Appearance Defect Detection and Localisation using a Lightweight CNN-based Detector

Rémi Cogranne, Lucien Derouet

Appearance Defect Detection and Localisation using a Lightweight CNN-based Detector

Rémi Cogranne*
Troyes University of Technology
Troyes, France
remi.cogranne@utt.fr

Lucien Derouet†
Troyes University of Technology
Troyes, France
lucien.derouet.2023@utt.fr

ABSTRACT

Automatic visual inspection plays a crucial role in many industrial sectors to assist human operators. This paper studies the general problem of automatic detection of appearance defects with application on wheels surface quality control. An original method is proposed combining image processing and deep learning. This method exploits geometrical knowledge of the manufactured product, which allows splitting the image into homogeneous zones, over which a dedicated lightweight deep learning network is trained to detect and locate anomalies with the highest accuracy. Additionally, the present paper also addresses the issue of training a supervised AI architecture with a limited availability and imbalance dataset containing 100, 000 images but only 1, 000 with defects. We show on this dataset that the proposed lightweight CNN can achieve a high detection rate for low false-positive rates, which is the main goal for applications in an operational context.

CCS CONCEPTS

• **Computing methodologies** → Simulation evaluation; **Supervised learning**; **Artificial intelligence**; *Computer vision problems*; **Visual inspection**; • **Information systems** → *Industrial Process control systems*; • **General and reference** → *Design and Evaluation*; • **Mathematics of computing** → *Nonparametric statistics.*

KEYWORDS

Defect detection, Visual inspection, Lightweight CCN, Image processing, Deep learning.

ACM Reference Format:
Rémi Cogranne and Lucien Derouet. 2025. Appearance Defect Detection and Localisation using a Lightweight CNN-based Detector. In *Proceedings of ACM International Conference on Computing and Artificial Intelligence (ACM ICCAI '25).* ACM, New York, NY, USA, 9 pages. https://doi.org/XXXXXXX.XXXXXXX

*Corresponding author.
†Lucien Derouet was with the UTT when the work presented in this paper was carried out.

ACM ICCAI '25, March 28–31, 2025, Kyoto, Japan

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

The increasing competition in various industrial sectors has led to a significant rise in production costs. Globalization has made it challenging for manufacturers in more economically developed countries (MEDCs) to remain competitive in terms of price. In response, many industries have focused on improving the quality of their products to differentiate themselves. While quality is often associated with performance, it also encompasses aesthetic aspects, which have a well-established impact on consumer purchasing decisions [19].

The manufacturing of wheels, the subject of this study, has not been immune to this trend. Recent research has shown that aesthetic aspects are even more important than environmental factors in car sales [11]. To meet this demand, quality control processes have increasingly taken into account the visual and aesthetic aspects of products. However, appearance quality control is complex due to the lack of precise standards or measurable quantities, and it often relies on human operators. This task is challenging, as a single operator must inspect many products for an extended period under strict time constraints, leading to non-reproducible, subjective, biased, and sometimes superficial or flawed visual inspections.

In an effort to assist human-based control, automatic visual inspection systems have emerged [5, 14, 15]. Such systems have been studied in various industrial sectors for appearance defect detection [1] in a wide range of materials and products, including fabrics [12, 16], nuclear fuel rods [8], printed circuit boards [17], and food [4].

1.1 State of the Art

Automatic visual inspection methods can be broadly categorized into three main types [5, 14–16]. The first type relies on a reference non-anomalous object and measures the divergence of each inspected object from this reference. When this divergence exceeds a given threshold, the inspected object is declared abnormal. This type of method [21] is highly dependent on experimental conditions and requires that all inspected objects be extremely similar.

The second type of method is based on prior statistical information about the non-anomalous object. Depending on whether it is aimed at detecting a specific type of anomaly [6, 7, 17, 22] or any deviation from the null hypothesis [26, 27], a dedicated optimal statistical test can be designed to decide whether observations are more likely drawn from the null or alternative statistical hypothesis. The main difficulty with this approach is modelling the observations with great accuracy, ensuring that all pixels from all images of non-anomalous wheels can be accurately described with a single model.

The third type of method relies on image processing tools. Initially based on simple techniques, such as morphological operations [2], recent advances in image representation have been exploited, including multi-resolution models [24] and sparse dictionary learning [18]. Originally, the goal of these tools was to increase the visibility of defects, but it has evolved to provide a homogeneous feature representation of inspected objects for use in supervised machine learning methods [10, 20]. Recently, the development of deep learning has been leveraged for visual inspection and non-destructive testing (NDT) [23, 30]. While it allows merging image processing and classification steps, its application is not straightforward in practical operational contexts, as we will study in this paper.

For a more detailed review of methods for automatic defect detection, the reader is referred to [5, 14–16].

1.2 Paper Contribution and Organization

This paper belongs to the third category, presenting an original and practical method for automatic visual inspection of wheels based on state-of-the-art deep learning methods. This allows benefiting from powerful image analysis and processing methods, achieving the highest detection accuracy. One of the originalities of this type of approach is that, based on the geometry of the inspected objects and knowledge of their design, it proposes to help the learning task by providing subsets of processed images that share similar characteristics in terms of contrast and content. Additionally, this application-oriented paper addresses the problem of designing a supervised detection task without much abnormal data while preventing overfitting to specific defects that can be found in the training dataset. The proposed method carefully takes into account the limited size of the dataset as well as processing time and real-time constraints to design a lightweight architecture that meets operational constraints and makes the interpretability of the detection easier by localizing the area of potential defects.

The main contributions of this paper are summarized below:

- The paper leverages the use of deep learning approaches for appearance defect detection, allowing for automatic adaptation to a wide range of defects in terms of shape, size, and location, and natural adjustment to varying acquisition conditions.
- The paper takes advantage of knowledge on inspected products, integrating them through a preliminary step of specific image analysis and processing operations to make non-anomalous objects as uniform as possible, easing the ensuing detection.
- The paper addresses the problem of imbalance in the training dataset due to the limited number of defective products. To this end we proposed an application specific simple yet efficient method for generating artificial defects based on the real ones and expertise of human controllers, while carefully preventing overfitting to specific defect generation.
- The paper relies on a lightweight architecture to meet real-time operational constraints and makes the interpretability of the detection easier by localizing the area of potential defects.
- The efficiency of the method is evaluated on very large datasets of generated as well as real appearance defects, confirming the relevance and sharpness of the proposed simple deep learning architecture.

We would like to acknowledge that this paper is clearly distinguished from our previous contributions [28, 29], which focused on online or sequential detection of coating intensity with maximal detection delay, and [25], which presented the whole automatic visual inspection system. While the present paper studies addressed the same problem as our prior works [26, 27], it relies on a fundamental different methodology: the prior works used a statistical linear parametric model to design a test based on hypothesis testing theory for appearance defect detection with false alarm constraints. This paper uses a deep learning methodology and aims to be more general, as it does not require a precise model of inspected objects while incorporating knowledge on their geometry.

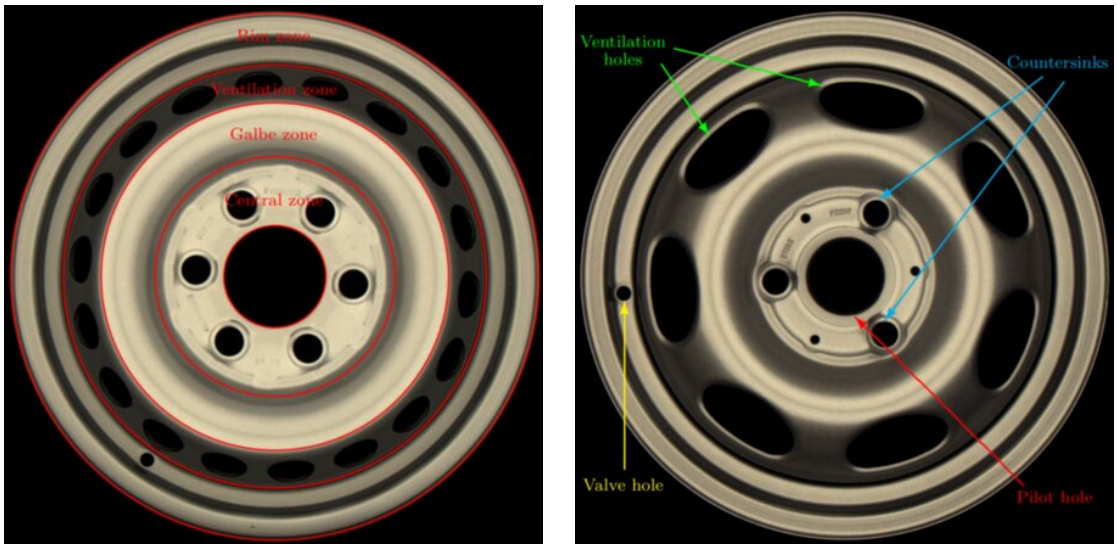
The rest of the paper is organized as follows. Section 2 briefly presents the operational context and clearly states the detection problem along with its practical constraints. Section 3 presents the inspected product and the preliminary steps of image processing used to help feed the deep network with homogeneous images. This section also presents the method for artificial generation of defects. Section 4 details the proposed architecture for detection of appearance defect detection and localization, focusing on the techniques used to prevent overfitting to generated defects. Numerical results are presented in Section 5, evaluating the ability of the proposed approach over real defects of very small "intensity" to show its relevance, sharpness, and generalization capacity with respect to defects size and shape. Finally, Section 7 concludes the paper by summarizing the main contributions and discussing closing remarks.

2 PROBLEM STATEMENT AND POSITION OF THE PAPER

The primary objective of this paper is to develop a high-accuracy detection method for appearance defects on wheel surfaces, with a focus on minimizing false alarm rates. This is crucial, as manual inspection is required for all defective objects, which can lead to production delays.

While wheels share a similar structure and geometry, each type exhibits unique characteristics. Figure 1 illustrates the various components of a wheel, and Figure 2 presents a range of wheel images with varying acquisition conditions. Note that the marking spot (surrounded in blue) is considered a defect on one wheel but not on another (painted in red). The four rightmost images demonstrate the variability in wheel models, including size, brightness, and localization of elements.

The high variability in wheel production and the small size of potential defects pose significant challenges for detection. The image resolution must be sufficiently large to capture minimal defects, and the appearance of defects can vary greatly, ranging from small paint drops to large marks. Given these constraints, designing an accurate detection method with a low false-positive rate is a challenging task.



(a) The four differents area of the wheel: Central, galbe, ventilation and rim zones.

(b) Specific elements that makes wheel surface more complex and variable depending on wheels model: ventillation holes are often present for improving cooling of the breaks, countersinks are use to attach the wheel to the vehicle, the pilot hole is used for positioning of the wheels and the valve hole is where the valve is place in order to inflate the tire.

Figure 1: Representation and description of the different elements of a wheel. Left: The four main zones from a wheel ; Right: the main geometrical elements that can be found in almost all wheels.

To address these challenges, we propose leveraging state-of-the-art deep learning-based classifiers. However, it is essential to balance the complexity of the architecture with operational constraints, such as limited training data and the need for fast inspection.

The large image size (several megapixels) and diversity of non-anomalous wheels in terms of content, contrast, and elements further complicate the detection task. To mitigate these difficulties, we propose a novel method that combines a lightweight deep learning-based detector with a carefully designed preprocessing step. This step involves analysing the object geometry to split the image into four homogeneous areas which share similar characteristics across all wheel types. This approach enables (1) high accuracy with a lightweight CNN, (2) adaptation to the variability of inspected products, and (3) identification of the area most likely to contain a defect.

Given the limited dataset of defective objects, we also propose a simple yet efficient method for training the CNN to detect a wide range of defects while preventing overfitting. This approach will be discussed in detail in the subsequent sections.

3 WHEEL IMAGE ANALYSIS AND PROCESSING

3.1 Data Preparation

To facilitate the deep learning classification task, we leverage our understanding of the wheel manufacturing process to pre-process the images. Specifically, we divide the image into four homogeneous regions: the central zone, the galbe zone, the ventilation holes area, and the rim, as illustrated in Figure 1.

For details on the proposed approach, we refer the reader to Tout et al. [25]. The pre-processing step is essential due to the constraints of the conveyor belt imaging system. Since the conveyor belt cannot be stopped for individual wheel imaging, and the size and geometry of the wheels vary, products are often located at slightly different positions within the original image.

To address this issue, we employ the Circular Hough Transform (CHT) to locate the pilot hole, which serves as a reference point for the wheel's centre. The Hough transform is a well-established technique in computer vision for detecting objects with known shapes [3, 9, 13]. Originally proposed for detecting straight lines and circles, it has been generalized to detect shapes of almost any type. In this case, since the pilot hole is a circle with a bounded radius, we can efficiently search for this specific shape within the central part of the image.

The location of the pilot hole is critical, as it enables us to (1) split the wheel into its main areas and (2) identify other key elements. Once the pilot hole's centre is located, we can use our knowledge of wheel geometry to divide the image into the four main areas.

However, to ensure that the image subparts are properly aligned, we also need to detect key elements, such as valve holes and countersinks. From the pilot hole location, we can start searching for countersinks, which have a bounded distance from the wheel centre and a bounded radius. We employ the CHT again, this time over the corresponding subpart of the image. The detection of countersinks facilitates the identification of valve holes, which are either aligned with one of the countersinks or located exactly between two countersinks, as shown in Figure 2. Using this information, we

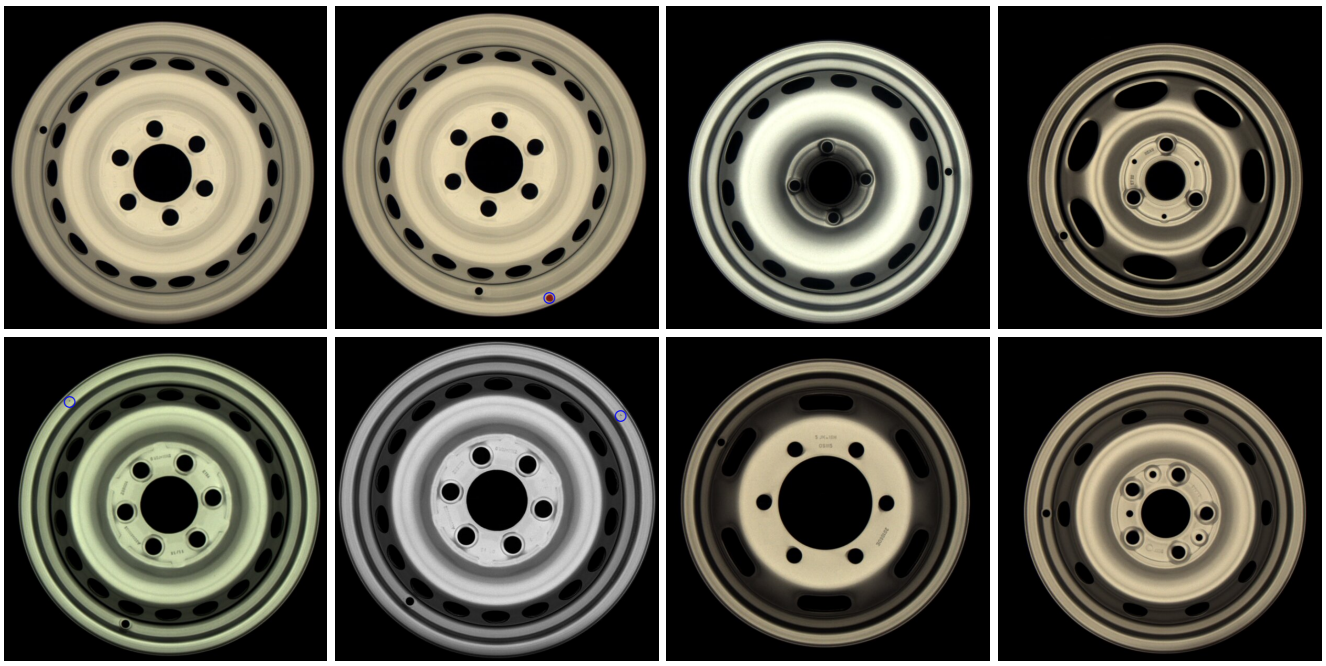


Figure 2: A few examples of wheels' images as captured by the automatic inspection system highlighting the wide diversity of wheels ; note especially the change in terms of size, shape, colours, and location of wheel elements. Also note the presence of some geometrical marks circled in blue, which are not defects: these are used to show the imbalance direction of the rim weight in order to align the centre of gravity by placement of the rim and disc parts.

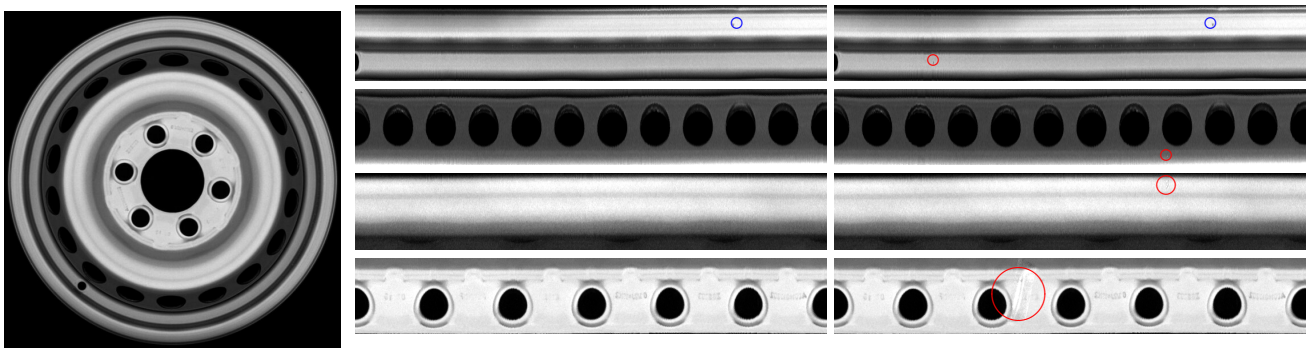


Figure 3: Examples of a wheel's image as captured by the automatic inspection system prior to being processed (left), after being processed and split into different areas (centre) and after the addition of "artificial" defects.

can search for a circle in a predefined zone that corresponds to the bounds of the valve hole distance from the wheel centre.

The final processing steps involve unfolding and resizing all the image subparts to the same size. These operations are performed in a similar manner. For resizing, we select the smallest image of each size based on a small dataset of image samples. Both unfolding and resizing are carried out without interpolation, as interpolation requires a prior low-pass filtering operation that may remove a significant portion of the appearance defects we aim to detect. Moreover, applying two low-pass filters in succession (first unfolding and then resizing) would smooth out the defects' edges, making their detection more challenging.

3.2 Appearance Detect Generation

With the pre-processing steps complete, we can now proceed to generate a dataset of representative examples for training and evaluating a supervised machine learning algorithm. This dataset must be exhaustive and representative of all possible real-life cases, including various types of wheels and defects. It shall be noted that constructing such a dataset is a challenging task, particularly in a real operational context. On the one hand, the dataset must be as comprehensive as possible, containing a large quantity of images representing all possible types of wheels and defects. On the other

hand, the dataset must be representative of the production process and customer demands, which can vary significantly.

The manual labelling of images is also a time-consuming and expensive process. Furthermore, the major difficulty in constructing a training dataset lies in the fact that the manufacturing process is well controlled, resulting in fewer than 1

To overcome this challenge, we propose generating artificial defects based on truly observed ones. To this end, we collected slightly more than 1,000 defective wheels and analysed them to extract information on the type of defect and its occurrence frequency. We also consulted with operators who provided valuable insights into their knowledge, particularly in terms of defect recognition and location. From this dataset, we manually extracted a subset of the most representative defects from the whole image of wheels.

Once the individual defect images were extracted, we flattened their background by selecting the boundary of the image not affected by the defect. We then fit a polynomial model to this boundary, which can be represented as follows:

$$f(x,y) = \sum_{i=0}^{d_x} \sum_{j=0}^{d_y} c_{i,j} x^i y^j, \tag{1}$$

where x and y are the local pixels' coordinates and $c_{i,j}$ represents the polynomial coefficients. We limit this analysis to a degree $d_x = d_y = 1$ for the smallest defect and used up to a degree of $d_x = d_y = 2$. The coefficient has been estimated using the ordinary least square estimation:

$$\mathbf{c} = \left(\mathbf{M}^\top \mathbf{M}\right)^{-1} \mathbf{M}^\top \mathbf{I}, \tag{2}$$

where $\mathbf{I}$ represents the defect-free pixels used for background estimation, put into a single column vector, and $\mathbf{M}$ represents the polynomial contributions corresponding to selected pixel locations put into a column-wise matrix.
The goal of background flattening operation is to allow the superimposition of those extracted defects with as few artefacts as possible.

A few examples of the defects extracted from the image of wheels with defects are presented in Figure 4. Note that the great variability in terms of shape, size and contrast.

It is also important to avoid always superimposing the same defect to prevent recognition of a specific pattern. Therefore the dataset of products with artificially generated defects has been created by modifying in a randomized manner the defect prior to their superimposition. More precisely, for each image of each subpart of wheels, we select first, in a random manner, the type of superimposed defect. Let us denote $d(x,y)$ the pixel value of the corresponding defect's image where the relative coordinates (x,y) are normalized in the range $[-1,1]$ for the image of the reference defects. The image is transformed to get the defect image $d'(x,y)$ using a rotation, a flip, a homothetic transform and a scaling:

$$d'((-1)^{f_x}x, (-1)^{f_y}y) = G \times \\ d\Big(E_x \cdot (x\cos(\theta) - y\sin(\theta)), E_y \cdot (x\sin(\theta) + y\cos(\theta))\Big), \tag{3}$$

where f_x and f_y are horizontal and vertical flipping factors, respectively, drawn from a Bernoulli distribution with $p = 1/2$. The scaling factor G is drawn from a uniform distribution between 0.25 and 2. The horizontal and vertical "enlargement" factors, respectively E_x and E_y, are also drawn from a uniform distribution between 0.25 and 2. The rotation angle θ is drawn from a uniform distribution over the set $[0, 2\pi]$.
All these parameters are selected randomly and independently for each new defect. This approach ensures that the generated defects have varying sizes, shapes, and orientations.
In addition to the randomized transformations, we also select a set of reference defects with different probabilities. The probability of each defect is based on our observations, experience from operators, and the difficulty of detection. We believe that including more defects that are hard to detect in the training dataset is essential for improving the robustness of the detection algorithm.

Figure 3 illustrates an example of the "data preparation pipeline". Starting from the left, the raw image of the wheel is obtained from the imaging system. The centre column shows the results of extracting the four subparts after flattening and resizing without filtering (artefacts are slightly visible). The right column shows the same subparts with artificial defects superimposed, surrounded by red circles. The reader can notice that the defects are small and hardly visible, reflecting the type of real appearance defects that can be observed in practice.
While this process allows us to create defects with different sizes and shapes, we acknowledge that it is hardly possible to represent exhaustively all possible appearance defects.

4 PROPOSED LIGHTWEIGHT DEEP LEARNING ARCHITECTURE FOR DETECTION AND LOCALIZATION

In recent years, deep learning has revolutionized various fields, including image processing, chess and go games, chatbots, and automatic visual inspection systems. The field of artificial intelligence is evolving at an ever-increasing pace, making it challenging to summarize the main breakthroughs achieved over the past decade. In this paper, we focus on image classification, which is the original topic that sparked the rise of deep learning. The ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) is an annual competition where scientists compete to classify objects and scenes from 1,000 classes. This competition remains a common benchmark for image classification, which is closely related to the problem of anomaly detection addressed in this paper.
The first competitions saw a rapid increase in the number of layers, leading to the term "deep" learning. For example, the VGG network used 16 and 19 layers in 2012-2013, while the ResNet network could implement 50, 100, or 150 layers in 2015. This seems to confirm the intuition that depth plays a crucial role in deep learning. The ResNet also addressed the problem of adding layers to an efficient network, which can sometimes degrade performance, while additional layers should, at worst, learn the identity function. To preserve performance, the ResNet introduced residual connections, which correspond to adding the identity in each new layer.

The question of the relevant size of the convolution kernel was studied in an interesting manner by the Inception Networks, which proposed different convolution kernel sizes in the layers followed by an aggregation step. More recently, as the size of CNNs kept increasing, the question of designing smaller yet efficient architectures was raised. Inspired by the lightweight MobileNet, the

Figure 4: Examples of defects extracted from real defective wheels and used to generate artificial superimposed defects.

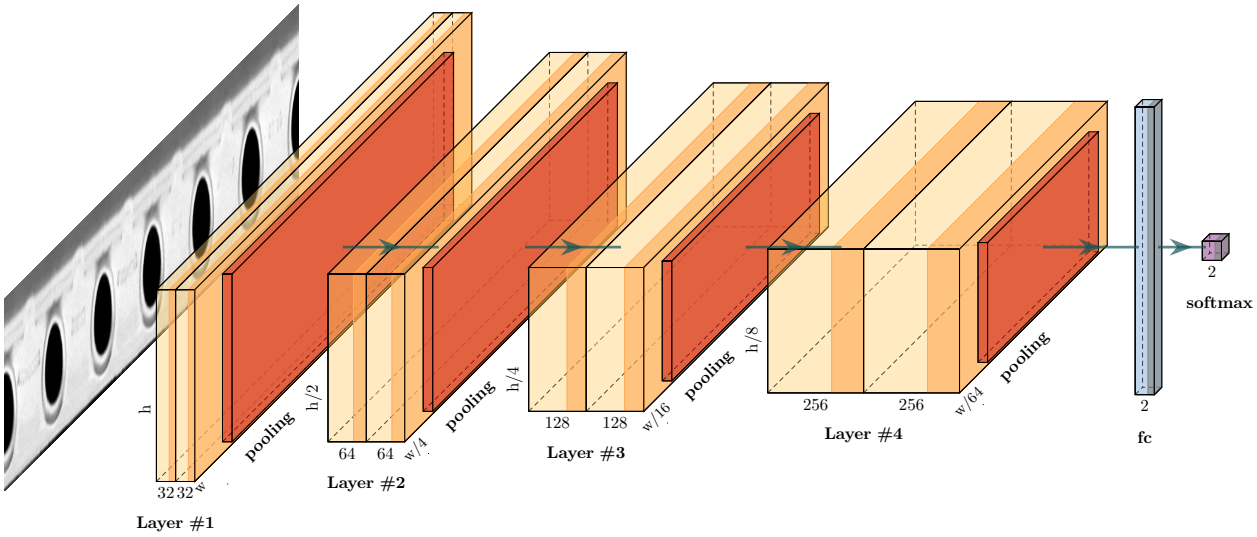


Figure 5: Illustration of the proposed lightweight CNN architecture proposed for appearance defect detection.

EfficientNet introduced the compound scaling strategy, which is a tradeoff between depth, width, and resolution scaling. This strategy allows for the design of architectures with an order of magnitude fewer parameters while preserving classification accuracy.

In this paper, we propose an ad hoc lightweight architecture that is inspired by these prior works and fits our operational application context. The proposed architecture is described in Figure 5: one can note the two consecutive convolution operations in each layer (represented in yellow). Also note the pooling operation, which is used to reduce the relative width of the image, whose height is much smaller.

Inspired by the EfficientNet method, we sought to find a tradeoff between depth, width, and resolution scaling, resulting in a rather shallow architecture with four layers. Each layer consists of the same elements: (1) convolution, (2) batch normalization, (3) max-pooling with a vertical stride of 2 and 4, respectively, and (4) ReLU activation. However, to minimize the number of parameters, we used the good old techniques from the VGG to include two consecutive convolution layers of size 3x3 instead of a kernel of size 5x5, with the same "receptive field" but fewer parameters and more non-linearity and abstraction.

The final classification is achieved by feeding the flattened final output to a fully connected layer followed by a softmax.

4.1 Defect Localization

We will address the practical implementation aspects, such as setting hyperparameters and techniques to prevent overfitting, in Section 5. Before moving to application and numerical results, we would like to describe the last major requirement of our application-oriented approach: once a wheel has been classified as defective, it is desirable to provide the localization of potential defects to indicate to the operator where to look.

However, it is well known that interpretability of deep learning algorithms remains a challenging problem, particularly due to the large number of parameters in most architectures. We have tried several methods, including the well-celebrated Yolo, but found that it requires training another deep learning architecture specifically for this task. Instead, we obtained better results using the Class Activation Map (CAM) method proposed in [31], which essentially consists in modifying the last classification layers. Therefore, in this paper, we have used a modified version of CAM, referred to as Grad-CAM, which allows extracting a heatmap based on the gradient of the prediction with respect to a given layer.

This approach has been extensively used in various applications, such as interpreting ocular medical checkups or rolling bearing inspections. Note that this method works very efficiently in our case, in large part due to the proposed lightweight CNN architecture. Indeed, with as many as 152 layers as one can find, finding relevant layers in which the gradient with respect to the final decision is interpretable is hardly possible. This was an important motivation in the design of a lightweight CNN architecture.

In this paper, we adapted this methodology as follows: first, the proposed lightweight CNN is trained, see the next Section 5 for details. We used the gradients of the prediction of the first

convolution for the penultimate (third) layer. To obtain a single signal map from 128 different channels, we used a weighted average: the average activation is calculated over every channel, which serves as the weights for computing a single heatmap from all channels.
For localization, we threshold the obtained activation heatmap using 70% of the max value, which was empirically selected to obtain good location accuracy. Simple morphological operations were used to remove outliers. More precisely, we used dilation and erosion and dilation (or opening and dilation) with a rectangle of size 3x3, and finally extracted the area with maximal non-zeros values. Last, for display purposes, we resized the image to the original input size using the simplest bilinear interpolation.
Though this method is quite simple, it allows obtaining very good results while requiring minimal additional computation.

5 NUMERICAL RESULTS AND VALIDATION

For benchmarking purposes, we used the ResNet50-v2, Inceptionv4, and EfficientNet-b0 models, which have 23M, 43M, and 5.3M trainable parameters, respectively. All models were trained in the same manner on an NVIDIA RTX 3090 GPU with a maximal batch size. We used the AdamW optimizer with a starting learning rate of 10^{-4} and a scheduler "Reduce on Plateau" with a reduction factor of 0.5 and patience parameter of 1. The number of epochs was set to 20, which was sufficient for convergence in all cases.
We employed a curriculum or transfer learning approach, where each model was first trained to recognize the different types of wheels. After a few epochs (typically 4-6), the model achieved excellent results, and the trained weights were used as input for the task of defect detection. We observed empirically that this approach generally provided slightly better results than starting from weights pre-trained on ImageNet. We believe that this may be due to the specific nature of the images used.

We trained a specific classifier for each part of the wheel using slightly more than 100,000 defect-free images (at least those that passed quality control by careful visual inspection) and the same number of images with an artificial defect added following the procedure described in Section 3.2. The images were divided into 70% for training, 20% for validation, and 10% for testing. To prevent overfitting, we used several methods, including data augmentation (horizontal and vertical flipping) and a rather harsh dropout, with a rate of 50% for Inception and ResNet and 30% for EfficientNet and the proposed lightweight CNN. Additionally, we added noise to the data after every layer, which helped but not significantly.
We tested the classifier on raw images, as acquired by the imaging device, as well as images of residuals, which were obtained by removing the estimated "denoised" content. To this end, we used the "content rejection" method developed in our prior works [26, 27], which was carefully designed for specific images of wheel parts and aimed at preserving most of the defects.

The results in terms of accuracy of defect detection are presented in Table 1 for all cases. It can be seen that while our model is extremely light compared to its competitors, it achieves overall similar performances. In general, EfficientNet performs slightly better (on both original images and residuals). Additionally, one can note that detection over the central zone seems challenging for all architectures, while the ventilation holes zone is not despite the many details it contains.

Table 1: Accuracy of the different architectures over artificial defects, real defects, original images and residuals ; the "Lightweight Model" corresponds to the architecture proposed in the present paper.

			Central	Galbe	Ventililation	Rim
Original images	Artificial defects	ResNet	0.9622	0.9835	0.9927	0.9787
		Inception	0.9655	0.9891	0.9935	0.9844
		EfficientNet	0.9719	0.9896	0.9951	0.9860
		Lightweight Model	0.9387	0.9791	0.9892	0.9705
Original images	Real defects	ResNet	0.7425	0.8884	0.7500	0.7690
		Inception	0.8023	0.9342	0.7969	0.8845
		EfficientNet	0.7659	0.8510	0.8103	0.8563
		Lightweight Model	0.8083	0.9227	0.8085	0.9186
Residuals images	Artificial defects	ResNet	0.9203	0.9877	0.9781	0.9754
		Inception	0.9403	0.9883	0.9829	0.9757
		EfficientNet	0.9585	0.9906	0.9879	0.9805
		Lightweight Model	0.9150	0.9876	0.9875	0.9726
Residuals images	Real defects	ResNet	0.8683	0.9055	0.9844	0.9431
		Inception	0.9042	0.9427	0.9844	0.8884
		EfficientNet	0.8862	0.9281	0.9401	0.9221
		Lightweight Model	0.9281	0.9675	0.9883	0.9636

Interestingly, it seems that using residual images instead of the original ones does not seem to help significantly. However, it is striking that when we evaluate all architectures on real defects, the use of residuals brings a significant robustness. Note that in all cases, the evaluation on real defects was carried out without any additional training (the architecture was trained on artificial defects and used "as is" on real defects).
Last but not least, we observed that our proposed architecture seems to achieve a much higher robustness, as it performs significantly better on real defects. We believe that this may be due to its very lightweight features that preserve better generalization capabilities.

6 DEFECTS LOCALIZATION

Before concluding the present paper, we will present the results on defects localization. Figure 6 shows an example of localization output using the proposed method based on Grad-CAM. It can be seen from this example that providing ground truth for the localization of defects is far from being straightforward; hence the limited number of images used to assess the accuracy of defect localization. For a more general result, we computed the Intersection over Union (IoU) between ground truth and localization results over 2,000 images labelled manually. The IoU represents the ratio of area on

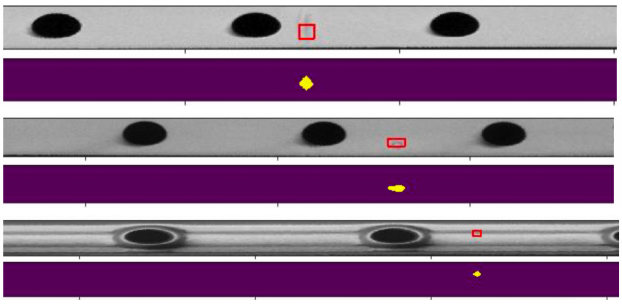


Figure 6: Outcomes from localization results for a few central zones.

which the localization results matches the reference over the total area.

More formally, for two areas $\mathcal{A}pred$ and $\mathcal{A}$ref, the IoU is defined as:

$$IoU(\mathcal{A}_{pred}, \mathcal{A}_{\text{ref}}) = \frac{\mathcal{A}_{pred} \cap \mathcal{A}_{\text{ref}}}{\mathcal{A}_{pred} \cup \mathcal{A}_{\text{ref}}}, \tag{4}$$

with $\cap$ and $\cup$ resp. the intersection and union operators (AND and OR binary operators).

Over the 2,000 images used for assessment, we obtained the following IoU:

- 81,38% on the central zone ;
- 93,43% on the galbe zone ;
- 87.80% on the ventilation zone ;
- 96,19% on the rim.

While these results are excellent, especially given the original goals to indicate the operator where to look, once again, we note that the central zone is the most challenging.

7 CONCLUSION AND CLOSING REMARKS

In the present paper, we have addressed a practical problem of appearance defect detection and localization on wheel surfaces. To achieve the highest possible accuracy, we proposed the use of deep learning, a highly effective approach in image processing. However, instead of feeding images to a complex network, we introduced an original method that incorporates a significant portion of image processing, specifically designed based on the knowledge of the manufactured product.

This approach has two main benefits. On the one hand, it greatly enhances the artificial intelligence architecture, allowing it to better understand the visual features of the product. On the other hand, it enables the method to meet the operational needs of the visual inspection system, including reliability, limited dataset, real-time operation, and defect localization.

Our proposed method represents an interesting tradeoff between the use of efficient deep learning solutions and the incorporation of knowledge on the visual inspection system it is designed for. By carefully taking into account the imbalance in the dataset and preventing overfitting to a very limited number of defects, our method offers a robust and reliable solution for appearance defect detection and localization on wheel surfaces.

In conclusion, our work demonstrates the effectiveness of combining deep learning with domain-specific knowledge to tackle complex problems in visual inspection. The proposed method has the potential to be applied to other areas where visual inspection is critical, and we believe that it will contribute to the development of more efficient and reliable visual inspection systems in the future.

REFERENCES

[1] Habibullah Akbar, Nanna Suryana, and Fikri Akbar. 2013. Surface defect detection and classification based on statistical filter and decision tree. *International Journal of Computer Theory and Engineering* 5, 5 (2013), 774.

[2] Alaknanda, R.S. Anand, and Pradeep Kumar. 2006. Flaw detection in radiographic weld images using morphological approach. *NDT & E International* 39, 1 (2006), 29 – 33. https://doi.org/10.1016/j.ndteint.2005.05.005

[3] Dana H Ballard. 1987. Generalizing the Hough transform to detect arbitrary shapes. In *Readings in computer vision*. Elsevier, 714–725.

[4] Tadhg Brosnan and Da-Wen Sun. 2004. Improving quality inspection of food products by computer vision—-a review. *Journal of food engineering* 61, 1 (2004), 3–16.

[5] C. H. Chen. 2015. *Handbook of Pattern Recognition and Computer Vision* (5th ed.). World Scientific Publishing Co., Inc., River Edge, NJ, USA.

[6] Rémi Cogranne, Guillaume Doyen, Nisrine Ghadban, and Badis Hammi. 2017. Detecting botclouds at large scale: A decentralized and robust detection method for multi-tenant virtualized environments. *IEEE Transactions on Network and Service Management* 15, 1 (2017), 68–82.

[7] Rémi Cogranne and Jessica Fridrich. 2015. Modeling and Extending the Ensemble Classifier for Steganalysis of Digital Images Using Hypothesis Testing Theory. *Information Forensics and Security, IEEE Transactions on* 10, 12 (December 2015), 2627–2642. https://doi.org/10.1109/TIFS.2015.2470220

[8] Rémi Cogranne and Florent Retraint. 2014. Statistical detection of defects in radiographic images using an adaptive parametric model. *Signal Processing* 96, Part B (2014), 173 – 189. https://doi.org/10.1016/j.sigpro.2013.09.016

[9] Richard O Duda and Peter E Hart. 1972. Use of the Hough transformation to detect lines and curves in pictures. *Commun. ACM* 15, 1 (1972), 11–15.

[10] Ce Ge, Jing Wang, Jingyu Wang, Qi Qi, Haifeng Sun, and Jianxin Liao. 2020. Towards automatic visual inspection: A weakly supervised learning method for industrial applicable object detection. *Computers in Industry* 121 (2020), 103232.

[11] Rebecca J. Hafner, Ian Walker, and Bas Verplanken. 2017. Image, not environmentalism: A qualitative exploration of factors influencing vehicle purchasing decisions. *Transportation Research Part A: Policy and Practice* 97 (2017), 89 – 105. https://doi.org/10.1016/j.tra.2017.01.012

[12] Elham Hoseini, Farnoush Farhadi, and Farshad Tajeripour. 2013. Fabric defect detection using auto-correlation function. *Int. J. Comput. Theory Eng* 5, 1 (2013), 114–117.

[13] Paul V.C. Hough. 1962. Method and Means for Recognizing Complex Patterns. U.S. Patent 3.069.654.

[14] Szu-Hao Huang and Ying-Cheng Pan. 2015. Automated visual inspection in the semiconductor industry: A survey. *Computers in industry* 66 (2015), 1–10.

[15] Christian Koch, Kristina Doycheva, Varun Kasi, Burcu Akinci, and Paul Fieguth. 2015. A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. *Advanced Engineering Informatics* 29, 2 (2015), 196 – 210. https://doi.org/10.1016/j.aei.2015.01.008 Infrastructure Computer Vision.

[16] A. Kumar. 2008. Computer-Vision-Based Fabric Defect Detection: A Survey. *IEEE Transactions on Industrial Electronics* 55, 1 (Jan 2008), 348–363. https://doi.org/10.1109/TIE.1930.896476

[17] Dejian Li, Shaoli Li, and Weiqi Yuan. 2020. Flexible printed circuit fracture detection based on hypothesis testing strategy. *IEEE Access* 8 (2020), 24457–24470.

[18] H. Y. Liao and G. Sapiro. 2008. Sparse representations for limited data tomography. In *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. 1375–1378.

[19] Yan Liu, Krista J. Li, Haipeng (Allan) Chen, and Subramanian Balachander. 2017. The Effects of Products' Aesthetic Design on Demand and Marketing-Mix Effectiveness: The Role of Segment Prototypicality and Brand Consistency. *Journal of Marketing* 81, 1 (2017), 83–102. https://doi.org/10.1509/jm.15.0315 arXiv:https://doi.org/10.1509/jm.15.0315

[20] Carlos Mera, Mauricio Orozco-Alzate, John Branch, and Domingo Mery. 2016. Automatic visual inspection: An approach with multi-instance learning. *Computers in Industry* 83 (2016), 46–54.

[21] Domingo Mery and Olaya Medina. 2004. *Automated Visual Inspection of Glass Bottles Using Adapted Median Filtering*. Springer Berlin Heidelberg, Berlin, Heidelberg, 818–825.

[22] Tan Nguyen, Hoang-Long Mai, Rémi Cogranne, Guillaume Doyen, Wissam Mallouli, Luong Nguyen, Moustapha El Aoun, Edgardo Montes De Oca, and Olivier Festor. 2019. Reliable Detection of Interest Flooding Attack in Real Deployment of Named Data Networking. *IEEE Transactions on Information Forensics and*

Security 14, 9 (Sept. 2019), 2470–2489. https://doi.org/10.1109/TIFS.2019.2899247

[23] Richard J. Pyle, Rhodri L. T. Bevan, Robert R. Hughes, Rosen K. Rachev, Amine Ait Si Ali, and Paul D. Wilcox. 2021. Deep Learning for Ultrasonic Crack Characterization in NDE. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 68, 5 (2021), 1854–1865. https://doi.org/10.1109/TUFFC.2020.3045847

[24] R. N. Strickland and He Il Hahn. 1997. Wavelet transform methods for object detection and recovery. *IEEE Transactions on Image Processing* 6, 5 (May 1997), 724–735.

[25] K. Tout, , F. Retraint, and R. Cogranne. 2017. Automatic Vision System for Wheel Surface Inspection and Monitoring. In *ASNT Annual Conference*.

[26] K. Tout, R. Cogranne, and F. Retraint. 2016. Fully automatic detection of anomalies on wheels surface using an adaptive accurate model and hypothesis testing theory. In *24th European Signal Processing Conference (EUSIPCO)*. 508–512. https://doi.org/10.1109/EUSIPCO.2016.7760300

[27] Karim Tout, Rémi Cogranne, and Florent Retraint. 2018. Statistical decision methods in the presence of linear nuisance parameters and despite imaging system heteroscedastic noise: Application to wheel surface inspection. *Signal Processing* 144 (2018), 430 – 443. https://doi.org/10.1016/j.sigpro.2017.10.030

[28] K. Tout, F. Retraint, and R. Cogranne. 2017. Wheels coating process monitoring in the presence of nuisance parameters using sequential change-point detection method. In *2017 25th European Signal Processing Conference (EUSIPCO)*. 196–200. https://doi.org/10.23919/EUSIPCO.2017.8081196

[29] K. Tout, F. Retraint, and R. Cogranne. 2018. Non-Stationary Process Monitoring for Change-Point Detection With Known Accuracy: Application to Wheels Coating Inspection. *IEEE Access* 6 (2018), 6709–6721. https://doi.org/10.1109/ACCESS.2018.2792838

[30] Jiaxing Ye and Nobuyuki Toyama. 2021. Benchmarking Deep Learning Models for Automatic Ultrasonic Imaging Inspection. *IEEE Access* 9 (2021), 36986–36994. https://doi.org/10.1109/ACCESS.2021.3062860

[31] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. 2016. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2921–2929.

submitted 05 Novembre 2024