stack of images of samples where images are acquired at a regular interval in a wavelength region [47]. As a result, and compared with traditional machine vision technology, the hyperspectral image contains a large amount of information which provides a more reliable object characterization. However, since hyperspectral images of full wavelength regions have excessive and redundant information, a selection process is applied to select the relevant wavelengths. Note that a hyperspectral imaging system can be developed by attaching a tunable filter with a monochrome camera [48].

In our system, it has been installed a line scan monochrome CMOS 4K camera with a standard GigE Vision interface [49] ($80Mb/s$) and TurboDrive technology [50]. To determine the proper wavelengths and remove all other unnecessary wavelength radiations [51], different tests were carried out in our laboratory with the different pizza brands and toppings. Of the possible wavelengths, ranging from $400nm$ to $1100nm$, the selected one was of $850nm$. In this study, the appropriate illumination system was also determined, obtaining a final camera configuration that works with a $35-30mm$ F4 M36 $\times$ 1 lens, a modular focus, adaptors and an Infrared Bandpass Filter $850nm$ and a linear IR 850 led light with a diffusion filter positioned behind the product that acts as a backlight. The details of the camera lens are given in Figure 6(a), where the speed, noise, and depth of field determine the acceptance limit values of exposure, gain, and aperture.

Moreover, to better inspect closure aspects such as closure dimensions, sealing features, and liner integrity for all brands and toppings with minimal time and without quality loss, a pixel binning process is applied to reduce the acquired images of 4096 pixels (4K) of $7.04\mu m \times 7.04\mu m$ pixel size to 2048 pixels (2K). To support the different features of the film packages, three acquisitions with different exposures are made for each acquisition line. In addition, to support the non-uniform density of the film and the non-transparent colors of the sealing, a technique based on pulse-width-modulation with a filter is applied to the light to obtain a range of intensity values instead of an on/off behavior (see Figure 6(b)).

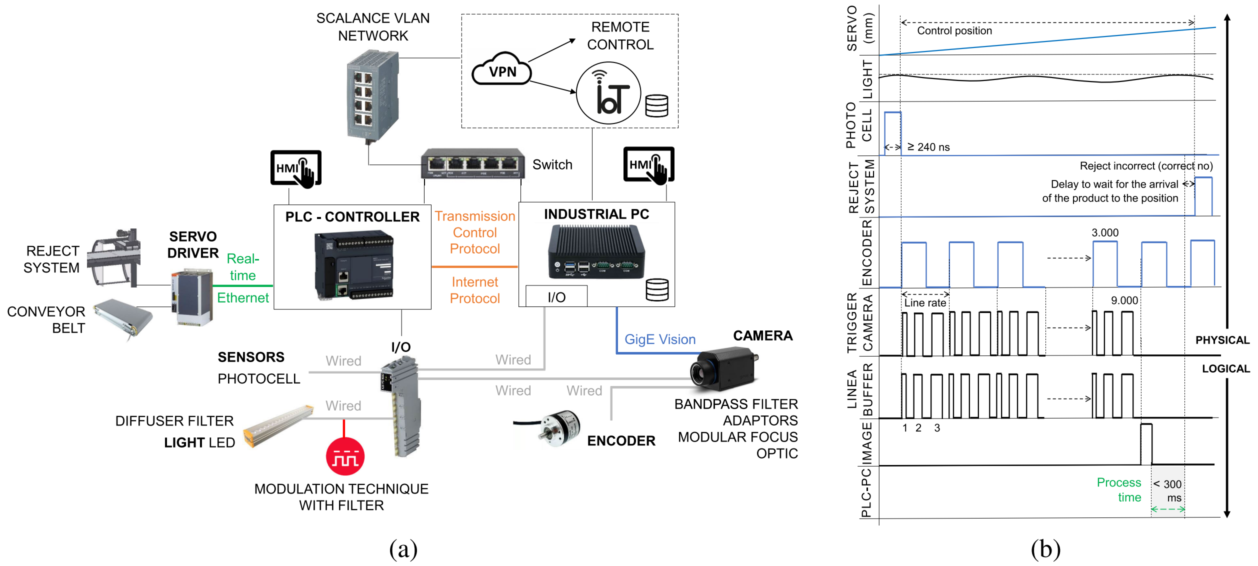


FIGURE 7. (a) Details of the communication between the components of the proposed system, and (b) physical and logical synchronization of the main components of the system.

B. GLOBAL SYSTEM CONFIGURATION

The machine vision system not only involves the described image acquisition system, but also different hardware components as illustrated in Figure 6(c), where a global view of the proposed system is presented. The system keeps the camera fixed while the pizza package is moved through a conveyor belt. A servo connected to a Programmable Logic Controller (PLC) controls the speed of the conveyor belt that transports pizzas from the entrance to the exit. The servo with the photocell information of the pizza entrance controls the position of the products and when the rejection system has to act. An encoder returns feedback about the pizza position which, together with the feedback of the photocell, triggers the line scan monochrome camera. The acquired image is processed by the software and is accepted or rejected as appropriate. To obtain high-quality images, an extremely uniform motion is required, which becomes a critical factor in the design. To do this, a synchronization has been carried out using an encoder in order to avoid image distortions.

Details of how all components are communicated are presented in Figure 7(a). The controller I/O is wired to a photocell, to the light, to the industrial PC I/O, and to the camera, and the encoder is wired to the camera. The industrial PC and the PLC communicate via Transmission Control Protocol/Internet Protocol. GigE Vision is used to communicate the camera with the PC, and real-time Ethernet is used to communicate the PLC and the Servo Driver. In addition, a detailed illustration of the physical and logical synchronization between the main components of the system is provided in Figure 7(b). It can be observed that the servo controls the position while the light remains constant. The photocell is active for a period of 240ns, that is when the acquisition process starts using the encoder information. Note that for each encoder pulse there are three camera triggers with different exposure times (9000 lines for 3000 encoder pulses). Once the images are obtained, time restrictions impose the process and, once the result is known, the reject system acts accordingly. If the product reaches the rejection system position without a known result, it is rejected.

IV. COMPUTER VISION SYSTEM SOFTWARE

All the hardware components that compose the machine vision system are controlled by a vision control software that processes the pizza images and determines how the components have to proceed to reject or accept the products. In the real scenario, this software is a custom-developed software created in our company for managing the different industrial vision systems that coexist in an industry [52].

Besides image processing strategies, a key component of the computer vision software is the applied CNN architecture. To select it, off-the-shelf and custom-developed architectures have been considered. Due to the importance of this selection, previous to the description of the software, we are going to present them.

A. CNN DESIGN ALTERNATIVES

To set the parameters of CNN-based solutions a computationally intensive training process is required. However, using transfer learning such a cost can be reduced since less labeled images and resources are needed [53]. In our case, to carry out the training, a dataset with 4558 images of 224 × 224 pixels with sealing information in their center has been collected and labelled. Then, the two following alternatives have been considered to design the CNN:

- The commercial off-the-shelf *pretrained-dl-classifier-resnet50.hdl* provided by MVTec Halcon [38]. This network has been trained using pictures from industrial applications and, using the transfer learning technique, it can be retrained for a new specific task with potentially

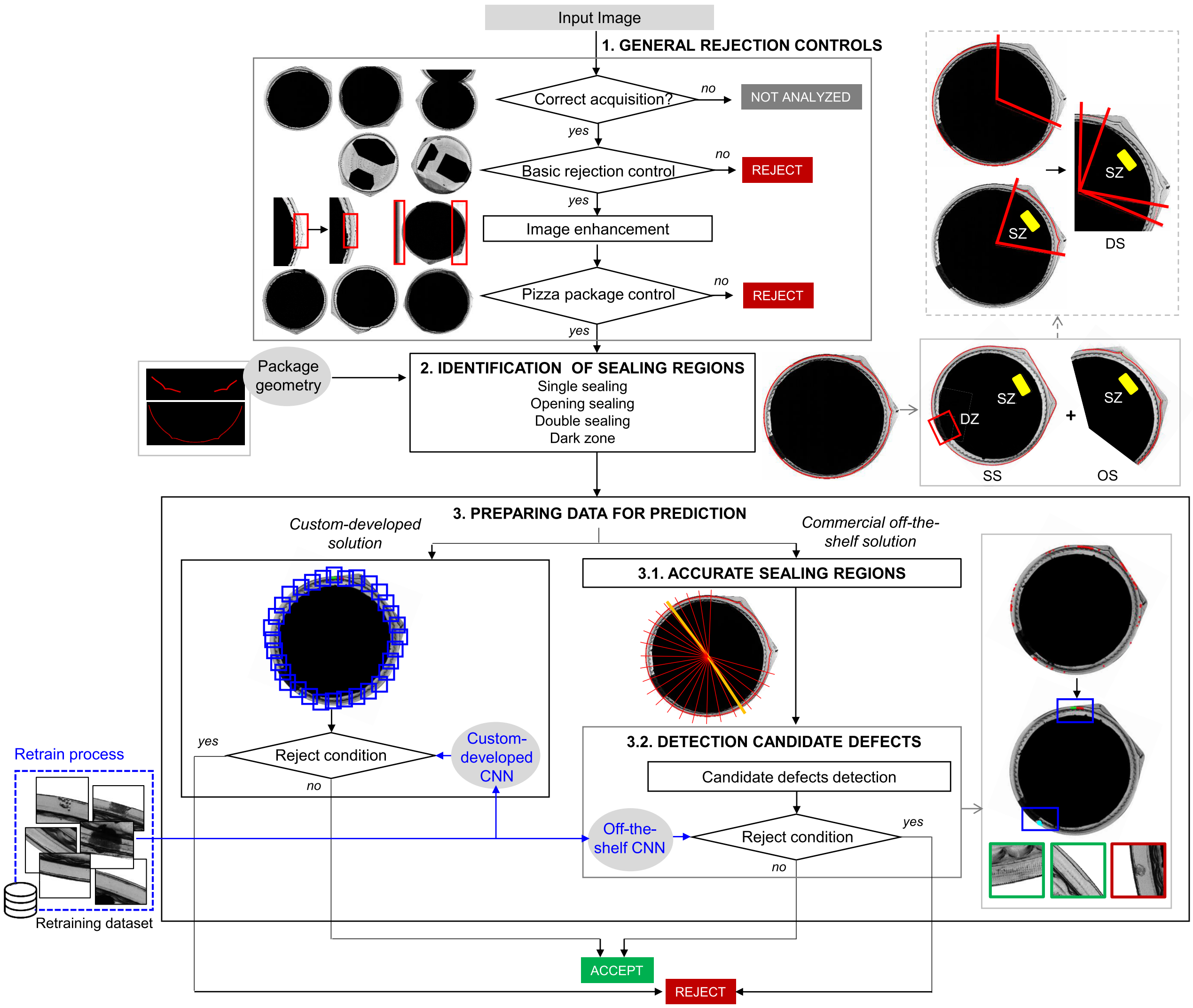


FIGURE 8. **Main steps of the three-phase software designed to automatically evaluate seal and closure of pizza packages. Representative images that illustrate the main actions carried out are also presented for each phase.**

and completely different classes. To apply it, the indications of Halcon were followed [54], [55] converting the collected images to three channels and real number as pixel format. Since the network imposes requirements on the images regarding image dimensions, gray value range and type, we retain the features of the images used by Halcon to pre-train the network.

- The custom-developed ResNet18 with 18 layers. The ResNet's deep-learning architecture was proposed by He *et al.* [56] and is characterized by the residual block and skip connections between blocks. This architecture alleviates the problem of gradient disappearance caused by increasing the network depth, and also improves accuracy by adding considerable depth. The model will be evaluated with pre-training in ImageNet [57]. This will be done using the Python programming language[1] by means of the PyTorch[2] and FastAI[3] optimized tensor libraries for deep learning, and considering a variable learning rate by using the optimizer based on the Adam's algorithm [58], [59]. The loss function used combines a Sigmoid layer and the Binary Cross-Entropy Loss BCELoss) into a single class.[4] A two-step training will be carried out: the first two epochs will run with frozen layers, while the next ones will run after the layers have been unfrozen. In this case, the collected images will be normalized without extra pre-processing.

B. SOFTWARE DESCRIPTION

The proposed software has been designed with the production cadence as the main limiting factor to be addressed. As illustrated in Figure 8, the software applies the three phases described in the next.

[1]https://www.python.org

[2]https://pytorch.org

[3]https://docs.fast.ai

[4]https://pytorch.org/docs/stable/generated/torch.nn.BCEWithLogitsLoss.html

1) PHASE 1: GENERAL REJECTION CONTROLS

This phase performs a set of controls to prevent some cases from entering the CNN. These controls are:

- The *correct acquisition control*, which analyzes whether pizzas are correctly displayed in the captured image by applying image thresholding and detecting whether there are foreground pixels in the top or bottom rows of the image. In this case, pizzas are not analyzed and are accepted by the system although they could be rejected. Our industrial partner suggested acting in this way as the rejection of these cases could dramatically increase the rejection rate. This first control detects the *cut* and *double cut* cases shown in Figure 5.
- The *basic rejection control*, which detects and rejects empty packages and some pizzas with out attached film (see Figure 2(a) and 2(f)). Thresholding is used to detect the pizza in order to compute its area in pixels.

After these two initial controls, an *image enhancement* step is applied to remove undesired vertical white or black lines and balance the gray level of the sealing zone, discarding light areas of the pizza and achieving a constant gray level. The image obtained is examined by the *pizza package control*, which checks the area of the pizza package to detect whether or not it is correct within a minimum and maximum acceptance range. Deformed pizzas, overlapping pizzas (*double*) and some pizzas with out attached film are rejected (see Figures 2(b-e)).

2) PHASE 2: IDENTIFICATION OF SEALING REGIONS

This phase identifies the single sealing, the opening sealing, the double sealing, and the dark zone regions (see Figure 1(d)). To proceed, some geometry information about the packaging is required, particularly a shape model and an input region, which are the same for all the brands. The regions are detected as follows:

- The single sealing is obtained from the contour of the area of the pizza package, which has been previously computed, by applying morphological erosion technique and considering the common parameters of the packages such as the sealing position and the sealing area. In addition, a search zone for the origin of the opening sealing coordinate system is also detected.
- To detect the opening sealing region, the transformation matrix that locates the origin of the shape model in the previously defined search zone is computed. This matrix is used to identify the opening sealing by transforming the input region, which is defined in the same coordinate system as the shape model.
- Once the single sealing and the opening sealing have been identified, the double sealing is automatically obtained using interpolation and union operations.
- To detect the dark zone only present in some brands such as $Brand_2$, different machine vision techniques can be applied [60]. In our implementation, a mirroring projection of the single sealing region is used to determine the search area of the dark zone.

3) PHASE 3: PREPARING DATA FOR PREDICTION

The preparation of the input data for the model depends on the selected CNN solution. In the case of a commercial off-the-shelf CNN, it is necessary to reduce the number of input images to improve the CNN performance in inference time. For this reason, a strategy has been devised to automatically select defects in the image that can lead to rejection. This strategy requires the following processes.

- *Accurate identification of sealing regions*. Depending on the position of the pizza during transportation, the backlight creates some shadows on the image that lead to an enlargement of the sealing contour. Moreover, the double sealing does not have the same shape and does not end in the same position for all brands. To accurately detect the sealing region, a set of cutting lines drawn from the center of the pizza to the boundary of the package are defined covering the single and double sealing regions. Two of these lines have to be in the boundary of both regions (one for each side of the package). To obtain these boundary lines, a rough segmentation of double sealing regions is first obtained using thresholding techniques. Then, to determine the exact position of the boundary lines, a binary search is performed, always taking two lines, one to the left and one to the right of the identified boundary, and evaluating the midline between them. The search, performed on both sides of the package, is repeated until a given angle precision is achieved. Each line is evaluated pixel-by-pixel from the outside to the inside of the package, considering the limits and tolerances defined a priori, as shown in Figure 9. In the case of single sealing lines (first five columns of Figure 9), the last white pixel outside the pizza background, the first black pixel considered as the sealing reference point (the first internal pixel of the sealing), and the first black pixel considered as the end of the sealing width (the second internal pixel after the first black pixel previously determined) are identified to determine the sealing movement to be applied. In the case of double sealing lines (the last two columns of Figure 9), a similar procedure is applied, but considering the two sealing lines. Once the sealing pixels are detected for each line, the preliminary detection regions obtained in the previous step are transformed to fit these new accurate points by applying scales, translations, morphological operators, and unions to each of the segments of the cutting lines taking the center of the pizza as a reference.
- *Candidate defects detection*. Candidate defects can be detected using thresholding, as they are darker than the correct sealing. They are pre-processed to generate a 224×224 pixel image with the defect at its center and the characteristics for the commercial off-the-shelf CNN previously described. These images are

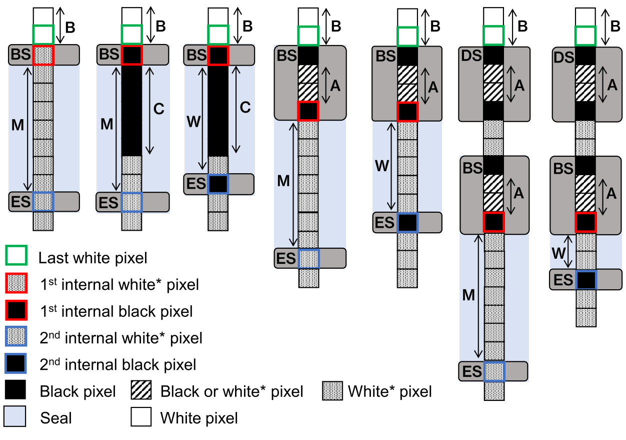


FIGURE 9. Different cases that can appear in the process of accurate identification of sealing regions, where *B* is the background, *M* is the maximum sealing width; *W* is the sealing width, less than *M*; *A* is a black or *white** pixel, where the distance in pixels between the first black pixel and the last black pixel is less than *N*; *N* is the acceptable distance between pixels to be considered in a single line; *white** is a white or gray pixel with an intensity lower than that of the black pixels in the sealing; *C* represents more than 18 consecutive black pixels (shadow case); *BS* is the beginning of the sealing; *ES* is the end of the sealing; and *DS* is the double sealing (a line to skip).

entered on the commercial off-the-shelf CNN, which classifies the pizza as rejected or not. In addition, to avoid processing all the regions detected as candidate defects, different filters are applied to directly discard those that do not have the possibility of being defects according to relevant features of shape and color. For instance, morphological operators such as opening, closing, dilation, erosion and fill up, intersections and unions, and area and color attributes via thresholding are used to identify relevant features and detect small bubbles, small food ingredients (see Figure 3(j-o)) or flour (see Figure 3(i)). By studying the shape of the candidate defects' boundaries, different marks (see Figure 3(a-e)) can also be discarded, such as sealing lines entering the inspection regions and small candidate defects. More complex candidate defects are analyzed using some contrast enhancement and partial derivatives of the Gaussian smoothing kernel [61]. In this way, the number of images to be classified is reduced and time restrictions can be satisfied.

For the custom-developed CNN, it is only necessary to create a set of 64 images with $224x224$ normalized pixels. These images will represent all parts of the sealing region in a mosaic mode.

V. RESULTS AND DISCUSSION

In this section, the different experiments that have been carried out to evaluate the proposed solution are presented. First, the commercial off-the-shelf and the custom-developed CNN performance will be compared in order to select the best architecture. Second, the selected solution will be compared with the manual procedure. Finally, the correctness of the automatic solution will be evaluated.

TABLE 2. Formulas of statistical features extracted from the confusion matrix where n_{TP}, n_{FP}, n_{TN} and n_{FN} represent the number of true positives, false positives, true negatives and false negatives, respectively.

Measure	Formula
Accuracy (A)	$\frac{n_{TP}+n_{TN}}{n_{TP}+n_{TN}+n_{FP}+n_{FN}}$
Precision (P)	$\frac{n_{TP}}{n_{TP}+n_{FP}}$
Recall (R)	$\frac{n_{TP}}{n_{TP}+n_{FN}}$
F-score (F)	$2 \times \frac{Precision \times Recall}{Precision + Recall}$

A. COMPARING DIFFERENT CNN ARCHITECTURES

1) SETUP

An industrial computer with 64GB of RAM, 9-core processor, and two GPU NVIDIA GeForce GTX 1080 Ti has been used for the training and the inference process. A set of 80% (3647 images) of the original dataset was used for the training, 10% (455 images) for the validation, and 10% (456 images) for the testing. To evaluate the performance of the classifiers, standard measures derived from the *confusion matrix* will be computed (see Table 2), including *accuracy* (the ratio of correct predictions to the total number of predictions), *precision* (the ratio of correct positive predictions to the total number of positive predictions), *recall* (the ratio of predicted positives to the total number of positive labels), and *F-score* (the harmonic mean of precision and recall). The training processes have been repeated five times and only the best results will be shown. The hyperparameters that need to be configured for the commercial off-the-shelf CNN solution before starting the training were set, as indicated by Halcon [62], with the following values: batch size to 64, epochs to 50, initial learning rate to 0.001, learning step per epoch to 2, learning step ratio to 0.9, and momentum to 0.8. For the custom-developed CNN solution, the hyperparameters were set with the following values: batch size of 64, epochs to 50, initial learning rate to 0.001, first and second momentum to 0.09 and 0.999, respectively, epsilon to 1e−05, and weight decay to 0.01. The number of trainable parameters involved in the training process for the custom-developed CNN is 11689512 and, in the case of commercial off-the-shelf CNN, there is no information about it. To avoid the overfitting risk, image processing methods were used by means of rotations and horizontal and vertical mirror displacements to achieve data augmentation when training the images, in order to obtain more strength and generalization.

2) SETTING THE NUMBER OF CLASSES AND CNN SELECTION

The design of the network requires determining the number of classes of the classifier. Two cases were considered: the classification of pizzas in two classes (accepted, with 3441 images; and rejected, with 1117 images) and

TABLE 3. Precision (P), recall (R), accuracy (A), and F-score (F) for the commercial off-the-shelf and custom-developed CNNs considering classifications into two classes (accepted and rejected) and five classes (accepted, rejected, machine marks, shadows, and small defects).

	Commercial off-the-shelf CNN								Custom-developed CNN			
	Two classes				Five classes				Two classes			
Measure	P%	R%	A%	F%	P%	R%	A%	F%	P%	R%	A%	F%
Rejected	97.6	97.6			95.0	93.4			97.6	99.2		
Accepted	99.1	99.1			98.1	97.4			99.7	99.1		
Machine marks					98.6	96.0						
Shadows					100	100						
Small defects					90.0	94.7						
Total	98.3	98.3	98.7	98.3	96.3	96.3	95.6	96.3	98.6	99.1	99.1	98.9

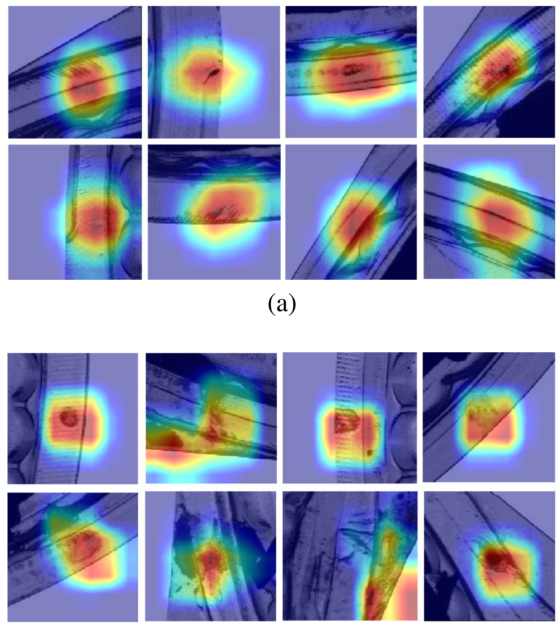
in five classes (*accepted*, with 1883 images; *rejected*, with 1117 images; pizzas with *machine marks*, with 714 images; pizzas with *shadows* due to humidity, fingerprints, or flour, with 94 images; and pizzas with *small defects* due to small bubbles or small food ingredients, with 954 images). To determine the best option, the commercial off-the-shelf CNN has first been tested considering two and five classes. The results obtained are shown in Table 3 and Table 4. For the commercial off-the-shelf CNN, the best results are obtained with the two-classes classification, with precision, recall, accuracy, and F-score values of 98.3%, 98.3%, 98.7%, and 98.3%, respectively. Second, the custom-developed solution has been tested with the two-classes classification, obtaining in this case precision, recall, accuracy, and F-score values of 98.6%, 99.1%, 99.1%, and 98.9%, respectively. Since these results are better than those of the commercial off-the-shelf CNN, this solution will be the one installed in the real scenario. Note that the custom-developed solution has a simpler structure with fewer layers and will consequently be faster than the commercial off-the-shelf solution (see Table 5).

Figure 10 shows the heat maps of some candidate images corresponding to accept and reject cases for the commercial off-the-shelf CNN. Note that the features for the classification of the sealing defects are obtained mainly from the center of the candidate defect images.

B. COMPARING THE AUTOMATIC SOLUTION WITH THE MANUAL PROCEDURE

To compare the manual and the automatic quality control of pizza packages, two scenarios were considered. The manual scenario, the actual real manual control, has two operators placed at the exit belts, outside the clean room, evaluating the pizzas. They had a white background with white light and they took pizzas from the belt one by one. Pizzas were transported at $0.5m/s$. The operators were unable to check all pizzas due to fatigue. The automatic scenario had the proposed solution, with the custom-developed CNN, installed in the clean room after the tray sealer and the cutting machines (see Figure 6(c)). The system examined the pizzas transported on a conveyor belt at $1.0m/s$. Non-rejected pizzas were transported out of the clean room with two exit conveyor belts.

For comparison purposes, the production department of our industrial partner collected reject data for one month



(a)

(b)

FIGURE 10. Heatmaps of candidate defect images for the commercial off-the-shelf CNN corresponding to (a) accept and (b) reject cases.

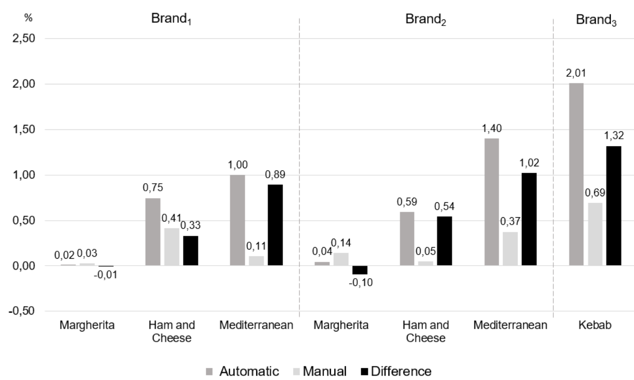


FIGURE 11. Automatic and manual rejection percentage (%) and their difference for different toppings of different brands.

considering the manual and automatic scenarios. The latter was supervised by two operators who checked the correctness of automatic rejections and acceptances carried out by the system. In Figure 11, the actual rejection percentage of

TABLE 4. Confusion matrices for the commercial off-the-shelf CNN for the two and five classes classification and for the custom-developed CNN for the two classes classification. The columns represent for each class the instances with its ground truth label and the rows the instances predicted to belong to this class.

	Commercial off-the-shelf CNN							Custom-developed CNN	
	Two classes		**Five classes**					**Two classes**	
Confusion Matrices	Rejected	Accepted	Rejected	Accepted	Machine marks	Shadows	Small defects	Rejected	Accepted
Rejected	120	3	113	0	0	0	8	120	1
Accepted	3	330	2	152	1	0	1	3	332
Machine marks			1	1	72	0	1		
Shadows			0	0	0	9	0		
Small defects			3	2	0	0	90		

TABLE 5. The execution and training runtimes for the commercial off-the-shelf and the custom-developed CNNs.

	Execution runtime (ms)			
CNN	Mean	Maximum	Minimum	**Training runtime**
Commercial off-the-shelf with candidate defects	223	276	15	52 minutes 47 seconds
Commercial off-the shelf without candidate defects	452	562	330	
Custom-developed	252	287	216	22 minutes 55 seconds

TABLE 6. Results in % of automatically rejected, accepted, and analyzed pizzas for mediterranean and chicken of *Brand*$_1$, in bold. For each case the results of manual supervision are presented in italics.

Automatically→	**Rejected**	**Accepted**	**Analyzed**	**Total**	*Manual ↓*
MEDITERRANEAN	**69**	**4013**	**4082**	**4092**	
Accepted		99.98	98.29	98.04	*4012*
False Negatives		0.02	0.02	0.02	*1*
Rejected	97.10		1.64	1.64	*67*
False Positives	2.90		0.05	0.05	*2*
Not analyzed				0.24	*10*
CHICKEN	**41**	**4724**	**4765**	**4766**	
Accepted		99.94	99.08	99.06	*4721*
False Negatives		0.06	0.06	0.06	*3*
Rejected	92.68		0.80	0.80	*38*
False Positives	7.32		0.06	0.06	*3*
Not analyzed				0.02	*1*

each scenario is presented considering margherita, ham and cheese, and mediterranean for *Brand*$_1$ and *Brand*$_2$, and kebab for *Brand*$_3$. The difference between both scenarios is also presented. It can be observed that in the case of margherita (*Brand*$_1$ and *Brand*$_2$), with a low rejection rate (0.03% for *Brand*$_1$ and 0.14% for *Brand*$_2$ in the manual case), the percentage of manual rejection is higher than the automatic one (0.02% for *Brand*$_1$ and 0.04% for *Brand*$_2$ in the automatic case). This is because operators rejected pizzas with small bubbles or small food ingredients in the pizza sealing that did not pass through the sealing (see Figure 3(j-o)), whereas the automatic system accepted them. In the case of ham and cheese and mediterranean for *Brand*$_1$ and *Brand*$_2$, and kebab for *Brand*$_3$, there was an opposite behavior as the automatic system detected more cases than the manual one (0.75% and 1.00% for ham and cheese and mediterranean of *Brand*$_1$, respectively, in the automatic case, and 0.41% and 0.11% in the manual case; 0.59% and 1.40% for ham and cheese and mediterranean of *Brand*$_2$, respectively, in automatic case, and 0.05% and 0.37% in the manual case; 2.01% for kebab of *Brand*$_3$ in the automatic case, and 0.69% in the manual case). When analyzing these results, it was noticed that the number of manual rejections was influenced by the number of accumulated rejections. It was observed that, when there were few rejections, operators tended to be more sensitive to rejecting pizzas that should have been accepted. On the contrary, when the number of rejections was higher, they tended to accept pizzas that they should have rejected. In the case of the automatic scenario, there was a constant behavior since the system did not take into account previous cases and, therefore, a more objective criterion was achieved.

C. CORRECTNESS OF THE AUTOMATIC SYSTEM

The supervision of the automatic system allows us to evaluate its correctness. Tables 6 and 7 present the results for the mediterranean and chicken pizzas of *Brand*$_1$ and *Brand*$_2$, and Table 8 presents those of the kebab and capricciosa pizzas of *Brand*$_3$. The number of rejected, accepted, analyzed, and total pizzas for each case evaluated is presented in bold in the row labeled as *Automatically*. In addition, for each category, the classification after manual supervision of the automatic results is presented showing the number of correctly acceptedpizzas, false negatives (non-rejected pizzas that have to be rejected), correctly rejected pizzas, false positives

TABLE 7. Results in % of automatically rejected, accepted, and analyzed pizzas for mediterranean and chicken of $Brand_2$, in bold. For each case the results of manual supervision are presented in italics.

Automatically→	Rejected	Accepted	Analyzed	Total	*Manual* ↓
MEDITERRANEAN	**43**	**4517**	**4560**	**4572**	
Accepted		99.98	99.02	98.78	*4516*
False Negatives		0.02	0.02	0.02	*1*
Rejected	98.82		0.92	0.92	*42*
False Positives	2.35		0.02	0.02	*1*
Not analyzed				0.25	*12*
CHICKEN	**37**	**3820**	**3857**	**3857**	
Accepted		99.97	99.01	99.01	*3819*
False Negatives		0.03	0.03	0.03	*1*
Rejected	97,30		0.93	0.93	*36*
False Positives	2.70		0.03	0.03	*1*
Not analyzed				0.00	*0*

TABLE 8. Results in % of automatically rejected, accepted, and analyzed pizzas for kebab and capricciosa of $Brand_3$, in bold. For each case the results of manual supervision are presented in italics.

Automatically→	Rejected	Accepted	Analyzed	Total	*Manual* ↓
KEBAB	**125**	**4710**	**4835**	**4840**	
Accepted		99.92	97.33	97.23	*4706*
False Negatives		0.08	0.08	0.08	*4*
Rejected	95.20		2.46	2.46	*119*
False Positives	4.80		0.12	0.12	*6*
Not analyzed				0.10	*5*
CAPRICCIOSA	**127**	**4211**	**4338**	**4338**	
Accepted		100.00	97.07	97.07	*4211*
False Negatives		0.00	0.00	0.00	*0*
Rejected	91.34		2.67	2.67	*116*
False Positives	8.66		0.25	0.25	*11*
Not analyzed				0.00	*0*

(rejected pizzas that have to be accepted), and non-analyzed pizzas. For all the cases presented, automatic rejection ranges from 0.86% to 2.93% of the total production. A similar behavior is obtained in the other cases not included in these tables, where, for $Brand_1$ and $Brand_2$, the automatic rejection ranges from 0.0% to 1.8, and for $Brand_3$, from 0.0% to 3.0%. When the system obtains rejection values higher than those presented, this indicates that the sealing system is not working properly and an alarm is triggered. Possible causes of this abnormal behavior could be that the machine that joins the film with the product is not clean or that it is necessary to replace parts due to wear, or that temperature and/or pressure are not properly adjusted (see Figure 3 (a-e) and (j)). Another cause could be that the distance between pizzas on the conveyor belt is not adequate (see Figure 2(d) and Figure 3).

Focusing on false positives, they range from 0.02% to 0.25% of the total production, reaching the maximum values in $Brand_3$, which has a film without constant density and without transparent sealing, unlike $Brand_1$, which has a transparent and constant package, and $Brand_2$, which only has a small dark-printed color zone in the package.

In the case of false negatives, the obtained values are between 0.0% and 0.08% of the total production, and the values of the non-analyzed pizzas are between 0.0% and 0.25%. These false negatives, corresponding to pizzas rejected in the manual inspection but accepted by the automatic one, were reexamined by the system and were rejected. This is due to the fact that these pizzas failed in the correct acquisition control since they were not displayed correctly in the processed image and were accepted by default by the system (see Figure 5).

The number of non-analyzed pizzas was strongly related to conflict situations of the conveyor belt caused by subsequent machines in the production line. False negatives are due to belt overload, which may generate images with a double cut or a cut. False positives are generated when the distance between pizzas is not adequate and the pizzas reach the rejection system before the system has processed the image. These pizzas are automatically rejected even though they could have been accepted. This situation leads to an increase in false negatives and false positives.

Considering all the data as a whole and the analyzed totals, the maximum and minimum precision values obtained are 99.95% and 99.74%, respectively, and the mean is 99.87%, with 99.9% for $Brand_1$, 99.95% for $Brand_2$, and 99.77% for $Brand_3$. Our industrial partner considered that the values obtained were very satisfactory.

VI. CONCLUSION AND FUTURE WORK

A key process in industrial pizza production is the packaging control, which checks the sealing, among other verifications, to ensure that the package closure is correct. By using computer vision techniques and deep-learning strategies, a system that automatically controls the closure and sealing of pizza packages has been proposed. A comparison process has been carried out to select the CNN architecture that best fits our scenario. A custom-developed ResNet18 CNN has been found to perform better than a commercial off-the-shelf solution. The system performs 100% inline inspection of pizzas and supports different pizza brands and toppings while satisfying the production cadence. Moreover, it adapts to films with different densities and serigraphies. To test the system, pizzas of different brands and toppings have been considered, obtaining very promising results with false positives ranging from 0.02% to 0.25% and false negatives between 0.0% and 0.08%. Furthermore, the system not only inspects all the pizzas, but also doubles the speed of the conveyor belts when compared to manual inspection. Our future work will be centered on the extension of the system to support packages of different shapes and materials leading to a more general solution. Moreover, since a key element of our proposal is the number images acquired by the computer vision system and two main components of this system are the cameras and the lights, which are in continuous evolution, our interest is focused on the evaluation of the capabilities of these new technologies to exploit and integrate them into our solution to obtain more general and efficient proposals able to detect as many features as possible. Finally, we want to expand our work by considering other networks to automatically adapt to new scenarios while reducing the degree of user interaction.

ACKNOWLEDGMENT

The authors acknowledge the members of the Vision Department (Research and Development) of TAVIL IND, S.A.U., for their advice and contribution to this work, which have been fundamental.

REFERENCES

[1] E. R. Davies, *Image Processing for the Food Industry*. Singapore: World Scientific, 2000.
[2] D.-W. Sun, *Computer Vision Technology for Food Quality Evaluation* (Food Science and Technology). New York, NY, USA: Academic, 2008.
[3] B. R. Mehta and Y. J. Reddy, *Industrial Process Automation Systems*, 1st ed. Oxford, U.K.: Butterworth-Heinemann, 2014.
[4] D. Sankowski and J. Nowakowski, *Computer Vision in Robotics and Industrial Applications* (Series in Computer Vision), vol. 3. Singapore: World Scientific, 2014.
[5] A. Raju, ''A survey on computer vision technology for food quality evaluation,'' *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 3297, no. 8, pp. 14860–14865, 2016.
[6] D.-W. S. Ma, J.-H. Qu, D. Liu, H. Pu, W.-H. Gao, and X.-A. Zeng, ''Applications of computer vision for assessing quality of agri-food products: A review of recent research advances,'' *Crit. Rev. Food Sci. Nutr.*, vol. 56, no. 1, pp. 113–127, 2016.
[7] R. Syed, S. Suriadi, M. Adams, W. Bandara, S. J. J. Leemans, C. Ouyang, A. H. M. ter Hofstede, I. van de Weerd, M. T. Wynn, and H. A. Reijers, ''Robotic process automation: Contemporary themes and challenges,'' *Comput. Ind.*, vol. 115, Feb. 2020, Art. no. 103162.
[8] X. Mai, H. Zhang, X. Jia, and M. Q.-H. Meng, ''Faster R-CNN with classifier fusion for automatic detection of small fruits,'' *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 3, pp. 1555–1569, Jul. 2020.
[9] R. Geirhos, D. H. J. Janssen, H. H. Schütt, J. Rauber, M. Bethge, and F. A. Wichmann, ''Comparing deep neural networks against humans: Object recognition when the signal gets weaker,'' 2018, *arXiv:1706.06969*.
[10] G. Aceto, V. Persico, and A. Pescape, ''A survey on information and communication technologies for industry 4.0: State-of-the-art, taxonomies, perspectives, and challenges,'' *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3467–3501, 2019.
[11] L. Bodenhagen, A. R. Fugl, A. Jordt, M. Willatzen, K. A. Andersen, M. M. Olsen, R. Koch, H. G. Petersen, and N. Krüger ''An adaptable robot vision system performing manipulation actions with flexible objects,'' *IEEE Trans. Autom. Sci. Eng.*, vol. 11, no. 3, pp. 749–765, Jul. 2014.
[12] Y. LeCun, Y. Bengio, and G. E. Hinton, ''Deep learning,'' *Nature*, vol. 521, pp. 436–444, Dec. 2015.
[13] A. M. Karadeniz, I. Arif, A. Kanak, and S. Ergun, ''Digital twin of eGastronomic things: A case study for ice cream machines,'' in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2019, pp. 1–4.
[14] S. Doltsinis, M. Krestenitis, and Z. Doulgeri, ''A machine learning framework for real-time identification of successful snap-fit assemblies,'' *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 1, pp. 513–523, Jan. 2020.
[15] H. Darvishi, D. Ciuonzo, E. R. Eide, and P. S. Rossi, ''Sensor-fault detection, isolation and accommodation for digital twins via modular data-driven architecture,'' *IEEE Sensors J.*, vol. 21, no. 4, pp. 4827–4838, Feb. 2021.
[16] J. Wang, Y. Ma, L. Zhang, R. X. Gao, and D. Wu, ''Deep learning for smart manufacturing: Methods and applications,'' *J. Manuf. Syst.*, vol. 48, pp. 144–156, Jul. 2018.
[17] G. Onoufriou, R. Bickerton, S. Pearson, and G. Leontidis, ''Nemesyst: A hybrid parallelism deep learning-based framework applied for Internet of Things enabled food retailing refrigeration systems,'' *Comput. Ind.*, vol. 113, Dec. 2019, Art. no. 103133.
[18] A. Banan, A. Nasiri, and A. Taheri-Garavand, ''Deep learning-based appearance features extraction for automated carp species identification,'' *Aquacultural Eng.*, vol. 89, May 2020, Art. no. 102053.
[19] V. Kakani, V. H. Nguyen, B. P. Kumar, H. Kim, and V. R. Pasupuleti, ''A critical review on computer vision and artificial intelligence in food industry,'' *J. Agricult. Food Res.*, vol. 2, Dec. 2020, Art. no. 100033.
[20] A. Nasiri, M. Omid, and A. Taheri-Garavand, ''An automatic sorting system for unwashed eggs using deep learning,'' *J. Food Eng.*, vol. 283, Oct. 2020, Art. no. 110036.
[21] S. N. Bairampalli, F. Ustolin, D. Ciuonzo, and P. S. Rossi, ''Digital moka: Small-scale condition monitoring in process engineering,'' *IEEE Sensors Lett.*, vol. 5, no. 3, pp. 1–4, Mar. 2021.
[22] T. Brosnan and D.-W. Sun, ''Improving quality inspection of food products by computer vision—A review,'' *J. Food Eng.*, vol. 61, no. 1, pp. 3–16, Jan. 2004.
[23] T. B. Moeslund, *Introduction to Video and Image Processing: Building Real System and Applications*, 1st ed. London, U.K.: Springer-Verlag, 2012.
[24] L. Li, K. Ota, and M. Dong, ''Deep learning for smart industry: Efficient manufacture inspection system with fog computing,'' *IEEE Trans. Ind. Informat.*, vol. 14, no. 10, pp. 4665–4673, Oct. 2018.
[25] C. Ge, J. Wang, J. Wang, Q. Qi, H. Sun, and J. Liao, ''Towards automatic visual inspection: A weakly supervised learning method for industrial applicable object detection,'' *Comput. Ind.*, vol. 121, Oct. 2020, Art. no. 103232.
[26] J. P. Yun, W. C. Shin, G. Koo, M. S. Kim, C. Lee, and S. J. Lee, ''Automated defect inspection system for metal surfaces based on deep learning and data augmentation,'' *J. Manuf. Syst.*, vol. 55, pp. 317–324, Apr. 2020.
[27] Y. Watanabe, H. Oku, and M. Ishikawa, ''Architectures and applications of high-speed vision,'' *Opt. Rev.*, vol. 21, no. 6, pp. 875–882, Nov. 2014.
[28] Q.-Y. Gu and I. Ishii, ''Review of some advances and applications in real-time high-speed vision: Our views and experiences,'' *Int. J. Automat. Comput.*, vol. 13, no. 4, pp. 305–318, 2016.
[29] Y. Zhang, Z. Jia, and Y. Dai, ''Real-time performance analysis of industrial serial production systems with flexible manufacturing,'' in *Proc. Joint 10th Int. Conf. Soft Comput. Intell. Syst. (SCIS) 19th Int. Symp. Adv. Intell. Syst. (ISIS)*, Dec. 2018, pp. 360–365.
[30] D.-W. Sun and T. Brosnan, ''Pizza quality evaluation using computer vision. Part 1: Pizza base and sauce spread,'' *J. Food Eng.*, vol. 57, no. 1, pp. 81–89, 2003.
[31] D.-W. Sun and T. Brosnan, ''Pizza quality evaluation using computer vision—-Part 2: Pizza topping analysis,'' *J. Food Eng.*, vol. 57, no. 1, pp. 91–95, Mar. 2003.
[32] C.-J. Du, D. Barbin, and D.-W. Sun, ''Chapter 19—Quality evaluation of pizzas,'' in *Computer Vision Technology for Food Quality Evaluation*, D.-W. Sun, Ed., 2nd ed. San Diego, CA, USA: Academic, 2016, pp. 465–485.
[33] C.-J. Du and D.-W. Sun, ''Shape extraction and classification of pizza base using computer vision,'' *J. Food Eng.*, vol. 64, no. 4, pp. 489–496, Oct. 2004.

[34] C.-J. Du and D.-W. Sun, "Pizza sauce spread classification using colour vision and support vector machines," *J. Food Eng.*, vol. 66, no. 2, pp. 137–145, Jan. 2005.

[35] C.-J. Du and D.-W. Sun, "Comparison of three methods for classification of pizza topping using different colour space transformations," *J. Food Eng.*, vol. 68, no. 3, pp. 277–287, Jun. 2005.

[36] M. Al-Sarayreh, M. M. Reis, W. Q. Yan, and R. Klette, "Potential of deep learning and snapshot hyperspectral imaging for classification of species in meat," *Food Control*, vol. 117, Nov. 2020, Art. no. 107332.

[37] S. Fan, J. Li, Y. Zhang, X. Tian, Q. Wang, X. He, C. Zhang, and W. Huang, "On line detection of defective apples using computer vision system combined with deep learning methods," *J. Food Eng.*, vol. 286, Dec. 2020, Art. no. 110102.

[38] MVTec. *MVTec Software GmbH HALCON—The Power of Machine Vision*. Accessed: Aug. 7, 2020. [Online]. Available: https://www.mvtec.com/products/halcon/

[39] T. MathWorks and I. *Pretrained Convolutional Neural Networks*. Accessed: Sep. 15, 2020. [Online]. Available: https://de.mathworks.com/help/deeplearning/ug/pretrained-convolutional-% neural-networks.html

[40] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[41] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[42] Q. Xie, D. Li, J. Xu, Z. Yu, and J. Wang, "Automatic detection and classification of sewer defects via hierarchical deep learning," *IEEE Trans. Autom. Sci. Eng.*, vol. 16, no. 4, pp. 1836–1847, Oct. 2019.

[43] S. Niu, B. Li, X. Wang, and H. Lin, "Defect image sample generation with GAN for improving defect recognition," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 3, pp. 1611–1622, Jul. 2020.

[44] M. Dudbridge, *Handbook of Seal Integrity in the Food Industry*. Chichester, U.K.: Wiley, 2016.

[45] S. A. Ashter, *Thermoforming Single Multilayer Laminates*. Oxford, U.K.: William Andrew Publishing, 2014.

[46] E. Bodenstorfer, J. Fürtler, J. Brodersen, K. J. Mayer, C. Eckel, K. Gravogl, and H. Nachtnebel, "High-speed line-scan camera with digital time delay integration," in *Proc. SPIE*, vol. 6496, pp. 165–174, Feb. 2007.

[47] N. Gat, "Imaging spectroscopy using tunable filters: A review," *Proc. SPIE*, vol. 4056, pp. 50–64, Apr. 2000.

[48] G. Polder, G. W. A. M. van der Heijden, L. C. P. Keizer, and I. T. Young, "Calibration and characterisation of imaging spectrographs," *J. Near Infr. Spectrosc.*, vol. 11, no. 3, pp. 193–210, Jun. 2003.

[49] AIA Global Vision Systems Trade Association. *GigE Vision*. Accessed: Dec. 10, 2020. [Online]. Available: https://www.visiononline.org/vision-standards-details.cfm?type=5/

[50] Teledyne DALSA. *Introducing TurboDrive*. Accessed: Dec. 10, 2020. [Online]. Available: https://www.teledynedalsa.com/en/learn/knowledge-center/turbodrive/

[51] P. Geladi, J. Burger, and T. Lestander, "Hyperspectral imaging: Calibration problems and solutions," *Chemometrics Intell. Lab. Syst.*, vol. 72, no. 2, pp. 209–217, Jul. 2004.

[52] N. Banus, I. Boada, P. Xiberta, and P. Toldra, "Design and deployment of a generic software for managing industrial vision systems," *IEEE Trans. Autom. Sci. Eng.*, early access, Jun. 2, 2021, doi: 10.1109/TASE.2021.3078787.

[53] J. Lu, V. Behbood, P. Hao, H. Zuo, S. Xue, and G. Zhang, "Transfer learning using computational intelligence: A survey," *Knowl.-Based Syst.*, vol. 80, pp. 14–23, May 2015.

[54] MVTec. *HALCON 20.05 Progress: HALCON Operator Reference*. Accessed: Aug. 7, 2020. [Online]. Available: https://www.mvtec.com/fileadmin/Redaktion/mvtec.com/documentation/refer% ence/reference_hdevelop.pdf

[55] MVTec. *HALCON 20.05 Progress: Solution Guide II-D—Classification*. Accessed: Aug. 11, 2020. [Online]. Available: https://www.mvtec.com/fileadmin/Redaktion/mvtec.com/documentation/solut% ion_guide/solution_guide_ii_d_classification.pdf

[56] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[57] Stanford Vision Lab. (2011). *ImageNet*. Accessed: Feb. 23, 2021. [Online]. Available: http://image-net.org

[58] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017, *arXiv:1412.6980*.

[59] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 2019, *arXiv:1711.05101*.

[60] E. R. Davies, *Computer Vision: Principles, Algorithms, Applications, Learning*, 5 ed. New York, NY, USA: Academic, 2018.

[61] B. M. Haar Romeny, *Front-End Vision and Multi-Scale Image Analysis*, 1st ed. Norwell, MA, USA: Kluwer, 2003.

[62] MVTec. *Deep Learning*. Accessed: Sep. 15, 2020. [Online]. Available: https://www.mvtec.com/doc/halcon/1905/en/toc_deeplearning.html

NÚRIA BANÚS received the B.Sc. and M.Sc. degrees in computer science from the Universitat de Girona (UdG), Girona, Catalonia, in 2016 and 2018, respectively. She is currently a Pre-Doctoral Researcher in an Industrial Doctoral Program with the Graphics and Imaging Laboratory and the Computer Science and Applied Mathematics Department, Universitat de Girona, and with the Research and Development Department, TAVIL IND, S.A.U., Girona, within the Computer Vision Team. She carries out research and development on image processing techniques applied to industrial processes.

IMMA BOADA received the M.Sc. degree in computer science from the Universitat Autònoma de Barcelona, Bellaterra, Catalonia, in 1992, and the Ph.D. degree in computer science from the Universitat Politècnica de Catalunya, Barcelona, Catalonia, in 2001. She is currently a Professor with the Computer Science and Applied Mathematics Department, Universitat de Girona, Girona, Catalonia, and a Researcher with the Graphics and Imaging Laboratory, Girona, where she carries out research and development on visualization, image processing, serious games, and e-learning. She has special interest on medical applications and the automation of industrial processes.

ANTON BARDERA received the M.Sc. degree in telecommunications engineering from the Universitat Politècnica de Catalunya, Barcelona, Catalonia, in 2002, and the Ph.D. degree in computer science from the Universitat de Girona, Girona, Catalonia, in 2008. He is currently an Associate Professor with the Computer Science and Applied Mathematics Department, Universitat de Girona, and a Researcher with the Graphics and Imaging Laboratory, Girona, where he carries out research and development on visualization, image processing, information theory, and biomedical applications. He has coauthored two books, more than 40 research articles in international journals, and many conference papers.

POL TOLDRÀ received the B.Sc. degree in industrial electronics from the Universitat de Girona, Girona, Catalonia, in 2001, and the M.Sc. degree in automatics and industrial electronics from the Universitat Rovira i Virgili, Tarragona, Catalonia, in 2004. He is currently the Head of the Research and Development Department, TAVIL IND, S.A.U., Girona. He has more than 20 years of experience in automation of industrial processes.