Article

Apparatus and Method of Defect Detection for Resin Films

Ruey-Kai Sheu [1], **Ya-Hsin Teng** [1], **Chien-Hao Tseng** [2] and **Lun-Chi Chen** [1,*]

1. Department of Computer Science, Tunghai University, Taichung 40704, Taiwan; rickysheu@thu.edu.tw (R.-K.S.); yahsin0606@gmail.com (Y.-H.T.)
2. National Center for High-Performance Computing, National Applied Research Laboratories, Hsinchu 30076, Taiwan; 0903049@nchc.narl.org.tw
* Correspondence: lunchi@thu.edu.tw; Tel.: +886-4-2359-0415

check for updates

Abstract: A defect inspection of resin films involves processes of detecting defects, size measuring, type classification and reflective action planning. It is not only a process requiring heavy investment in workforce, but also a tension between quality assurance with a 50-micrometer tolerance and visibility of the naked eye. To solve the difficulties of the workforce and time consumption processes of defect inspection, an apparatus is designed to collect high-quality images in one shot by leveraging a large field-of-view microscope at 2K resolution. Based on the image dataset, a two-step method is used to first locate possible defects and predict their types by a defect-shape-based deep learning model using the LeNet-5-adjusted network. The experimental results show that the proposed method can precisely locate the position and accurately inspect the fine-grained defects of resin films.

Keywords: microscale; defect inspection; convolution neural network; plastic resin films

1. Introduction

Defects considerably affect the quality of plastic products, and defect inspections are conducted for quality control. Defects can stain the surfaces of plastic products if quality control is not performed thoroughly, which could lead to a decrease in the sales of the product because of unfavorable impression on the customer, and consequently cause losses to the company. The most common technique of plastic product fabrication is injection molding, which involves producing parts by melting granular plastic and injecting them into a mold, and subsequently cooling the product. However, granular plastic defect inspection is difficult because the size of the defects is at microscale. Furthermore, the inside of the granular plastic can have defects. Therefore, granular plastics can be first compressed into plastic resin films. This increases microscale defect visibility. Thus, the inspection can be conducted easily. Industrial inspection systems for defect detection can mostly be divided into two types, namely the traditional procedure and the automatic procedure. The traditional procedure, which is currently prominent, involves observing whether plastic resin film quality is satisfactory through human vision by using a magnifier as the inspection tool. This inspection procedure has the following problems:

1. Because numerous items must be inspected, high-intensity repetitive inspection can cause fatigue and lethargy in inspectors. The quality of inspected products is not guaranteed in such scenarios. This method is prone to workplace injuries because the inspecting staff can injure their eyes.
2. Human vision cannot precisely discern how large any defect is; humans have at best an approximate awareness of whether a defect is present. This may cause the inspection process to miss some defects. Furthermore, it is nearly impossible to standardize human vision; each inspector's eyesight will have its own peculiarities.

3. In some cases, the manufacturer may want to tolerate no defect larger than 25 µm. However, human vision is limited to objects larger than approximately 50 µm; this limit is not acceptable for high-quality products.
4. Defect detection is not immediate. Using human vision for the inspection process is considerably time consuming and inefficient. The average number of plastic resin films that can be checked by a person in a day is approximately 12 to 15.

Automatic procedures, including machine vision-based methods of defect detection, have gained considerable popularity for their high speed, high precision, and real-time performance. Several strategies for defect detection have been explored in numerous studies. These machine vision-based methods are used in the inspection of many industrial products, such as metal [1], fabric [2], steel [3], car surfaces [4], and light-emitting diode chips [5]. The methods of surface defect detection can be divided into three types, as depicted in Figure 1, namely traditional image processing, feature extraction with machine learning, and deep learning methods.
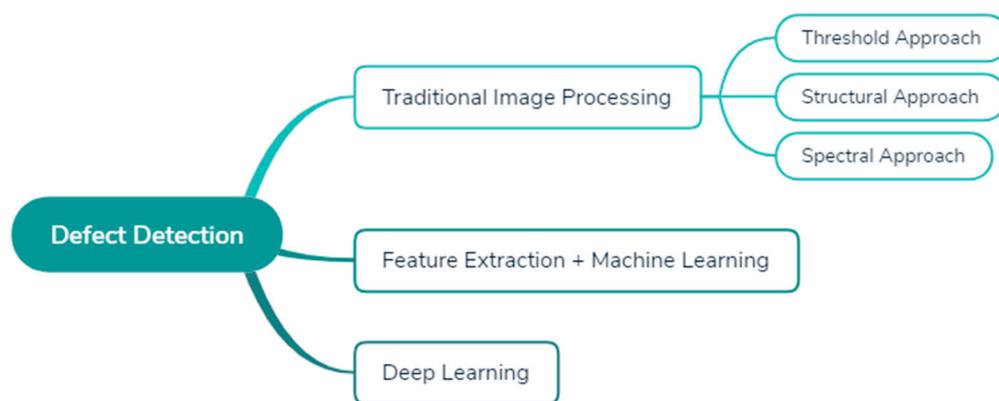
Figure 1. Different approaches to defect detection.

Primitive attributes are used to detect defects in traditional image processing methods. These methods can be divided into three approaches, namely threshold, structural, and spectral. The threshold approach involves transforming grayscale images to binary images, separating the background and defects by applying various types of thresholds such as adaptive thresholding method [6] and the Otsu [7,8] method. The structural approach includes edge [9], skeleton [10], and morphological operations [11]. Fourier [12], wavelet [13], and Gabor transforms [14] are used in the spectral approach. However, it is difficult to classify various nondefect objects using traditional image processing because nondefect items are not easily filtered.

In feature extraction using a machine learning method, defect detection has two stages, namely feature extraction and machine learning, that is, using a feature extraction algorithm to extract various features from images, then using machine learning techniques to determine the pattern or relationship among different classes. Shumin [15] proposed a novel fabric defect detection technique using a histogram of oriented gradient (HOG) for counting occurrences of gradient orientation in localized portions of an image, AdaBoost was used to select a small set of HOG data for support vector machine (SVM) to classify fabric defects. Kuang et al. [16] proposed a method for bamboo strip defect detection using a set of features based on local binary pattern (LBP) and gray level co-occurrence matrix (GLCM). LBP is used to describe the local texture features of the image and GLCM is used to characterize the texture of an image by calculating how often pairs of pixels with specific values in a specified spatial relationship occur in an image. After extracting the features, an SVM was proposed to classify defects. Chang et al. [17] proposed a method for defect detection on the compact lens. He segmented objects by applying weighted Sobel filters and watersheds and then used the SVM for classification. Watershed is a transformation that considers the images as topographic maps, the brightness of the images representing its height and finds the lines that run along the tops of ridges. Watershed can

segment two objects that are close to each other. Zhou et al. [18] proposed a surface defect for a vehicle body by using a multiscale Hessian matrix fusion method to determine defect regions and SVM to classify defects.

Deep learning methods have achieved excellent results in many fields [19–21] and surface defect detection is one of them. Several defect detection methods based on convolutional neural networks (CNNs) have been proposed. Arikan et al. [22] proposed a CNN model for classification, setting two classes: defect or nondefect. This model was designed to handle capacity and real-time speed requirements. A generative adversarial network (GAN) was used to generate more data. Wang et al. [23] proposed a fast and robust CNN-based defect detection model by using CNNs with a sliding window to localize the product damage. The sliding window is time consuming and is not sufficiently efficient for our data. Mei et al. [24] designed a multiscale convolutional denoising autoencoder network (MSCDAE) model for fabric defect detection by using reconstructed image patches with the model at multiple Gaussian pyramid levels and synthesized results from these pyramid levels. Defective regions in the reconstruction residual maps were generated using the CDAE networks. This model can be trained with only a small set of defect-free samples and can deliver excellent performance. These two methods can only localize and detect defects but cannot classify the type of defects.

An object detection method is required for resin films because an image can contain various objects at the same time. Object detection is used to detect objects inside an image and many studies have been conducted on object detection. Li et al. [25] proposed an improved you-only-look-once (YOLO) network to detect six types of defects on steel strip surfaces. Cha et al. [26] proposed a structural visual inspection method based on faster R-CNN to detect five types of defects, namely steel delamination, steel corrosion, bolt corrosion, and concrete cracks. Yuan et al. [27] proposed a modified segmentation method and deep neural network to detect defects on the cover glass of mobile phones. GAN was used to generate more data for the deep learning network to overcome the problem of small amount of data. Ferguson et al. [28] used a defect detection system based on the mask region-based CNN (mask R-CNN) architecture to detect the casting defects on the GDXray dataset. Excellent performance is achieved based on transfer learning, using weights pre-trained on the ImageNet dataset, and then trained the defect detection system on the COCO (Common Objects in Context) dataset. Wen et al. [29] proposed an object detection method to detect defects on bearing rollers, using CNN to extract the features of the defects, then classified the defects, and calculated the position of the defects simultaneously. Chen et al. [30] proposed a novel vision-based method in which deep CNNs (DCNNs) are applied in the defect detection of the fasteners on the catenary support device. The system cascades three DCNN-based detection stages, including single shot multibox detector (SSD) and YOLO to localize the cantilever joints and their fasteners. Then, a classifier was used to classify defects. Li et al. [31] proposed a surface defect detection model based on the SSD network that was combined with a MobileNet to detect the sealing surface of an oil chili to achieve real-time and accurate detection. The Hough circle transform was applied to detect the oil chili. However, while above mentioned methods proposes object detection networks that calculated the position and classifies defects in one whole network, traditional image processing methods are sufficient for finding the objects because of the simple background in the proposed system. Therefore, object detection networks such as YOLO, faster R-CNN are not required for our inspection systems. The proposed method in this is mostly inspired by the following methods. Song et al. [32] proposed a deep CNN-based technique for detecting micro defects on metal screw surfaces by using traditional image processing methods to detect metal screws, then using a CNN network to classify whether a metal screw was defective. Tao et al. [33] designed a cascaded autoencoder architecture to obtain accurate and consistent defect detection results under complex lighting conditions and ambiguous defects. The autoencoder could distinguish the nonbackground objects and only required the basic thresholding method to separate nonbackground objects and the background. A deep convolution network was used to classify various types of defects.

In the proposed method, we used a small microscope to observe data. Two types of plastic resin films were present in our data, namely those with transparent and white backgrounds. Figure 2

illustrates how the resin films appear under the microscope. Various types of nonbackground objects are present on the surface; scratches (Figure 2b) and bubbles (Figure 2c) are not defined as defects. Pollutants in the industrial environment, nondefective items, such as dust (Figure 2d) may appear on the inspected surface. Different nonbackground objects have distinct features. Therefore, an image processing method to segment the nonbackground objects was proposed, and a classification module was then used to classify the defects. The results prove that the proposed method exhibited performance.


(**a**)　　　　　　　　　　　　　　　　　　　(**b**)

(**c**)　　　　　　　　　　　　　　　　　　　(**d**)
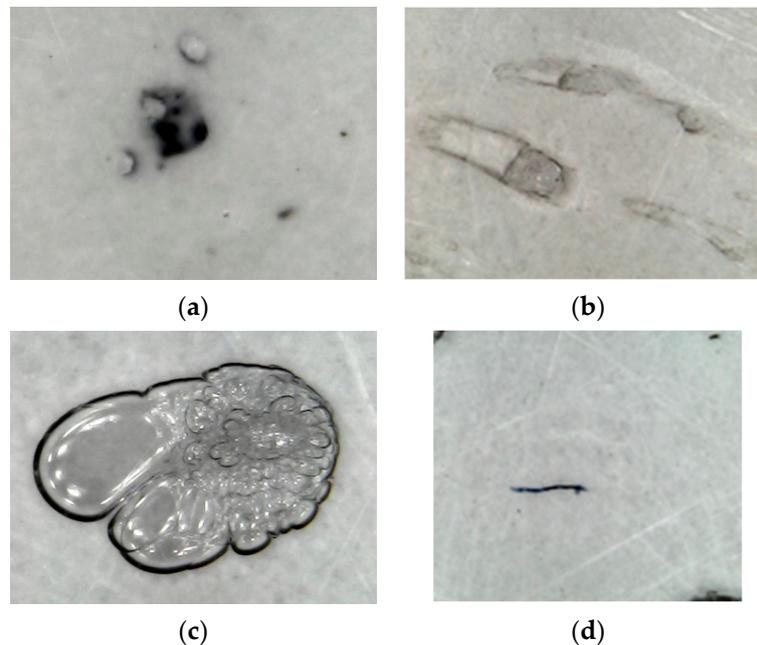
Figure 2. Plastic resin film under the microscope. (**a**) Defect, (**b**) scratch, (**c**) bubble and (**d**) dust.

However, although the method exhibited excellent results, some disadvantages were observed under the microscope. First, the microscope FOV was small—we need to capture hundreds of images to compose one plastic resin film. Therefore, this method does not immediately raise efficiency. It is also difficult to scan the plastic resin film completely without missing any parts; we have no method to consider whether any parts have gone unscanned. Second, identifying defects and determining the corresponding positions is difficult under the microscope. The aforementioned method is not suitable for our study because our final goal is to increase the number of inspected samples. Therefore, we proposed a machine-assisted method with a 2K-resolution camera. One plastic resin film can be captured completely by one picture using the machine; in other words, the machine can speed up the process and increase the number of samples that can be inspected in a day. Figure 3a,b depicts the transparent-background and white-background plastic resin film surfaces under a 2K-resolution camera. The transparent and white plastic resin film results vary considerably under the 2K-resolution camera compared with the data from the microscope. The defects depicted in Figure 3(a1,b1) are mostly small and dot-like, whereas the scratch and dust are shown in Figure 3(a2,b2) and Figure 3(a4,b4), respectively, are similar to lines. Bubbles depicted in Figure 3(a3,b3) are circular. The difference between each type of nonbackground objects is not as obvious in the microscale version. However, only a few images are required for composing one plastic resin film.
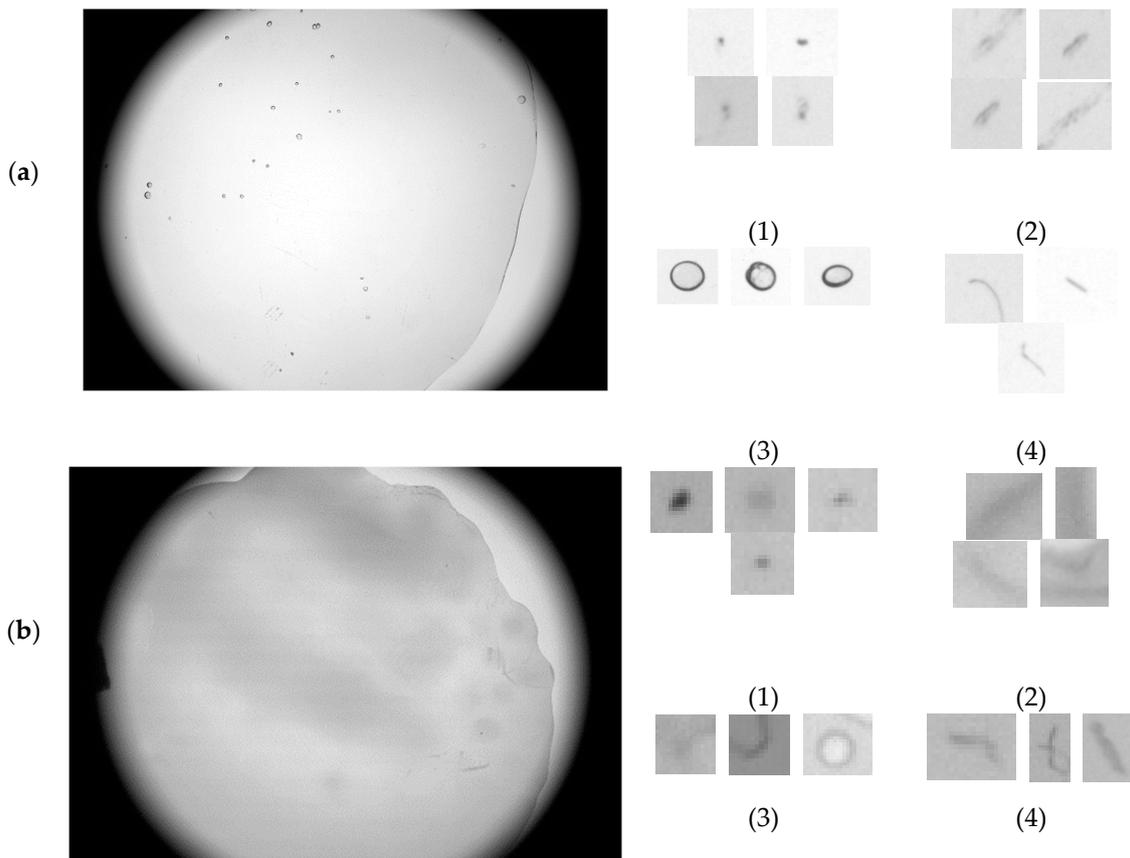
Figure 3. Plastic resin film under a 2K-resolution camera. (**a**) White plastic resin film, (**b**) transparent plastic resin film, (**a1,b1**) defects, (**a2,b2**) scratches, (**a3,b3**) bubbles, (**a4,b4**) dust.

A microscale defect inspection architecture that could automatically identify defects for plastic resin films has been presented in this paper. The proposed method on the 2K-resolution camera was modified based on the microscope method. The microscale defect inspection architecture consists of two steps. In the first part of the architecture, an image preprocessing method for the detection part, segments, and local nonbackground objects was proposed. In the second part, a classification module is used to classify the objects. Compared with traditional procedure, the proposed method has the following advantages:

1. With the same standards for defect detection, the proposed method can obtain the precise area and location of defects.
2. The proposed method has high precision and high speed, which speeds up the inspection process and reduces labor cost. The average amount of plastic resin films that can be checked in a day increases.

The rest of paper is organized as follows: In Section 2, the overall system and proposed approach are described in detail. Experimental results are presented in Section 3 and the discussion is presented in Section 4. Finally, the conclusion is provided in Section 5.

2. Materials and Methods

In this section, the proposed method is discussed in detail. The microscope method is described in Section 2.1, and Section 2.2 describes the method using a 2K-resolution camera. The 2K-resolution camera method is based on the microscope method with some adjustments.

2.1. Microscope Method

Figure 4 illustrates the overall architecture of the method under the microscope. The method includes two stages, namely image processing and classification. During the image processing stage, the traditional image process was used for detecting nonbackground objects, then nonbackground objects that were considered as the input of the classification model were extracted and cropped. In the classification stage, a CNN was proposed to classify various types of nonbackground objects. The system overview is presented in Section 2.1.1, and the image processing stage is depicted in Section 2.1.2 and the classification stage is in Section 2.1.3.
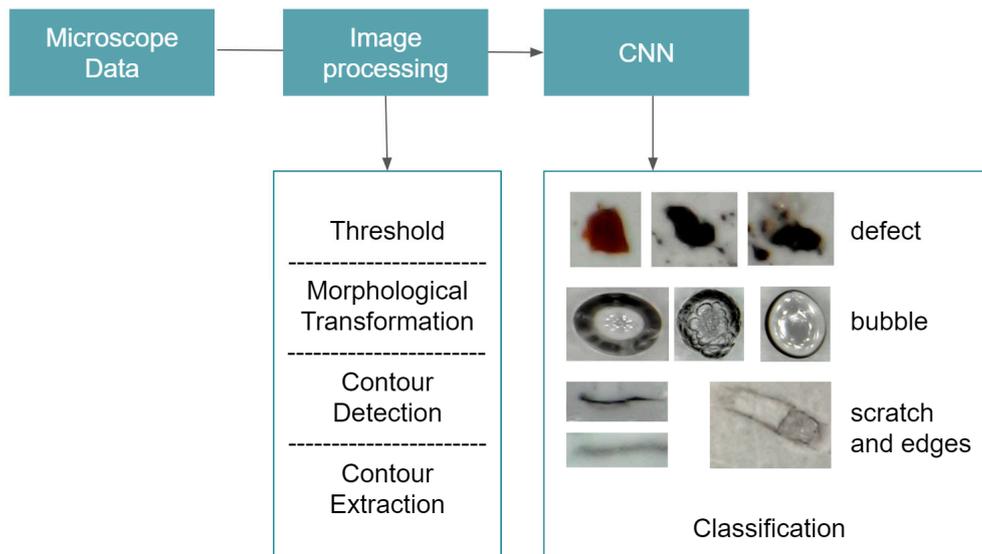

Figure 4. Architecture of the method under the microscope.

2.1.1. System Overview

Figure 5 depicts the sketch map of the inspection process under the microscope. The image obtained by the microscope was sent to the computer and processed immediately. Because of its small FOV, the slide should be moved through all parts of the plastic resin film.
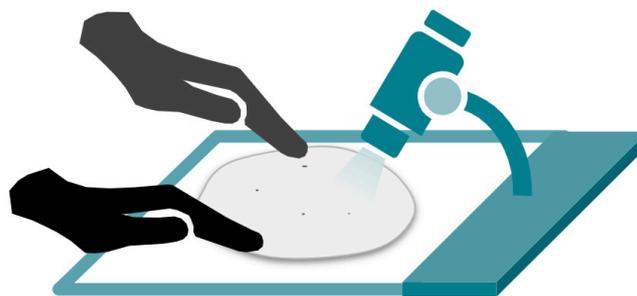

Figure 5. Sketch map of the inspection process under the microscope.

2.1.2. Image Processing

The images captured by the microscope can contain more than one type of object. Therefore, rather than using the whole image to train the classifier, objects should be detected and cropped out first. To segment and crop the nonbackground objects, the image processing methods are required in

this method: Threshold, blur transform, morphological transformations, and contour extraction. In the threshold part, adaptive thresholding method was used. The formula is presented as follows:

$$\text{dst}(x,y) = \begin{cases} \text{maxValue}, & \text{if } \text{src}(x,y) > T(x,y) \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where the threshold value *T(x,y)* is a mean of the blocksize × blocksize neighborhood of (x,y) minus some constant c. Before applying adaptive threshold, Gaussian blur was used on the grayscale image for improving performance. We did several experiments with different combinations of the parameters, observed the data to find the most suitable parameters for our plastic resin film, afterwards obtaining the threshold image. The experiment details are provided in Section 4.1.

The threshold image was then passed through a morphological transform, using the closing technique for closing small holes inside the foreground objects. Closing is dilation followed by erosion. Dilation increases the white region parts in the binary image, whereas erosion is the opposite and erodes them away. Dilation and erosion are defined as follows:

$$\text{dilation} = A \oplus B = \bigcup_{b \in B} A_b \tag{2}$$

$$\text{erosion} = A \ominus B = \bigcap_{b \in B} A_{-b} \tag{3}$$

where A is our threshold image and B is our structuring element, A_b is the translation of *A* by *b* and A_{-b} denotes the translation of *A* by −*b*. Closing is typically useful in closing small holes inside an object and connecting broken areas. Closing can successfully connect broken areas on our data and assist in determining nonbackground objects.

During contour detection, the findContours function in OpenCV was used to locate objects, which extracts objects with areas more than certain pixels. In this study, 25 µm was set the defect threshold. Objects were extracted by their bounding box and cropped out of the image for further steps. Figure 6 depicts the image processing stage.
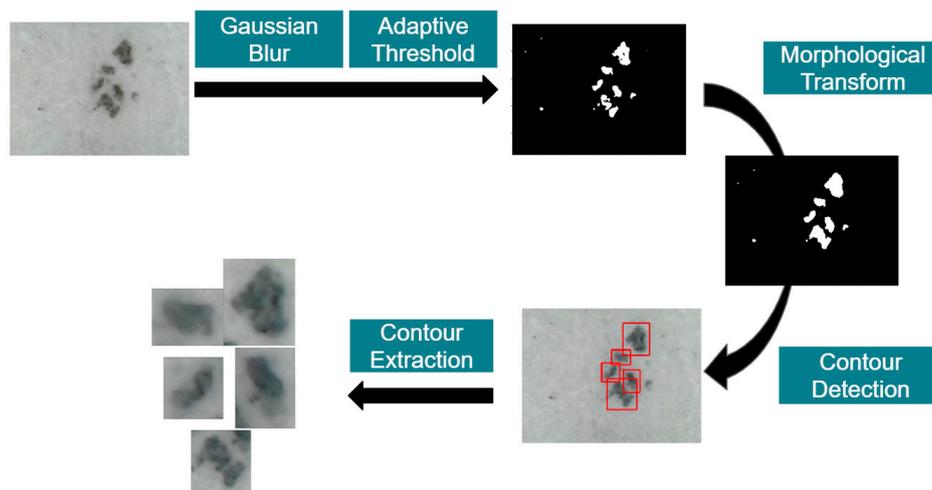
Figure 6. Overview of the image processing stage of the method under the microscope.

2.1.3. Classification

CNN learns the hierarchy of features from the input image and is commonly used in image classification, object detection, natural language processing, among others. A CNN architecture typically consists of some convolution and pooling layers that are fully connected at the end. The convolution layer is the first layer that is used extracting features from the image; filters are defined to

perform multimatrix convolution with the image and obtain the feature map. The pooling layer is used for reducing the number of parameters; the number of feature reduction is defined by the filter size. Pooling is of three types, namely max, average, and sum pooling, depending on the type of values that should be maintained. The fully connected layer flattens the results into the most basic neural network. Usually, a softmax activation function is used at the end for classification to output the probability of the class.

In the classification stage, the LeNet-5-adjusted network was proposed; objects are classified into their categories because of the observed defects having different colors. Colors do not provide information of the classification model; therefore, data is first transferred to grayscale images before training on CNN. The cropped out objects are resized to a size of 50 × 50 pixels. The 50 pixels size was determined by the average size of the data. In the CNN model, the input layer limits the data size, therefore we determine the size of the input layer by averaging. If the average is used, the overall difference of each contour will not be too large, hoping to keep their feature. Figure 7 illustrates the architecture of the classification module.
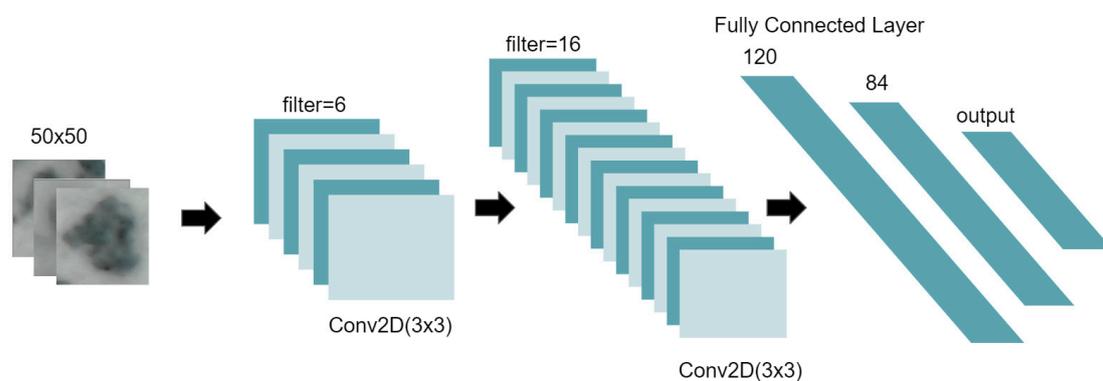

Figure 7. LeNet-5-adjusted CNN network.

The size and style of the target image are similar to the MNIST dataset (Modified National Institute of Standards and Technology database) [34], therefore, LeNet-5 [34], with excellent performance in the MNIST dataset is used as the baseline creation model in the microscale defect detection. Average pooling can result in several features being disregarded. However, all features are crucial for our data. Therefore, average pooling was discarded and superior results were obtained in the microscale defect detection.

2.2. 2K Resolution Camera

Because of the small FOV of the microscope, not only are we required to slide the plastic resin film many times to capture the complete image but we must also be careful not to miss any spot of the plastic resin film. It is difficult to determine accurate positions of the defects on the plastic resin film. Even if we know where a defect approximately is, human vision may not be able to perceive it, because it is too small for detection using human vision. The microscope method is not sufficiently efficient for the plastic resin film because it is time consuming. Therefore, we proposed a machine-assisted method with a 2K-resolution camera. Although accuracy using the microscope method is higher, the proposed method is more efficient than the microscope method and it is considerably easier to locate the defects on the plastic resin films. Only one image is required to be captured to obtain the complete plastic resin film. Figure 8 depicts the architecture of the proposed method using a 2K-resolution camera. This method consists of three stages, namely image adjustment, image processing, and classification. The image adjustment stage improves performance during comparison using the image processing stage. The image processing stage and classification stage are mostly the same that in the microscope method, with the exception of a comparison part to filter noise and reduce the amount contours.
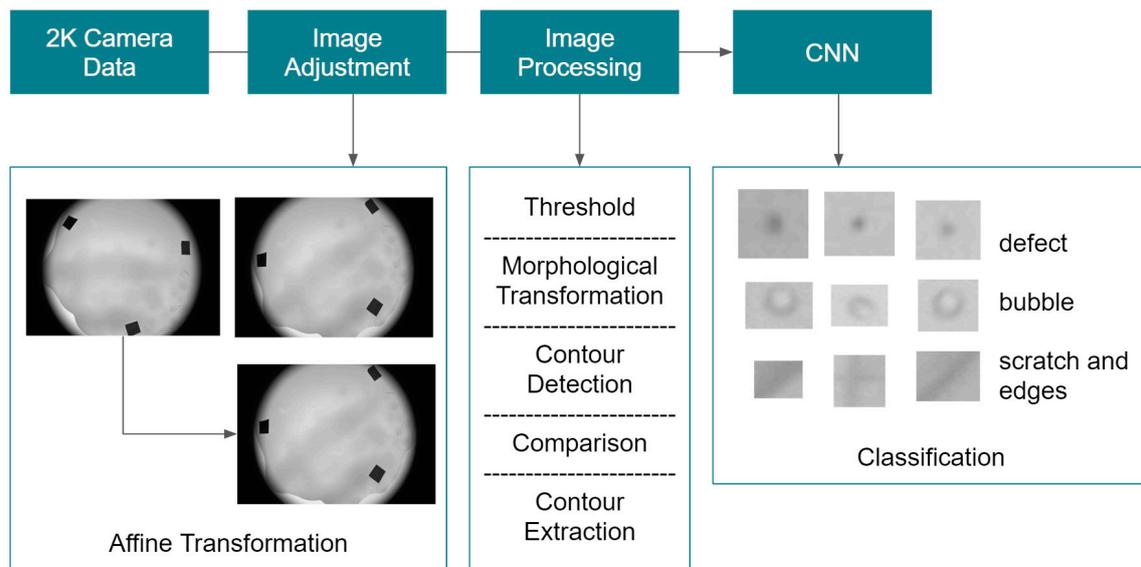
Figure 8. Architecture of the method with the 2K-resolution camera.

2.2.1. System Overview

Figure 9 depicts the sketch map of the 2K-resolution camera. The key component of the system is the VS-LDA series lens from vs. technology corporation [35], which has a 2K resolution (5472 × 3648 pixels) and has been fixed at a shooting distance of 12.5 cm on our machine. This series lens is best known for its low distortion, even when using extension tubes and is designed to support a wide range of magnification, wide angle (*WD*), and depth of field. The resolution of the camera can reach 762 dpi (30 pixels/mm), which satisfies more than 25-μm defect detection requirement. A light regulator was used to control light sources to ensure the same light condition was used throughout the study. This minimized the influence of light on the data. A dust removal fan was used to filter dust on plastic resin films.
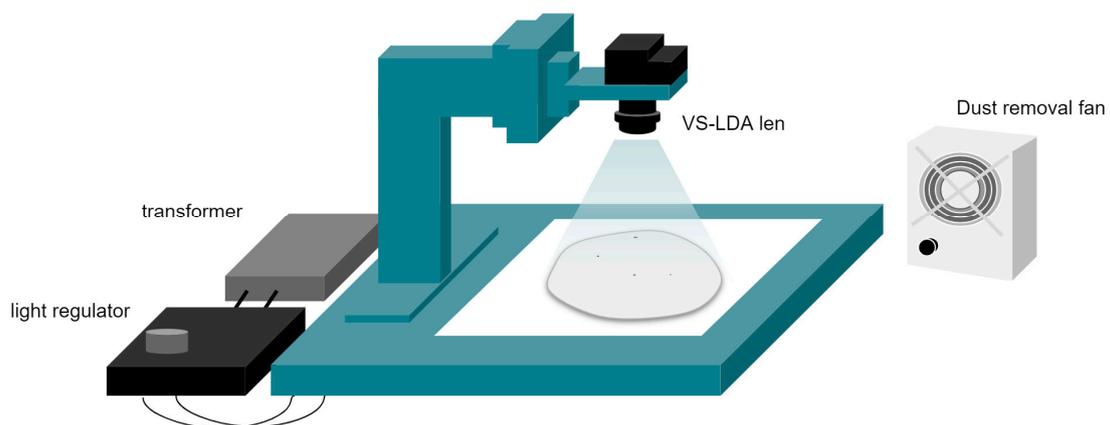
Figure 9. Sketch of the inspection process using the 2K-resolution camera.

2.2.2. Image Adjustment and Image Processing

Normally, the inspection process of the plastic resin film is not conducted in a dust-free environment. Therefore, the plastic resin film can contain dust. Thus, identifying defects is difficult because dust and defects (scratches and edges) appear similar under the 2K-resolution camera. This can decrease the classification performance. Therefore, a comparing component was introduced to filter noise and avoid dust before contour extraction. In this approach, three images on the same plastic resin film

were used, the contours of each image were determined and compared, and a fan was used to remove dust each time before capturing an image.

If the contour's location was different than that of other contours, then the contour was mostly dust or fiber. The comparison between the contours must be in the same location or at least close, we used Intersection of Union (IoU) to examine whether they are the same object, which is defined as follows:

$$\text{IoU} = \frac{Area\ of\ Overlap}{Area\ of\ Union} \tag{4}$$

when we evaluate the IoU of the bounding box of the detected objects, the bounding box is defined during contour detection. We defined IoUs that are more than 0.3 as the same object. However, it is impossible to capture three images at the same position. Manual adjustment of the picture to the exact position is time consuming and reduces efficiency. Therefore, affine transformation is used. Affine transformation is any transformation that can be expressed through matrix multiplication. Rotation, translation, or scaled operations can be expressed using an affine transformation. Affine transform is the relation between two images, usually shown as a 2×3 matrix. The formula used for image adjustment is expressed as follows:

$$\text{T} = \text{M} \cdot [x, y, 1]^T = \begin{bmatrix} a_{00} & a_{01} & b_{00} \\ a_{10} & a_{11} & b_{10} \end{bmatrix} \cdot [x, y, 1]^T = \begin{bmatrix} a_{00}x + a_{01}y + b_{00} \\ a_{10}x + a_{11}y + b_{10} \end{bmatrix} \tag{5}$$

where M is the affine transform matrix, *(x,y)* is the original image position, and T is the new position after applying the affine transform. To apply the transform to an image, *M* must be defined first. Our goal was to apply the affine transform to transform one image into another image, which required three corresponding positions in the two images. Thus, the relation between the two images can be obtained and can be transferred from one to the other. Three same points can be obtained from both images to achieve the affine transform matrix, the transform can then be applied to one of the images, and transform into the other image. The affine transformation is depicted in Figure 10.
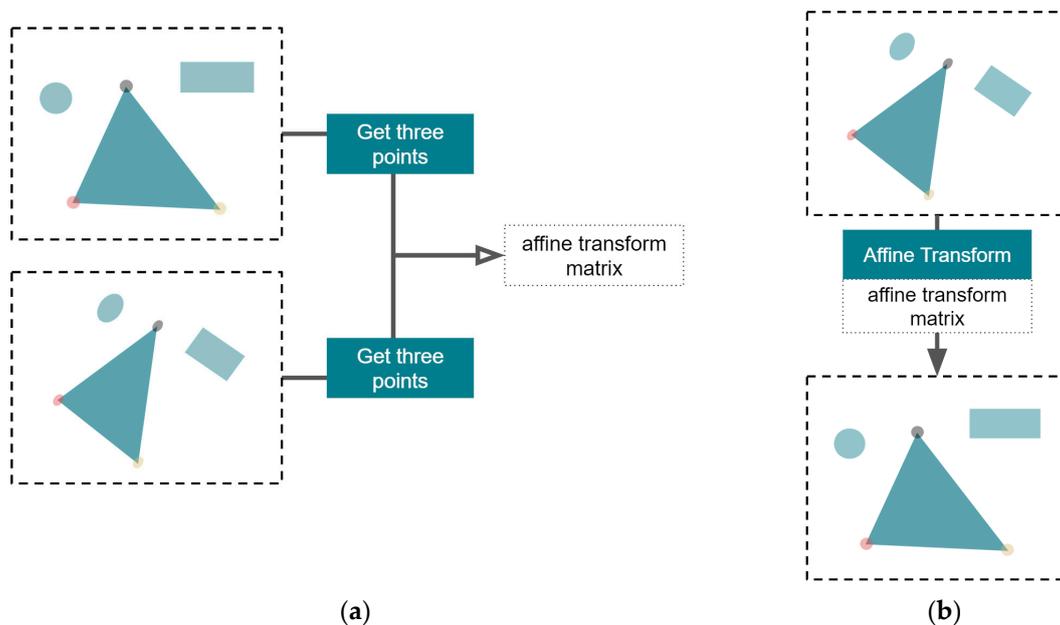
Figure 10. Affine transformation. (**a**) The first step: obtain the affine transform matrix, (**b**) the second step: apply the transform to the image.

Therefore, three images were captured, and the affine transform was applied to two of the images making them exactly the same, and then the locations of the contours were compared to determine whether the objects were dust. To achieve the affine transform matrix, three exact points are required.

Three stickers were placed on the plastic resin film. Some simple thresholding technique and contour detection could easily determine the center position.

After the images had been transformed to the same position, image processing was performed, and the bounding box of the contours in all three images was obtained after contour detection. IoU more than 0.3 was counted as the same contour as mentioned before. The contours were extracted only if all three images have them. The results indicated the method can efficiently filter most of the dust and reduce the number of contours. Figure 11 illustrates the details of both the image adjusting stage and image processing stage.
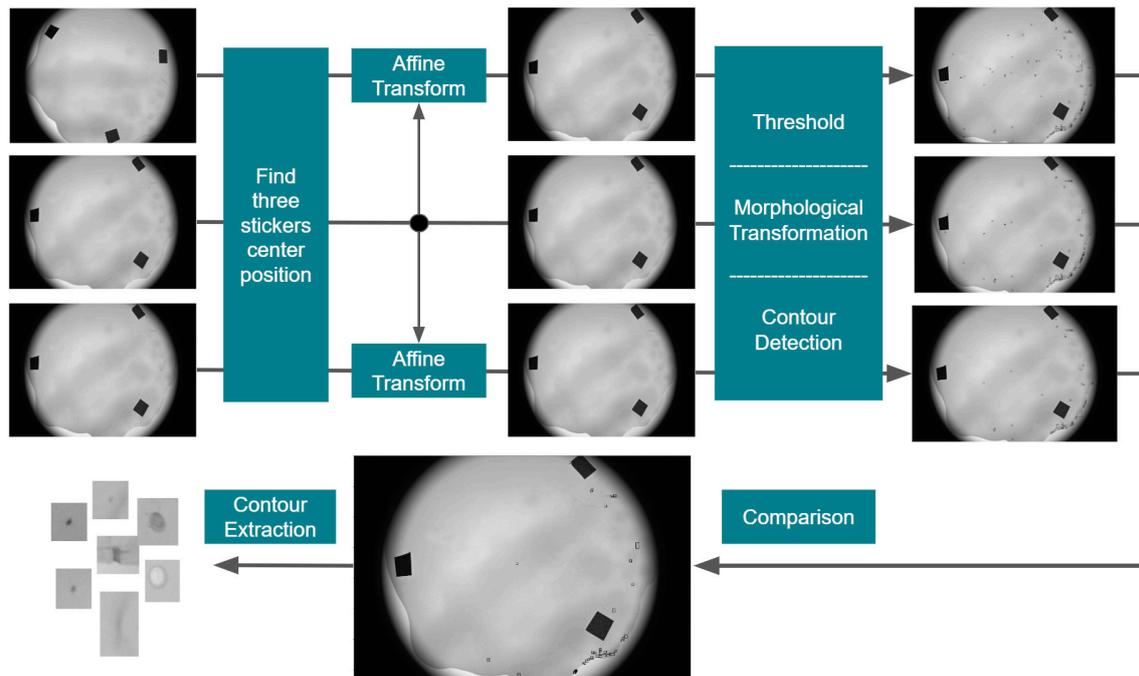

Figure 11. Overview of the image adjusting and image processing stage in the 2K-resolution camera method.

2.2.3. Classification

The LeNet-5-adjusted CNN network was used as the classification model, which is the same as the microscope method. The input size was changed to 52×52 pixels, the size was determined by the average size of the data

3. Experiments and Results

In this section, we discuss the experimental setup before training on the LeNet-5-adjusted network and the parameters that have been used during training. The microscope method is discussed in Section 3.1 and the 2K-resolution camera is discussed in Section 3.2. The recall was used to evaluate the performance of the classification task, which is calculated as follows:

$$\text{Recall} = \frac{T_P}{T_P + F_N} \tag{6}$$

while recall is our main standard, we also want to gain a high-precision F1-score. Their formulas are shown as follows:

$$\text{Precision} = \frac{T_P}{T_P + F_P} \tag{7}$$

$$\text{F1} - \text{Score} = \frac{2 \times T_P}{2 \times T_P + F_N + F_P} \tag{8}$$

T_P (true positive) indicates the number of defects classified correctly, F_N (false negative) indicates defects classified into wrong categories and F_P (false positive) indicates the number of nondefects classified as defects. The research goal is to reduce the number of suspicious defects of human judgment, but also not to overkill the true positive—therefore, the recall needs to be high. However, while the recall needs to be high, we also want to prevent having too low precision. Therefore, we would also consider precision and F1-Score.

3.1. Microscope Method

3.1.1. Experiment Setup

For the adaptive threshold parameters, the block size was set to 205 and the constant *c* was set as 30 during the image processing stage by observing the defects and confirming if all the defects were captured. Contour extraction only extracts objects with areas larger than 400 pixels because under the microscope, a 25-μm defect is approximately that size.

The dataset of plastic resin film images was provided from a production line of a plastic component using a microscope. All the defect components were previously inspected using an expert examiner. During the image processing stage, we cropped 237 defects and 2457 bubbles because such irregularities would undermine the training of the neural network. Therefore, some modifications were incorporated, including rotation and flipping. In this case, the aforementioned operation increased the number of defects to 2796.

For the classification task, 100 images were used for testing and the remaining images were used for training. The LeNet-5-adjusted network was trained for 100 epochs and the batch size was set to 100.

3.1.2. Evaluation of the CNN Model

To verify the performance of the proposed CNN, in this section, we compared the inspection results of this model with those of other CNN networks. The Arikan et al. [22] proposed SURFnet, which was inspired by the VGG network [36] configurations and residual learning [37]. It has nine convolution layers and one fully connected layer. Each convolution layer contains a batch normalization and using the parametric rectified linear unit as the activation function. The LeNet-5 architecture consists of two pairs of convolution and average pooling layers, followed by a flatten layer and two fully connected layers. Both network epochs were set to 200 and the batch size was set to 32 during training.

Table 1 depicts the performance of the defect class with various networks. We mainly focused on whether all defects were detected, evaluating the recall of the defect class. The results prove that LeNet-5 has high precision, and their recall is the lowest. This proves that superior performance can be obtained if the average pooling layer was removed. This is because extracted objects may be only approximately 1 pixel. Average pooling can cause loss of vital features. For SURFnet, although the precision is high (up to 1), the recall is not as high as the LeNet-5-adjusted network.

Table 1. Performance of classification with different CNN networks of the defect class.

Network	Precision	Recall	F1-Score
SURFnet	**1.00**	0.84	0.91
Lenet-5	0.93	0.81	0.86
Lenet-5-adjusted	0.91	**0.91**	**0.91**

3.2. K Resolution Camera Method

3.2.1. Implementation Detail

The inspection experiment system was developed using Python 3.6.7, with Keras as the deep learning platform. The following results were obtained using a server with Intel(R) Core(TM) i7-8700K CPU and NVIDIA TITAN-V with 11 GB of memory as graphic processing unit.

3.2.2. Experimental Setup

During the image adjustment stage, the center location of the three stickers was obtained by applying some simple threshold transform and object extraction. The adaptive threshold parameter was selected by observing the defects and confirming if all the defects were detected. We set the blocksize to 55 and the constant c to 19. For contour extraction, we extract objects that areas are above 1 pixel while limiting the size under 500 pixels since the size of 25 μm under the 2K-resolution camera is approximately that size and all the observed defects are under 500 pixels.

Table 2 lists the average number of objects in one plastic resin film before and after comparison. The result indicated that it can successfully reduce most of the dust and a considerable number of objects. More objects were observed against the transparent background than against the white background because more bubbles were present in it.

Table 2. The average number of objects observed in one plastic resin film before and after comparison.

Type of Plastic	Before	After
Transparent background	260	172
White background	54	23

Because the transparent- and white-background plastic data was considerably different on the 2K-resolution camera, we trained them separately. Both networks were trained for 200 epochs, and the batch size was set to 32 on the transparent-background plastic and 16 on white-background plastic. The number of output classes were different. The transparent background has three classes, namely bubbles, defects, and scratches, whereas the white background only has two classes, combining bubbles and scratches into one. For the training process, we considered 177 bubbles, 31 defects, and 37 scratches and edges on a transparent background, and 72 bubbles and scratches, 32 defects on a white background. Some augmentations were made to balance the data.

3.2.3. Evaluation of the CNN Model with Feature Selection Algorithm

To verify the performance, we compare our CNN model with some traditional methods, in our experiment we use two feature selection algorithms with SVM or MLP. Histogram of oriented gradient (HOG) is for counting occurrences of gradient orientation in localized portions of an image while local binary pattern (LBP) is a powerful feature for texture classification. The quantization of the gray values in HOG is twelve and the number of circularly symmetric neighbor set points for LBP is eight. The MLP consists two hidden layers, the first layer contains of 12 units while the second contains 8. For the transparent-background data, the output layer contains two output variables while the white-background data contains three output variables. The MLP model were trained for 50 epochs, and the batch size was set to 32. Table 3 displays the performance of the transparent-background plastic and Table 4 displays the performance of the white-background plastic.

As shown in Tables 3 and 4, the LBP + SVM and HOG + SVM mostly have low recall, not only that, their precision and f1-score did not show good results, either. In Table 3, while LBP + MLP has a rather good result on the transparent-background data compare to the other methods, our CNN models still has better results. HOG + MLP has a better recall than our model, however their precision is too low. In Table 4, three of the methods have the same result on the recall with our CNN model; however, the

precision is lower than ours. It shows that it is difficult to fully classify the features of defects only by gradient features and a classifier. Our method is better compared to the feature selection algorithms.

Table 3. Performance of the defect class on transparent-background plastic with various feature selection algorithms.

Method	Precision	Recall	F1-Score
LBP + SVM	0.54	0.77	0.64
HOG + SVM	0.55	0.69	0.61
LBP + MLP	0.70	0.80	0.75
HOG + MLP	0.47	**0.89**	0.61
Lenet-5-adjusted	**0.77**	0.84	**0.80**

Table 4. Performance of the defect class on white-background plastic with various feature selection algorithms.

Method	Precision	Recall	F1-Score
LBP + SVM	0.36	**0.97**	0.62
HOG + SVM	0.40	0.63	0.49
LBP + MLP	0.42	**0.97**	0.59
HOG + MLP	0.63	**0.97**	0.76
Lenet-5-adjusted	**0.78**	**0.97**	**0.86**

3.2.4. Evaluation of the CNN Model with other CNN Networks

To verify the performance of the proposed CNN network, in this section we compared the inspection results of this model with those of other CNN networks. Table 5 lists the performance of the transparent-background plastic and Table 6 displays the performance of the white-background plastic.

Table 5. Performance of the defect class on transparent-background plastic with various CNN networks.

Network	Precision	Recall	F1-Score
Lenet-5	**0.84**	0.74	0.78
SURFnet	0.78	0.84	**0.80**
Lenet-5-adjusted	0.77	**0.84**	**0.80**

Table 6. Performance of the defect class on white-background plastic with various CNN networks.

Network	Precision	Recall	F1-Score
Lenet-5	0.45	0.90	0.60
SURFnet	**0.83**	**0.97**	**0.89**
Lenet-5-adjusted	0.78	**0.97**	0.86

As depicted in Table 5, the original LeNet-5 does not exhibit high recall (0.74), whereas both SURFnet and the proposed CNN exhibited a recall of 0.84. Table 6 lists the same concept. Both transparent and white plastic on LeNet-5 do not perform better recall, and the precision on the white background is notably low, and even though SURFnet on both plastics had the same recall compared with the proposed CNN, the CNN structure of SURFnet is considerably more complex.

4. Discussion

During the development of the proposed method, a number of experiments were performed to deliver high performance. In this section, details during experimentation are discussed. The image processing stage is discussed in Sections 4.1 and 4.2 and the hyperparameters during the training of the CNN model are discussed in Section 4.3. Finally, implementation issues are discussed in Section 4.4

4.1. Threshold

When selecting the suitable thresholding method, the Otsu method and the basic thresholding method did not provide a satisfactory performance on both kinds of data. On the microscope data, the low-quality image causing too much noise on the image, therefore Otsu did not performance great. As for the 2K-resolution camera data, the objects were small and not sufficiently obvious for Otsu to obtain. The basic threshold, in which a pixel value was selected and set as the boundary line did not provide satisfactory results because the microscope data and 2K-resolution camera has varying light at the center. Therefore, the adaptive threshold is the most suitable approach among others. There are different kinds of adaptive thresholding method, such as mean, Niblack [38] and Sauvola [39]. The adaptive threshold sets the boundary line as shown in Equation (3), different adaptive thresholding methods determine the $T(x, y)$ value differently. The Niblack and Sauvola formula is shown as below.

$$T(x, y) = m(x, y) \cdot K \cdot s(x, y) \tag{9}$$

$$T(x, y) = m(x, y) \cdot \left[1 + K \cdot \left(\frac{s(x, y)}{R} - 1 \right) \right] \tag{10}$$

where m (x,y) is the mean of the blocksize × blocksize neighborhood, s (x,y) is the standard deviation. The parameter K gets positive values, and R is the dynamic range of standard deviation. The results after using different thresholding methods are shown in Figure 12. As shown in Figure 12a,b, nonuniform light source causes bad performance on the basic thresholding method. In Figure 12c, Otsu cannot get any objects on the plastic resin film. Although Niblack and Sauvola shows better performance compared to the two methods mentioned above, these methods gain a lot of unrequired noises, making it difficult for the contour detection afterwards. In addition, many objects are fragmented while using Sauvola. Their computation time is also rather slow compared to the mean adaptive threshold. Therefore, we chose mean adaptive threshold.

(**a**)

**Figure 12.** *Cont.*

(**b**)

(**c**)

(**d**)
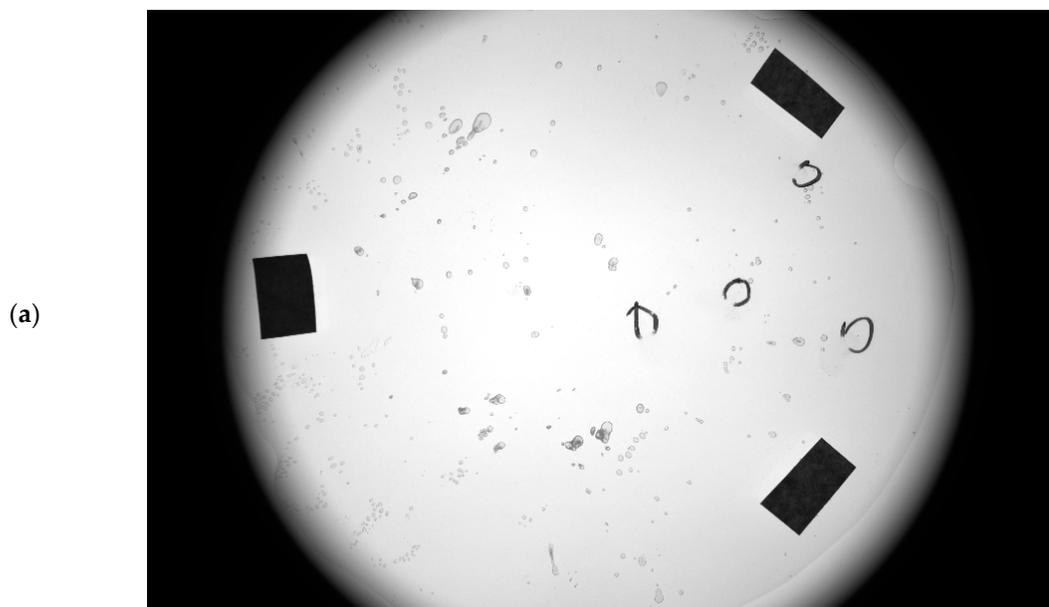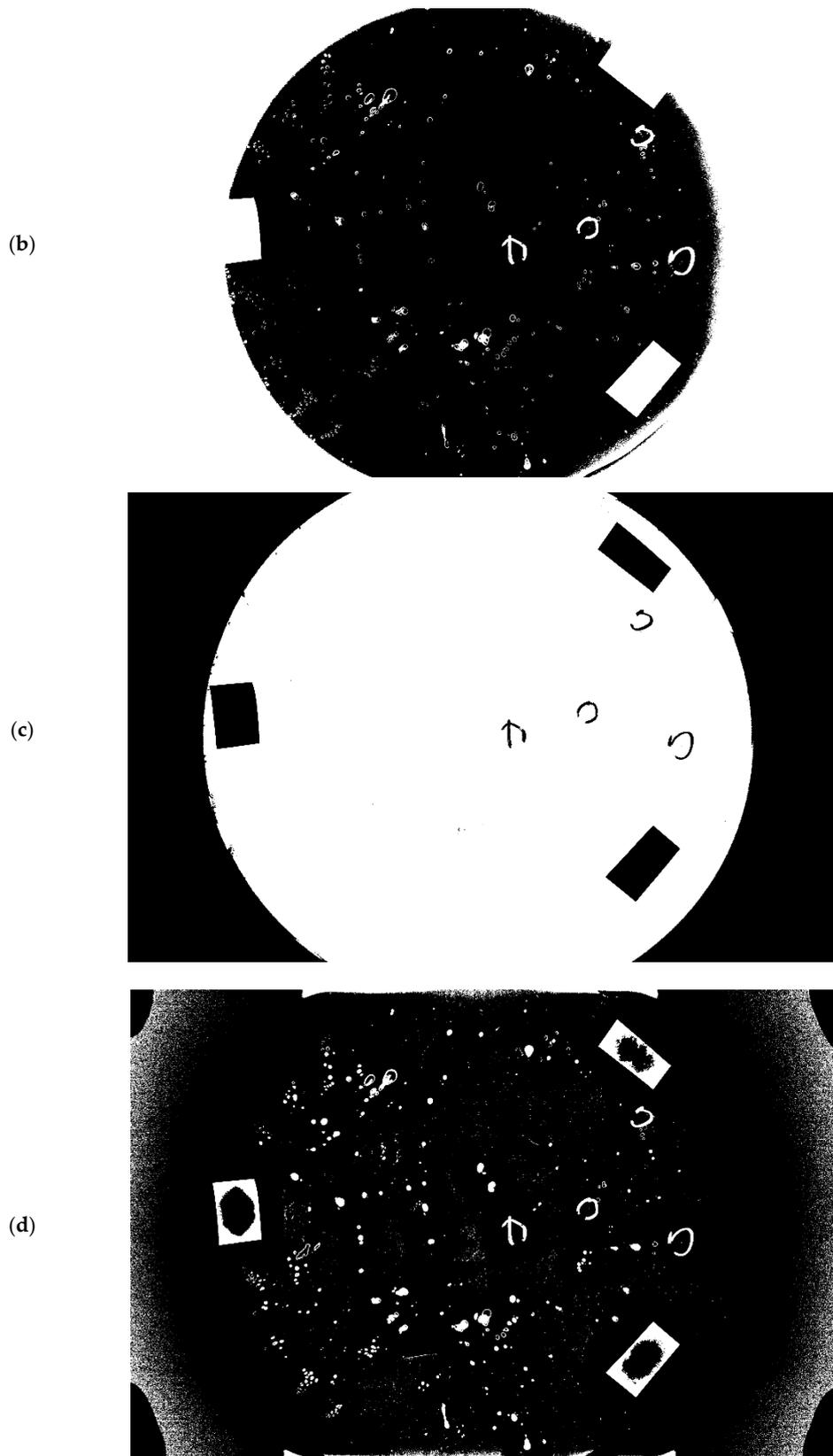
Figure 12. *Cont.*
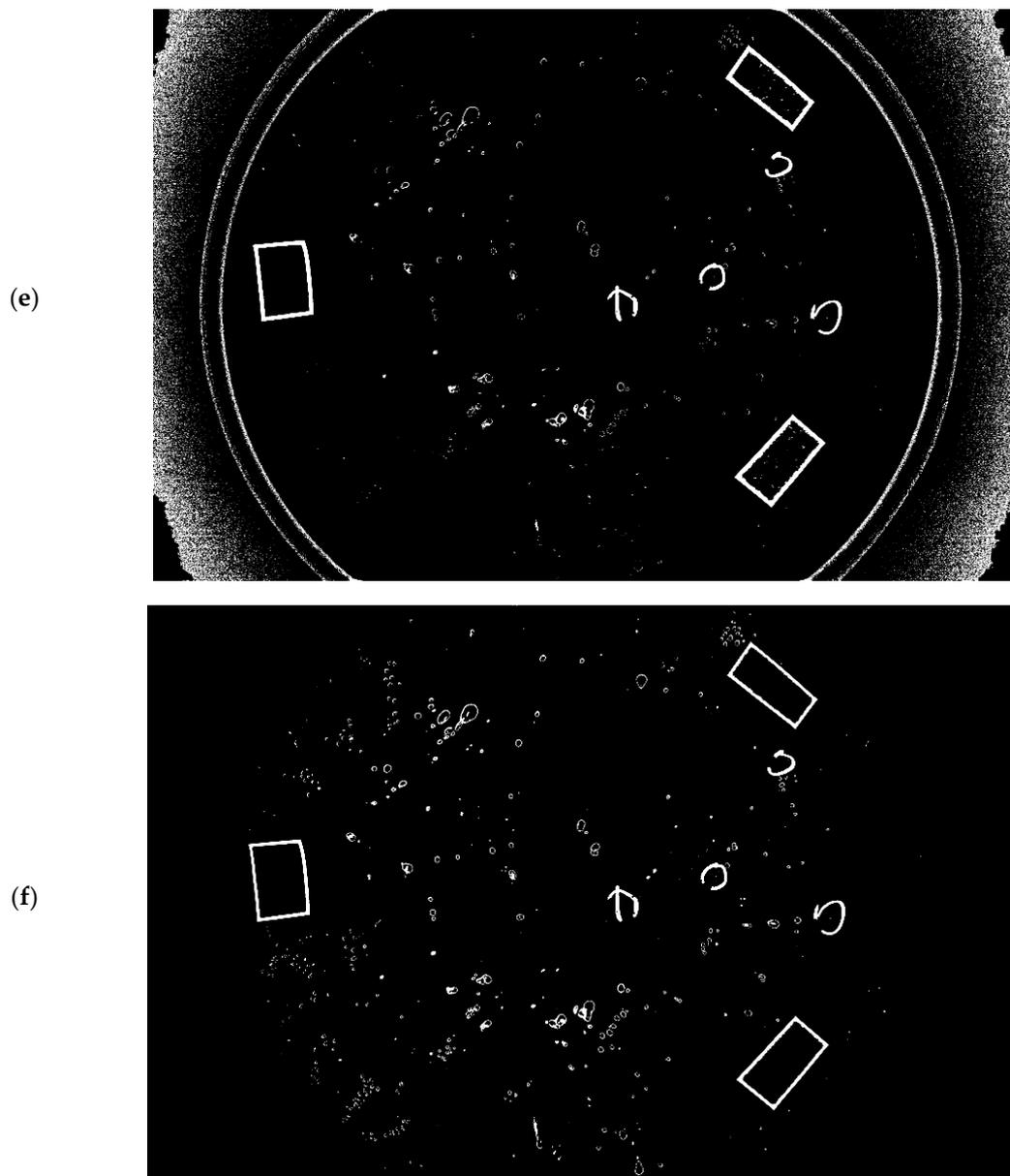
Figure 12. Results on different thresholding methods. (**a**) Origin image, (**b**) global threshold, (**c**) Otsu, (**d**) Niblack, (**e**) Sauvola, (**f**) mean.

However, the parameters for the mean adaptive threshold must be adjusted because of the data difference and light effect, which requires time. Finding the best parameters for the adaptive threshold consumed the most amount of time in the experiment. We mainly attempted the combination and compared each of them. Figure 13 shows the comparison between different parameter values.

Here, x, y are the parameters of the adaptive threshold where x is the constant c and y is the blocksize, the total number of the missing ground true defect are depicted in Figure 13a and the total number of contours that were found are illustrated in Figure 13b. The figure shows that only the parameter with blocksize 55 and c 5 has gotten all the defects, however while our goal is to find all the defects as many as possible, we also want to reduce the number of objects being detected. Therefore, we observe the data under four different combination; each combination has a different amount of missing ground true, the objects being detected on the threshold image are shown in Figure 14. It is obvious that blocksize 55, c 5 is not a suitable parameter; as shown in Figure 14d, it is not suitable for finding the objects on the plastic resin film, Figure 14b will detect a lot of noises. Other than that due to

observation, the parameters used on Figure 14c would cause some same scratches being detected as different objects, these sub-scratches have similar features as the defects, which would cause it hard to classify between defect and scratches. Therefore, because of the aforementioned results, a blocksize of 55 and a constant *c* of 19 were set.
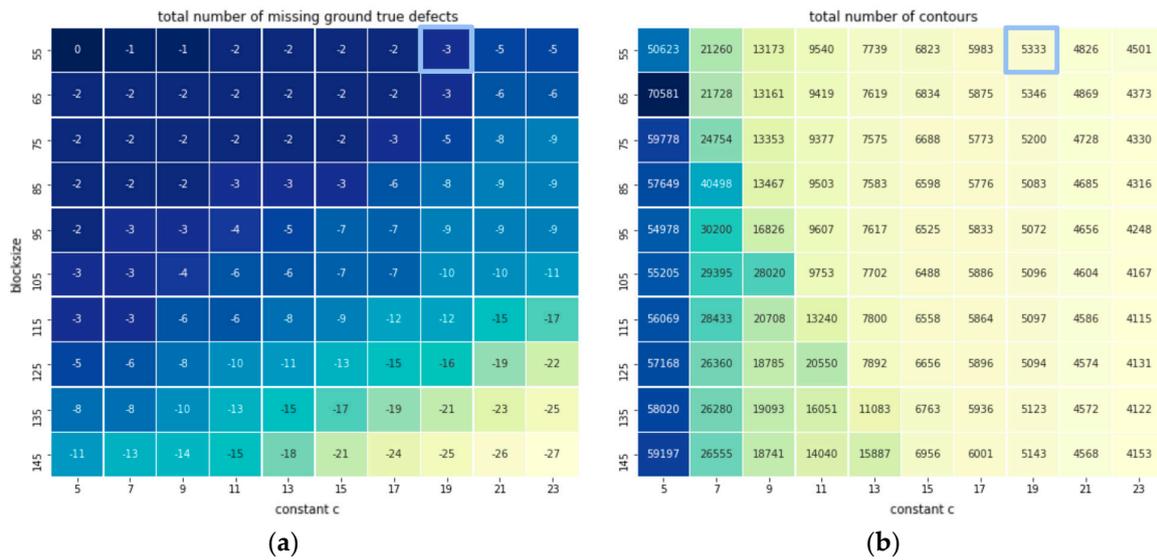
Figure 13. Different combinations of the adaptive threshold parameters on the 2K-resolution camera: (**a**) The total number of missing ground true defects, (**b**) The total number of contours.

Parameters	Blocksize 55 × 55, c 19	Blocksize 55 × 55, c 17	Blocksize 55 × 55, c 9	Blocksize 55 × 55, c 5
Number of missing defects	3	2	1	0
Case1				
Case2				
	(**a**)	(**b**)	(**c**)	(**d**)

Figure 14. Contour detection result using different adaptive threshold parameters. (**a**) Blocksize 55 × 55 pixels, c 19 (**b**) Blocksize 55 × 55 pixels, c 17 (**c**) Blocksize 55 × 55 pixels, c 9 (**d**) Blocksize 55 × 55 pixels, c 5.

4.2. Contour Extraction

During contour extraction, when the minimum bounding box of the contours in the microscope data, we did not apply the minimum on the camera data. This is mainly because we detected the possibility of objects with area less than 10 pixels. The minimum bounding box that was used to extract the object would then be difficult to train during the classification stage.

As depicted in Figure 15, if we selected to extract the object by the minimum bounding box of the contour, limited background details could be compared. Because our input for the LeNet-5-adjusted network must be of the same size, loss of attribute features occurred during resizing, which caused the accuracy to decrease during classification. In this case, we expanded the bounding box. The extracted objects attribute features were more obvious (Figure 14b), we expanded the area to approximately 5, 10, and 15 pixels; then we trained them on the LeNet-5-adjusted network. The best results were obtained for 15 pixels during the classification stage.
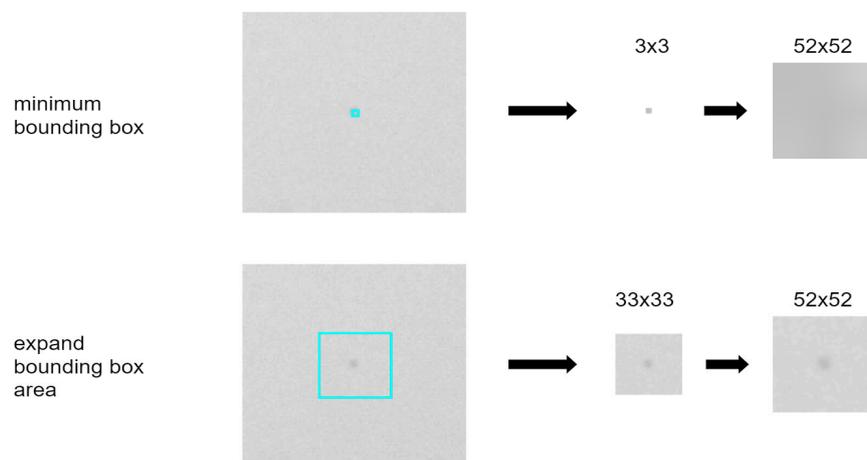

Figure 15. Data difference in depicting the difference between expanding the bounding box area or not.

4.3. Hyperparameter

Fine adjustments were performed under the light camera. When creating the model, dropout and pooling layers were used. However, these did not have a considerable effect. The batch size was switched between 16, 32, and 64. Excellent results were obtained for the white-background plastic with a batch size of 16, whereas excellent results were obtained with a batch of 32 for transparent background. During tuning, we added different poolings, but did not achieve better results than that using the nonpooling model. The use of the dropout layer did not achieve better results on the data.

4.4. Implementation Issues

Camera data issue: 2K resolution does not have sufficient quality for the plastic resin film; the nonbackground object is still considerably blurry under the camera and sometimes it is difficult to distinguish between a bubble and a defect. Many defects were less than 5 pixels across and the CNN could not be trained to detect them. Raising the platform and lowering the FOV, the defects increased to 10 to 15 pixels and exhibited a superior result in classification. However, three images were required to compose one plastic resin film. This leads to other problems such as localization of the defects in one combined image and ensuring every part of the plastic resin film is captured in the three images. Therefore, we did not raise the height of the platform, and although classification was not better and the nonbackground objects were not clearer, it was considerably efficient during the inspection process.

Data labeling: Considerable time was required for training the data, as every object was cropped to be classified as the right class, most of the time microscope was required to classify the object into the correct class. Furthermore, expanded area of 5, 10, and 15 pixels exhibited different data. This is

time consuming and careful scrutiny is required at this stage because this strongly affects the efficiency of our classification model.

　　During the classification of the camera data, unstable results were obtained for the white background when divided into three categories, and bubbles and scratch would be categorized as defects as shown in Table 7a and Figure 16a. Both models were trained on the same data with the same hyperparameters. Therefore, the bubble and the scratch classes were combined, which exhibited superior results as depicted in Table 7b and Figure 16b.

Table 7. Performance of the Lenet-5-adjusted trained on white-background plastic. (**a**) CNN trained in three classes, (**b**) CNN trained in two classes.

| (a)			(b)		
Type of Class	Precision	Recall	Type of Class	Precision	Recall
Bubble	0.07	0.71	bubble_scratch	0.99	0.91
Defect	0.80	0.93	defect	0.78	0.97
Scratch	0.93	0.16			

(**a**)　　　　　　　　　　　　　　　　　　　　　　　　　　(**b**)
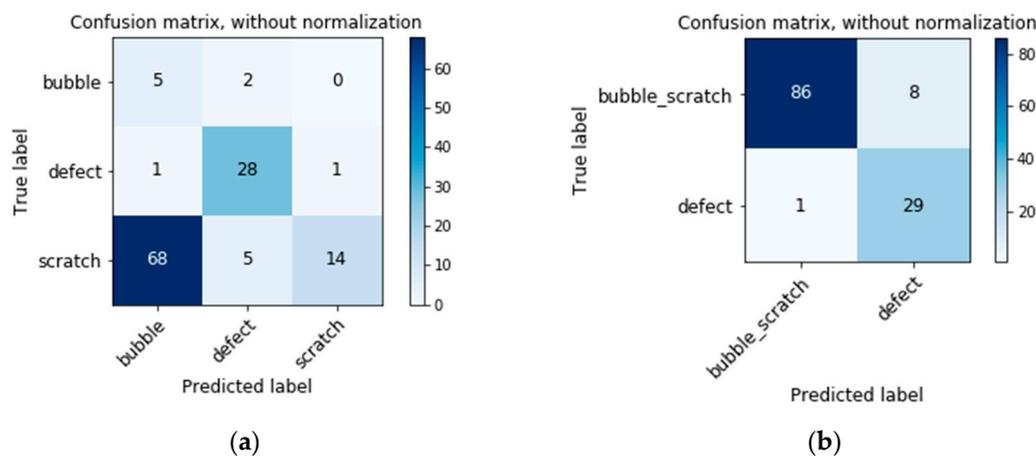
Figure 16. Confusion matrix of the Lenet-5-adjusted trained on white-background plastic. (**a**) CNN trained in three classes, (**b**) CNN trained in two classes.

　　However, while white-background plastic has better results while dividing into two categories, the transparent-backgound plastic does not. As shown in Table 8 and Figure 17, the recall of the defect class is higher when dividing into three categories. Therefore, during training the transparent-background plastic data, we choose to divide the data into three categories: defect, bubble and scratches.

Table 8. Performance of the Lenet-5-adjusted trained on transparent-background plastic. (**a**) CNN trained in three classes, (**b**) CNN trained in two classes.

| (a)			(b)		
Type of Class	Precision	Recall	Type of Class	Precision	Recall
bubble	0.94	0.89	bubble_scratch	0.97	0.97
defect	0.77	**0.84**	defect	**0.79**	0.81
scratch	0.48	0.62			

(**a**)   (**b**)
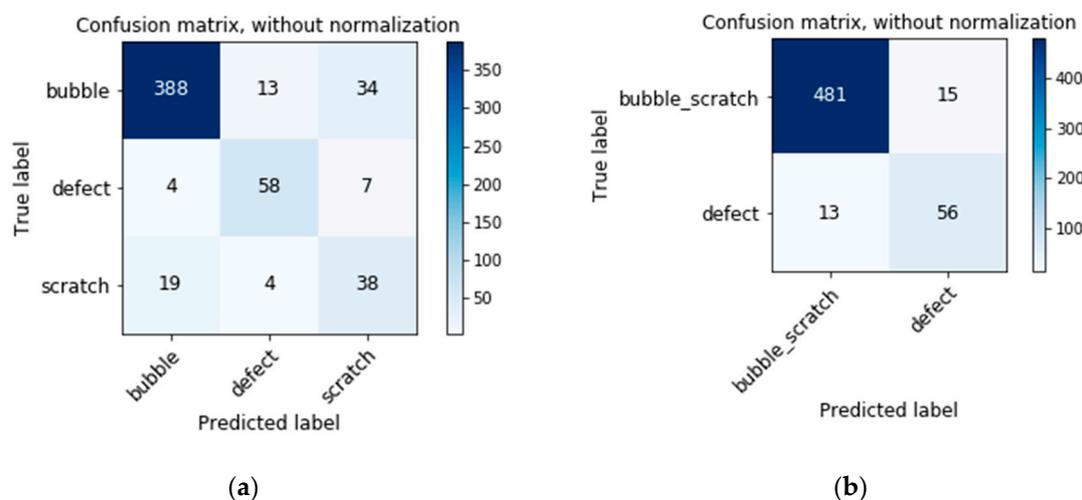
Figure 17. Confusion matrix of the Lenet-5-adjusted trained on transparent-background plastic. (**a**) CNN trained in three classes, (**b**) CNN trained in two classes.

4.5. Increasing Accuracy

To increase time efficiency, the 2K-resolution camera was selected, and although recall is our priority, precision should also increase. Therefore, the data was doubly verified by applying the 2K-resolution camera first, and then using the microscope method.

5. Conclusions

In this paper, an automatic method that could identify nonbackground objects and classify objects was proposed by using a traditional image processing method to detect nonbackground objects, and a CNN network to classify the defects. Based on the proposed method, the inspection was converted to a segmentation and classification problem. Traditional image processing detected objects; a comparison part was used to filter dust on the plastic resin film. The results indicated that the process was successfully able to filter out a sufficient amount of dust. A LeNet-5-adjusted network was proposed for the classification, and the results exhibited excellent accuracy, proving that an average pooling of the data can lead to loss of vital features. The transparent-background data has an 84% recall and a 77% precision, while the white-background data has a 97% recall and a 78% precision. The proposed automatic defect detection with machine vision-based method can improve the disadvantages of the traditional procedure. The proposed method exhibits uniform measurements, unlike the traditional method in which the criteria can vary for each inspector. Whereas human vision for the inspection process is quite time consuming and inefficient, the proposed method is fast. Using the traditional method, the average number of plastic resin films that can be checked by a person in a day is approximately 12 to 15. However, with the proposed method, the amount of time needed to check one plastic resin film can be down to 3 min. In the proposed method, the labor costs of the inspection process can be reduced, the workload of the inspectors can be lightened, and productivity can be increased.

Author Contributions: Methodology, R.-K.S., Y.-H.T. and L.-C.C.; software, Y.-H.T. and C.-H.T.; supervision, R.-K.S. and L.-C.C.; Writing – original draft, R.-K.S., Y.-H.T. and L.-C.C.; Writing – review & editing, R.-K.S., C.-H.T. and L.-C.C. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Tikhe, C.; Chitode, J. Metal surface inspection for defect detection and classification using Gabor filter. *Int. J. Innov. Res. Sci. Eng. Technol.* **2014**, *3*, 13702.

2. Şeker, A.; Peker, K.A.; Yüksek, A.G.; Delibaş, E. Fabric defect detection using deep learning. In Proceedings of the 2016 24th Signal Processing and Communication Application Conference (SIU), Zonguldak, Turkey, 16–19 May 2016; pp. 1437–1440.

3. Soukup, D.; Huber-Mörk, R. Convolutional neural networks for steel surface defect detection from photometric stereo images. In Proceedings of the International Symposium on Visual Computing, Berlin/Heidelberg, Germany, 8–10 December 2014; pp. 668–677.

4. Molina, J.; Solanes, J.E.; Arnal, L.; Tornero, J. On the detection of defects on specular car body surfaces. *Robot. Comput. Integr. Manuf.* **2017**, *48*, 263–278. [CrossRef]

5. Lin, H.; Li, B.; Wang, X.; Shu, Y.; Niu, S. Automated defect inspection of LED chip using deep convolutional neural network. *J. Intell. Manuf.* **2018**, *30*, 2525–2534. [CrossRef]

6. Zhang, J.; Guo, Z.; Jiao, T.; Wang, M. Defect Detection of Aluminum Alloy Wheels in Radiography Images Using Adaptive Threshold and Morphological Reconstruction. *Appl. Sci.* **2018**, *8*, 2365. [CrossRef]

7. Yuan, X.-C.; Wu, L.-S.; Peng, Q. An improved Otsu method using the weighted object variance for defect detection. *Appl. Surf. Sci.* **2015**, *349*, 472–484. [CrossRef]

8. Gao, J.; Wang, Z.; Liu, Y.; Jian, C.; Chen, X. Development of OLED panel defect detection system through improved Otsu algorithm. In Proceedings of the International Conference on Intelligent Robotics and Applications, Montreal, QC, Canada, 3–5 October 2012; pp. 127–134.

9. Spinola, C.G.; Canero, J.; Moreno-Aranda, G.; Bonelo, J.M.; Martin-Vazquez, M. Real-time image processing for edge inspection and defect detection in stainless steel production lines. In Proceedings of the 2011 IEEE International Conference on Imaging Systems and Techniques, Penang, Malaysia, 17–18 May 2011; pp. 170–175.

10. Chen, J.; Jain, A.K. A structural approach to identify defects in textured images. In Proceedings of the 1988 IEEE International Conference on Systems, Man, and Cybernetics, Beijing, China, 8–12 August 1988; pp. 29–32.

11. Datta, A.; Kumar, J. Detection of Defects in Fabric by Morphological Image Processing. *Woven Fabr. Eng.* **2010**. [CrossRef]

12. Bai, X.; Fang, Y.; Lin, W.; Wang, L.; Ju, B.-F. Saliency-based defect detection in industrial images by using phase spectrum. *IEEE Trans. Ind. Inform.* **2014**, *10*, 2135–2145. [CrossRef]

13. Borwankar, R.; Ludwig, R. An Optical Surface Inspection and Automatic Classification Technique Using the Rotated Wavelet Transform. *IEEE Trans. Instrum. Meas.* **2018**, *67*, 690–697. [CrossRef]

14. Hu, G.-H. Automated defect detection in textured surfaces using optimal elliptical Gabor filters. *Optik* **2015**, *126*, 1331–1340. [CrossRef]

15. Shumin, D.; Zhoufeng, L.; Chunlei, L. Adaboost learning for fabric defect detection based on hog and SVM. In Proceedings of the International Conferenceon Multimedia Technology, Hangzhou, China, 26–28 July 2011.

16. Kuang, H.; Ding, Y.; Li, R.; Liu, X. Defect detection of bamboo strips based on LBP and GLCM features by using SVM classifier. In Proceedings of the 2018 Chinese Control and Decision Conference (CCDC), Shenyang, China, 9–11 June 2018; pp. 3341–3345.

17. Chang, C.-F.; Wu, J.-L.; Chen, K.-J.; Hsu, M.-C. A hybrid defect detection method for compact camera lens. *Adv. Mech. Eng.* **2017**, *9*, 1–19. [CrossRef]

18. Zhou, Q.; Chen, R.; Huang, B.; Liu, C.; Yu, J.; Yu, X. An Automatic Surface Defect Inspection System for Automobiles Using Machine Vision Methods. *Sensors* **2019**, *19*, 644. [CrossRef]

19. Islam, S.S.; Dey, E.K.; Tawhid, M.N.A.; Hossain, B.M.M. A CNN Based Approach for Garments Texture Design Classification. *Adv. Technol. Innov.* **2017**, *2*, 119–125.

20. Suh, W.-H.; Park, J.-W.; Kim, J.-R. Traffic Safety Evaluation Based on Vision and Signal Timing Data. *Proc. Eng. Technol. Innov.* **2017**, *7*, 37–40.

21. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; Springer: Cham, Switzerland, 2016; pp. 21–37.

22. Arikan, S.; Varanasi, K.; Stricker, D. Surface Defect Classification in Real-Time Using Convolutional Neural Networks. *arXiv* **2019**, arXiv:1904.04671.

23. Wang, T.; Chen, Y.; Qiao, M.; Snoussi, H. A fast and robust convolutional neural network-based defect detection model in product quality control. *Int. J. Adv. Manuf. Technol.* **2017**, *94*, 3465–3471. [CrossRef]

24. Mei, S.; Wang, Y.; Wen, G. Automatic Fabric Defect Detection with a Multi-Scale Convolutional Denoising Autoencoder Network Model. *Sensors* **2018**, *18*, 1064. [CrossRef]

25. Li, J.; Su, Z.; Geng, J.; Yin, Y. Real-time detection of steel strip surface defects based on improved yolo detection network. *IFAC PapersOnLine* **2018**, *51*, 76–81. [CrossRef]

26. Cha, Y.-J.; Choi, W.; Suh, G.; Mahmoudkhani, S.; Büyüköztürk, O. Autonomous Structural Visual Inspection Using Region-Based Deep Learning for Detecting Multiple Damage Types. *Comput. Aided Civ. Infrastruct. Eng.* **2018**, *33*, 731–747. [CrossRef]

27. Yuan, Z.-C.; Zhang, Z.-T.; Su, H.; Zhang, L.; Shen, F.; Zhang, F. Vision-based defect detection for mobile phone cover glass using deep neural networks. *Int. J. Precis. Eng. Manuf.* **2018**, *19*, 801–810. [CrossRef]

28. Ferguson, M.K.; Ronay, A.; Lee, Y.T.; Law, K.H. Detection and Segmentation of Manufacturing Defects with Convolutional Neural Networks and Transfer Learning. *Smart Sustain. Manuf. Syst.* **2018**, *2*. [CrossRef]

29. Wen, S.; Chen, Z.; Li, C. Vision-Based Surface Inspection System for Bearing Rollers Using Convolutional Neural Networks. *Appl. Sci.* **2018**, *8*, 2565. [CrossRef]

30. Chen, J.; Liu, Z.; Wang, H.; Nunez, A.; Han, Z. Automatic Defect Detection of Fasteners on the Catenary Support Device Using Deep Convolutional Neural Network. *IEEE Trans. Instrum. Meas.* **2018**, *67*, 257–269. [CrossRef]

31. Li, Y.; Huang, H.; Xie, Q.; Yao, L.; Chen, Q. Research on a Surface Defect Detection Algorithm Based on MobileNet-SSD. *Appl. Sci.* **2018**, *8*, 1678. [CrossRef]

32. Song, L.; Li, X.; Yang, Y.; Zhu, X.; Guo, Q.; Yang, H. Detection of Micro-Defects on Metal Screw Surfaces Based on Deep Convolutional Neural Networks. *Sensors* **2018**, *18*, 3709. [CrossRef]

33. Tao, X.; Zhang, D.; Ma, W.; Liu, X.; Xu, D. Automatic Metallic Surface Defect Detection and Recognition with Convolutional Neural Networks. *Appl. Sci.* **2018**, *8*, 1575. [CrossRef]

34. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]

35. Machine Vision Lenses & Security Surveillance Optics | vs. Technology. Available online: https://vst.co.jp/en/ (accessed on 1 May 2019).

36. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.

37. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

38. Niblack, W. *An Introduction to Image Processing*; Prentice-Hall: Englewood, NJ, USA, 1986; pp. 115–116.

39. Sauvola, J.; Pietikäinen, M. Adaptive document image binarization. *Pattern Recognit.* **2000**, *33*, 225–236. [CrossRef]