applied sciences

MDPI

Article

Industrial Application of AI-Based Assistive Magnetic Particle Inspection

Julien Baumeyer [1,2,*], Hermine Chatoux [2], Arnaud Pelletier [1] and Patrick Marquié [2]

[1] CMPhy, 26 Rue Paul Sabatier, 71530 Crissey, France; arnaud.pelletier@cmphy.fr
[2] Laboratoire ImViA Dijon, 64 Rue Sully, 21000 Dijon, France; hermine.chatoux@u-bourgogne.fr (H.C.); patrick.marquie@u-bourgogne.fr (P.M.)
* Correspondence: julien.baumeyer@u-bourgogne.fr

Abstract: Magnetic Particle Inspection (MPI) is one of the most used methods in Non-Destructive Testing (NDT), allowing precise and robust defect detection on industrial-grade manufactured parts. However, human controllers perform this task in full black environments under UV-A lighting only (with safety glasses) and use chemical products in a confined environment. Those constraints tends to lower control performance and increase stress and fatigue. As a solution, we propose an AI-based assistive machine (called "PARADES") inside the hazardous environment, remotely manipulated by a human operator, outside of the confined area, in cleaner and safer conditions. This paper focuses on the development of a complete industrial-grade AI machine, both in terms of hardware and software. The result is a standalone assistive AI-based vision system, plug-and-play and controller-friendly, which only needs the usual power supply 230 V plug that detects defects and measures defect length. In conclusion, the PARADES machines address for the first time the problem of occupational health in MPI with an industrial standalone machine which can work on several parts and be integrated into current production lines. Providing cleaner and healthier working conditions for operators will invariably lead to increased quality of detection. These results suggest that it would be beneficial to spread this kind of AI-based assistive technology in NDT, in particular MPI, but also in Fluorescent Penetrant Testing (FPT) or in visual inspection.

Keywords: NDT; magnetic particle inspection; deep learning; object detection; dataset; vision system; quality control; hardware; software; transfer learning; faster R-CNN

check for updates

Citation: Baumeyer, J.; Chatoux, H.; Pelletier, A.; Marquié, P. Industrial Application of AI-Based Assistive Magnetic Particle Inspection. *Appl. Sci.* **2024**, *14*, 1499. https://doi.org/10.3390/app14041499

Academic Editor: Andrea Prati

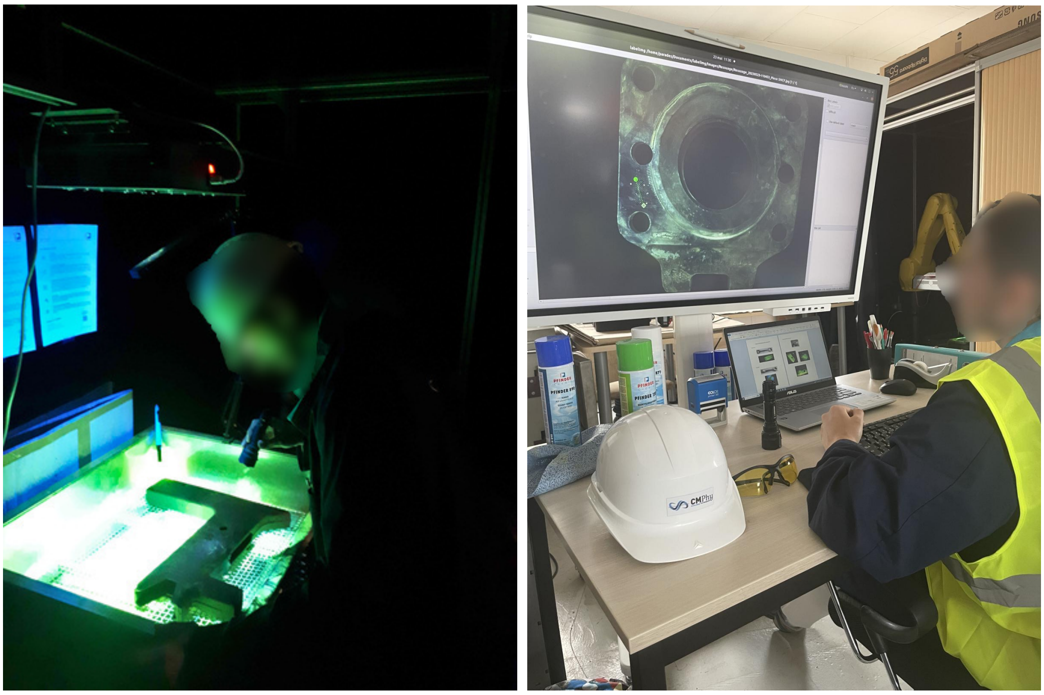
Received: 31 December 2023
Revised: 7 February 2024
Accepted: 8 February 2024
Published: 12 February 2024

1. Introduction

NDT is widely used in industries such as manufacturing, construction, aerospace, automotive and energy, among others. It helps ensure the quality [1] and safety of materials and structures by detecting flaws or defects early in the manufacturing or operational processes [2]. Some of the most used NDT methods are Magnetic Particle Inspection (MPI) and Fluorescent Penetrant Testing (FPT) because they are quick to implement, efficient and cost-effective. In this study, we will only explore MPI.

This method can suffer from lower control performance and increase stress and fatigue for human controllers due to the specific working condition, as depicted in Figure 1a. In order to overcome those drawbacks, a strong demand for a semi-automatic vision system has been expressed by our customers, leading to the design of the PARADES machine.

Customers' requirements concentrate on exceeding human controller performance in terms of percentage of defect detection, false detections and accuracy regarding defect length measurement. Parts' defect lengths and widths are summarized from several customer's requirements as well as the maximal system mechanical footprint. A summary of the important PARADES machine specifications is given in Table 1.



(**a**) (**b**)

Figure 1. MPI control without and with PARADES machine. (**a**) Human controller in a usual black, hazardous confined area. (**b**) Human controller remotely controlling defects with PARADES machine.

Table 1. PARADES machine important specifications.

Specification	Value
Percentage of Detection (PoD)	>90%
Max average FP per image	>1
MAE on defect length measurement	<3 mm
Scene area size	1200 × 800 mm
Minimal UV irradiance on scene area	1500 μW/cm^2
Maximal vision system size	700 × 400 × 150 mm
Min defect length	3 mm
Min defect width	50 μm

The assistive "PARADES" machine was briefly described in [3] and updated information can be found on the official company website (https://www.cmphy.fr/nos-produits/276/systeme-de-vision/, accessed on 7 February 2024). It consists of a standalone vision system with industrial-grade mechanical enclosure outer layer, powered from a standard 230 V socket. The operator can remotely use an external display as well as keyboard/mouse inputs as depicted in Figure 1b. A comparative control performed by a human operator with and without a PARADES machine is depicted in Figure 1.

Our main contributions are the following:

1. We developed an industrial machine for defect detection that can be integrated into every production line.
2. We proposed a way to design a customized lighting panel with specific simulation and optimization software.
3. We implemented a tool to measure defect length no matter the distance between the part and the PARADES machine.

The next sections will describe the related work (Section 2) and the PARADES machine, both in terms of hardware (Section 3) and software (Section 4). Hardware includes a custom UV-A lighting panel (Section 3.1), the RGB vision system (Section 3.3) as well as the depth sensor (Section 3.4) and an embedded AI-ready computer. The software section will detail the AI-based defect detection module (Section 4.2.1), the dedicated dataset in Section 4.1 (which was built using the PARADES machine itself) and the defect length measurement feature (Section 4.3). The results are then presented (Section 5), followed by the conclusion and future works.

2. Related Work

A substantial number of works have been found in the literature on the MPI topic. For instance, [4] shows an industrial application of a conventional computer vision algorithm on a rails wheelset. The non-standalone system built around a camera and a small lighting system must be coupled with an external PC for image processing, moreover, PoD results are not given and only qualitative examples are shown. Another example of industrial automatic defect detection in MPI is given by [5], showing material selection for specific image capturing on MPI. As stated in this article, optical filters are mandatory, and thus were used with the PARADES machine. A mobile machine based on a robot is depicted in [6] for use with MPI control on wide parts. It shows an interesting embedded magnetic excitation material as well as an MPI suspension dispenser. The defect-detection feature is not developed and shows the need for an external computer in order to process acquired images, due to the embedded Raspberry Pi Zero low computational performance.

The study in [7] is another example of an industrial machine used to detect defects with the MPI technique along with computer vision. It uses hardware as well as the usual computer vision algorithms to detect flaws. The paper focuses on detection speed rather than Percentage of Detection (PoD) and shows an insufficient detection accuracy (around 50%) for flaws less than 10 mm.

An experimental hardware setup is presented in [8] and aims to improve defect-detection precision using a conventional computer vision algorithm based on Hough transform. It visually shows how the false detections are removed but quantitative metrics are not given. Another approach was taken in [9], in order to find the 3D reconstruction of cracks on parts. The result is quite similar detection conditions as in PARADES, with defect size representing 1/4000 of the entire image with a few pixels for the defect width (7 pixels in their case). The scene area seems a lot smaller than ours, with a 7-pixel defect width at 2046 × 2040 full resolution compared to the PARADES high-definition >96 Mp camera system.

An experimental platform is detailed in [10] and uses conventional image processing (i.e., a Canny operator); however, it cannot be used in a production line. Another example of an AI-based defect-detection system is presented in [11], which uses a CNN coupled with a set of cameras to find defects on a specific crankset. A 3000 × 4096 pixel camera is used to feed a ResNet34 CNN, which has been trained on 100 × 100 image patches. This study shows a very high score on classification and a 0.9 mAP@0.5 (mAP computed with IoU score over 0.5) for object detection. A segmentation AI-based method is proposed in [12].

Study [13] shows a good 96.5% recall with 91.7% precision on industrial bearing using a custom mobileNetV3 CNN but the hardware setup is not presented.

To summarize, a lot of work has been done on MPI using image processing and deep-learning methods, achieving better performance than human controllers while lowering risks and improving working conditions. However, most of the studies concentrate on developing the software side with a laboratory minimal hardware setup most of the time. The literature is missing industrial-grade production-ready hardware for MPI automatic defect detection. The goal of this paper is to present the PARADES machine, which is designed as a versatile platform with industrial production use in mind and with a focus on AI-ready image acquisition. The PARADES machine aims to achieve the best detection performance in MPI but might be sufficiently generic for a variety of industrial computer vision tasks.

3. Hardware

Fluorescent Magnetic Particle Testing uses 365 nm-based UV illumination. The PARADES machine is built around a custom designed lighting system that matches the international aeronautics standard limit of 1500 $\mu W/cm^2$. Indeed, international standard

[14] ISO 9934-1:2015 requires a minimum of 1000 μW/cm^2, but the PARADES machine will also be used by aeronautics manufacturers and so needs to reach those specific standards. Airbus Helicopter, Rolls-Royce, Turbomeca and other manufacturers have defined their own standards. For example, a minimum of 1500 μW/cm^2 is required (Airbus Helicopter EI070, Turbomeca TS-00616, Paris, France) . In terms of spectrum, the requirement of the ISO standard [15] (ISO 3059:2012) matches aeronautics manufacturers and the most restrictive requirement is made by Airbus Helicopter. In this case, the UV peak must be within 365 and 370 nm. The FWHM's most restrictive criterion is given by Rolls-Royce RRES-90061 (Derby, UK) as a maximum of 20 nm. The PARADES specification is directly derived from those standards, so the minimal lighting irradiance will be 1500 μW/cm^2 with a peak between 365 and 370 nm and an FWHM <20 nm, in order for PARADES to be able to be used in a variety of domains and by many customers. In the next section, we describe our custom lighting panel based on flood and narrow beam UV spots. As explained in the previous section, the illuminated scene size is 600 × 400 mm. First, we will test two versions of UV spots in order to characterize the relative emitted irradiance regarding angle. Then, multi-spot lighting simulation software which has been specifically developed is presented. Next, we propose to optimize UV spot location in order to fulfill the 1500 μW/cm^2 requirement on the 600 × 400 mm scene with a minimum number of spots, based on spot characterization and a simulation algorithm.

3.1. Wavelength Measurement

As explained in the previous section, the standards requirement, in terms of spectrum, is a peak between 365 and 370 nm with FWHM less than 20 nm. This validation was performed using a calibrated Hamamatsu mini spectrometer model TG-C9404CAH (Hamamatsu, Japan) with specifications given in Table 2.

Table 2. Spectrum measurement device Hamatsu TG-C9404CAH.

Feature	Value
Spectral response range	200 to 400 nm
Spectral resolution (FWHM) (typ.)	1 nm
Spectral resolution (FWHM) (max.)	2.2 nm

The measured spectrum is depicted in Figure 2 and shows a peak located at 369 nm with an FWHM of 11 nm. The results are summarized in Table 3.

Table 3. Spectrum data of narrow and flood UV spot compared to requirements.

Characteristic	Requirement Value	Ours
Main spectrum peak (nm)	365 < peak < 370 nm	369 nm
FWHM (nm)	<20 nm	11 nm

The characterized UV spots are within the standard requirement; they can then be used in the PARADES Machine.

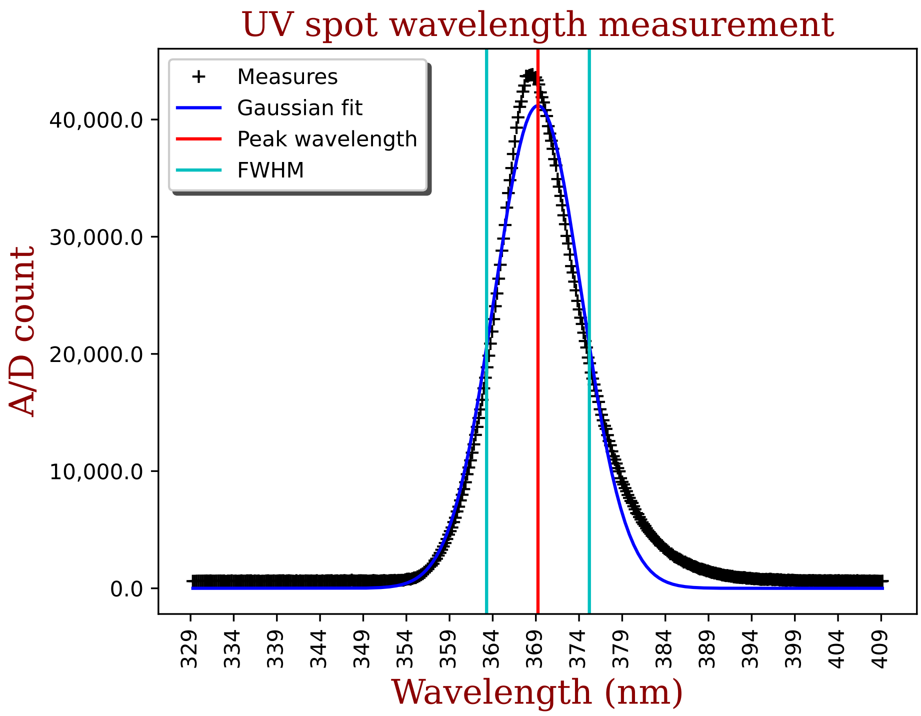


Figure 2. Measured spectrum of both narrow and flood UV spot used in the PARADES machine.

3.2. Custom UV Lighting Panel

Industrial UV spots are used in order to light a surface of 600 × 400 mm with a minimum irradiance of 1500 µW/cm^2. A few 365 nm UV spots have been used in the PARADES machine. Two versions (flood and narrow beam) were used in a specific combination in order to reach the 1500 µW/cm^2 requirement on all of the 600 × 400 m scenes.

3.2.1. Characterization Setup

The setup consists of a tripod supporting the UV spot to be characterized, over a graduated line where a radiometer is moved along at a defined vertical distance of 68 cm above the scene area. The setup is depicted in Figure 3. A measurement is done using a radiometer every 1 cm, ranging from 0 to 40 cm. A horizontal radiometer position from spot beam axis is then converted to an angle using the distance between the spot and the scene surface.

The radiometer used to measure UV light irradiance is a calibrated Pfinder UVLuxCHECK.

3.2.2. Measures and Characterization

Figures 4 give the measured and the interpolated data both with narrow (Figure 4a) and flood spots (Figure 4b). Those figures represent the relative irradiance versus the angle from vertical direction. As we can see, the flood spot (Figure 4b) has a wider emitting beam angle (35°) than the narrow beam spot (19°) shown in Figure 4a.

The flood spot emission feature is interpolated using a quadratic function clipped from 35°. The narrow spot is characterized using a SplineTransformer with 14 knots and degree 3. Despite the 68 cm distance between the radiometer and the spots, those measurements are then converted into an angle, allowing the use of those spots at any vertical distance.

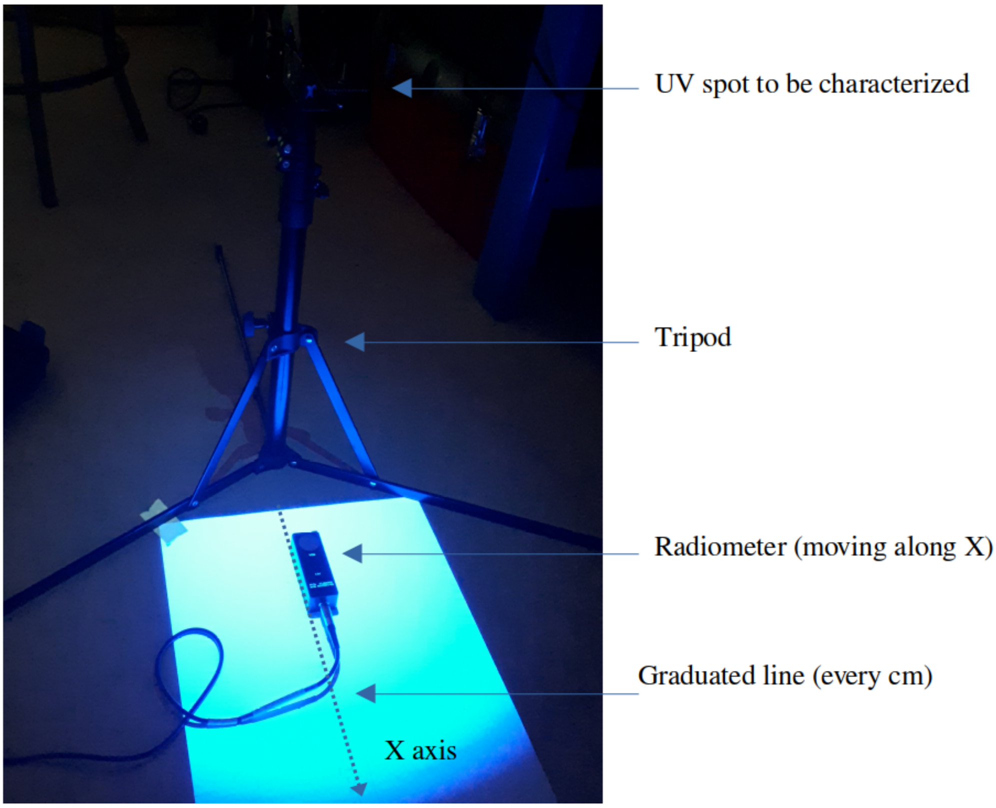


Figure 3. Hardware material setup to characterize UV spots.

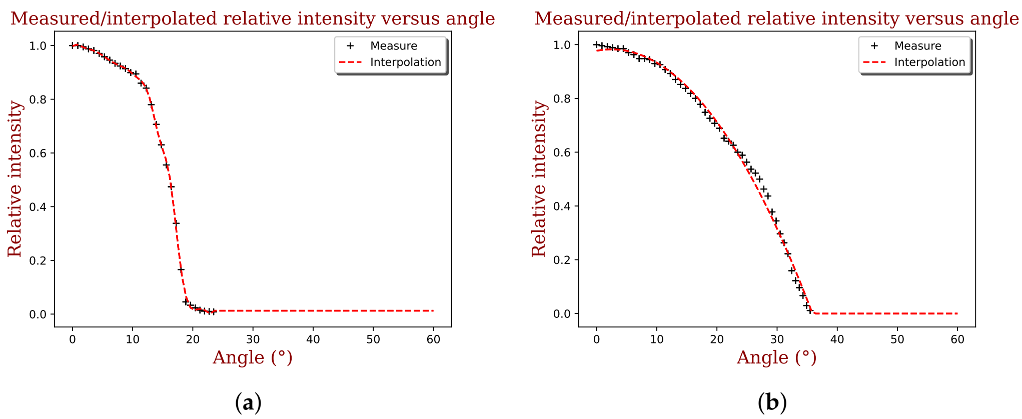


Figure 4. Measured and interpolated narrow and flood spot characteristics. (**a**) Relative irradiance regarding angle for narrow beam UV spot. (**b**) Relative irradiance regarding angle for flood beam UV spot.

The estimation error used is RMSE between measured and interpolated data. The narrow and flood spot RMSEs are given in Table 4 and show an error around 5 μW/cm^2 which is low compared to the requested 1500 μW/cm^2. Basically, an RMSE less than 10 μW/cm^2 is considered as acceptable.

Table 4. RMSE error for flood and narrow beam UV spot characterization.

Spot type	RMSE (μW/cm^2)
Flood	5.25
Narrow	6.52

3.2.3. UV Lighting Simulation

Based on those results, we are able to estimate the amount of UV light received at any location (X and Y) of our studied surface (e.g., 600 × 400 mm) by either a narrow or a flood spot located in a parallel plane distant from a defined distance (50 cm in our case). This is given by the spot characterization performed in the previous section. The next step is to estimate the received amount of light when a given set of narrow or flood spots are used with their own spatial disposition. Please note that in our case, all spots are located on the same surface, parallel to the illuminated surface and 50 cm distant. Basically, the UV irradiance received by a X,Y point from several spots is the sum of the irradiance received by each spot at this location. This computation is done for every point of a 1 × 1 mm grid pattern of the 600 × 400 mm studied surface. In order to perform the computation, simulation software written in Python has been specifically developed and uses the spots characterization. A screenshot of a simulation performed with our software is shown in Figure 5.

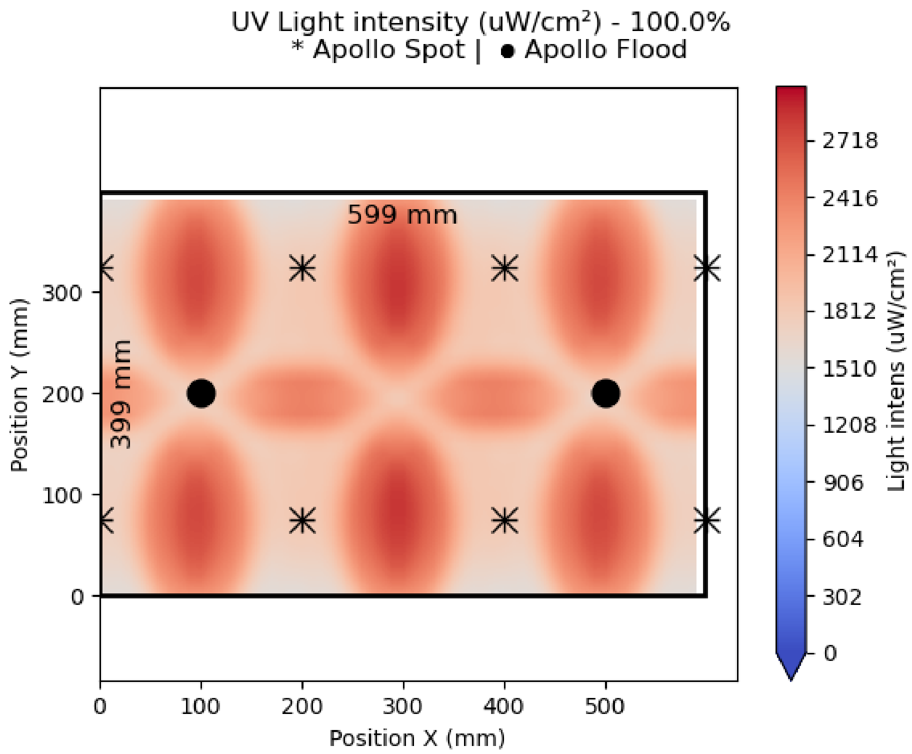


Figure 5. Two-dimensional visualization of UV light received.

3.2.4. UV Spots Disposition Automatic Optimization

Achieving 1500 µW/cm^2 on every point inside our studied surface of 600 × 400 mm is not an easy task. Trying and testing several spot dispositions in order to reach the minimum UV light specification will require a lot of tests with unlimited combinations of:

- Spot type (narrow, flood);
- Spot number;
- Spot disposition (X and Y).

To obtain a working disposition (all of the 600 × 400 mm surfaces above 1500 µW/cm^2), we used an optimization algorithm based on our light simulator results which gives us a suitable spots disposition (X and Y for each spots). Spot number and spot type will be tuned manually. Optimization is done using a combination of "Simulated annealing" and the "Nelder–Mead" algorithm, both implemented with Scipy library in Python. The optimized output variables are the (XY) location of each spot. All spots are considered on a fixed horizontal plane, on the same Z location as the camera system (e.g., 50 cm) above the illuminated surface.

In our tests, a minimum of 10 spots are required, with the following mix:

- 8 narrow spots;
- 2 flood spots.

The validation step was performed using the same Pfinder UVLuxCHECK radiometer, measuring the absolute received UV irradiance on a 10 cm × 10 cm grid pattern under the lighting system at 50 cm distance (Figure 6), covering a surface of 600 × 400 mm. The results show a perfect similarity with the simulated one, since no measured points were less than 1500 μW/cm^2.

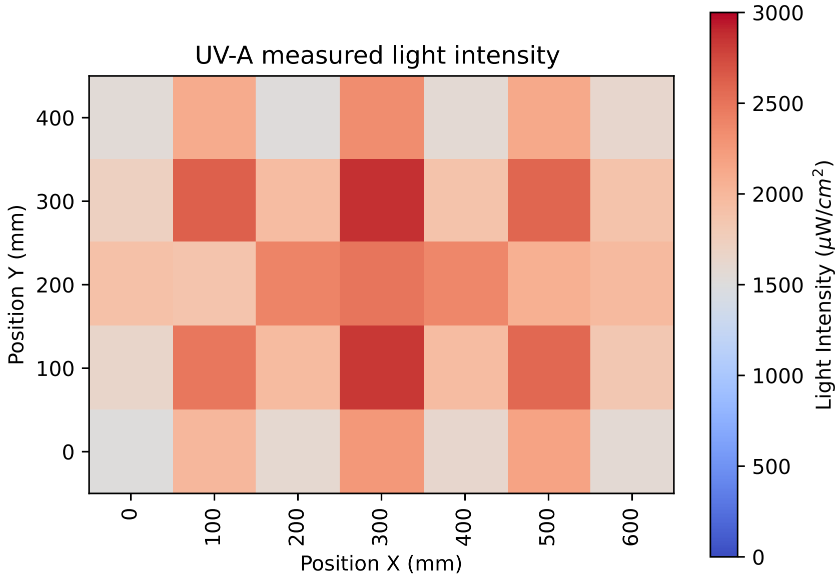


Figure 6. Measured UV light irradiance under the lighting system at 50 cm distance.

3.3. High-Definition Camera Imaging System

The machine must be able to detect defects with length above 3 mm with at least 50 μm width. In order to achieve this requirement, we state the hypothesis that we need at least one pixel representing the defect width (e.g., 50 μm).

Contrary to the initial scene size requirement (1200 × 800 mm), we decided to reduce this value to 600 × 400 mm due to hardware and performance limits. Considering we need at least 1 pixel on the defect width, this leads to:

- 50 μm on 400 mm → 8000 pixels;
- 50 μm on 600 mm → 12,000 pixels

We need at least a sensor with 8000 × 12,000 (96 M) pixels, so we choose a camera sensor with more than 100 M pixels fulfilling the requirement.

3.4. Depth Camera System

In addition to defect detection, a human controller needs to measure the defect length. Defects are located on a part's surface. Due to the complex and unpredictable 3D shape of the part, the distance between the defect and the 2D camera could be different from 50 cm, leading to an over or under length estimation, if length is computed from a 2D image only with a fixed pixel-to-distance ratio. In order to overcome this issue, a depth camera was used to obtain the previously unknown distance between the defect and the camera system, enabling an accurate defect length measurement. The integration of the depth sensor inside the vision system is depicted in Figure 7.

The depth camera is able to provide 2D images (Figure 8a) as well as a depth array (Figure 8b). Both data can be merged using the manufacturer SDK. A custom method to fill the missing depth point information has been implemented as shown in Figure 8c.

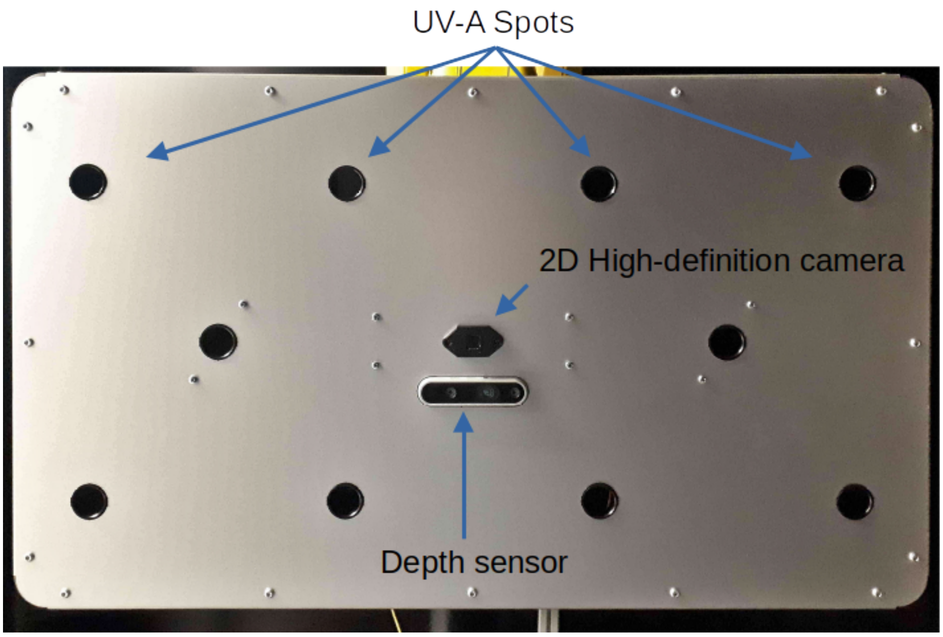
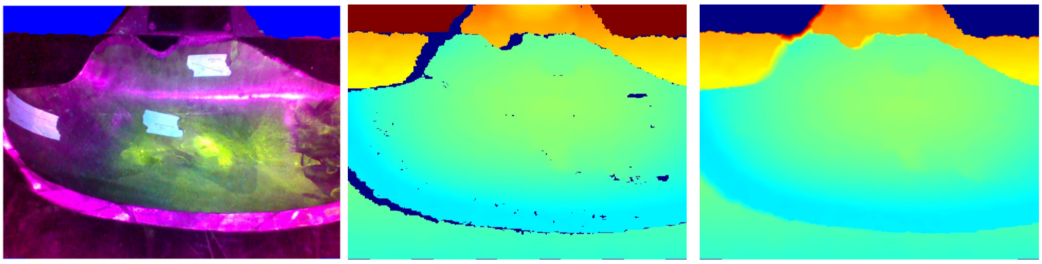


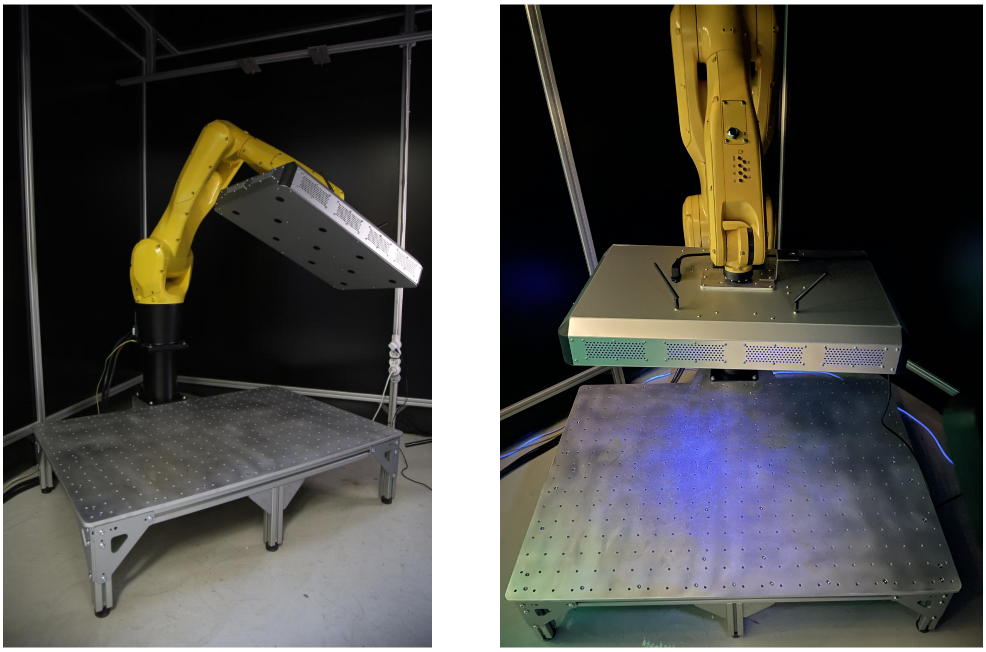
Figure 7. Integration of depth sensor along with 2D camera and UV-A spots in PARADES standalone vision system.



(**a**) (**b**) (**c**)

Figure 8. Output images from depth camera system. (**a**) 2D image. (**b**) Colorized depth image. (**c**) Processed depth image.

The complete hardware setup is depicted in Figure 9a,b.



(**a**) (**b**)

Figure 9. Pictures of PARADES machine ready for production. (**a**) PARADES vision system. (**b**) PARADES vision system with UV lighting on.

4. Software

The hardware detailed in the previous section permits us to acquire MPI images in a controlled and repeatable way, thus enabling efficient AI-based defect detection. The images acquired by our system are >100 MP where defect length is about 3 $\rightarrow$ 20 mm with a minimal width of 50 µm. Only one class representing defects is used which changes from the usual paradigm in object detection where images are usually smaller (640 $\times$ 480 pixels in COCODataset) with more classes (80 classes in COCODataset). In this case, objects to be detected are only four or five times smaller than the image size. Those characteristics remain the same with the PascalVOC dataset with 207 images of size 500 $\times$ 375 pixels with 20 classes. With our hardware, in our use-case, a small defect (50 µm $\times$ 3 mm) is $\sim$8000 times smaller than the image size, leading to a lot of missing detections if using current unchanged object-detection models. An example of an acquired MPI processed part is given in Figure 10. The part size is 80 $\times$ 80 mm in the 600 $\times$ 400 mm scene and as observed in this figure, a defect is present on the part with a size of $\sim$20 $\times$ 20 mm. In addition, none of the trained object-detection models have labels for cracks and none of the image datasets used during training have fluorescent colorimetry, contrary to our acquired images. Indeed, usual crack datasets such as in [16] represent cracks in visible light but not using UV-based fluorescence. Moreover, cracks appear a lot bigger in the acquired image than defects seen with the PARADES machine. Based on those observations, a PARADES dedicated dataset is mandatory and so must be created. For that purpose, a custom UV-based MPI dataset is first built using PARADES hardware and will be detailed in the next section. Then, a specifically tuned object-detection module based on Faster R-CNN will be trained and presented. Finally, a defect length measurement feature is detailed.

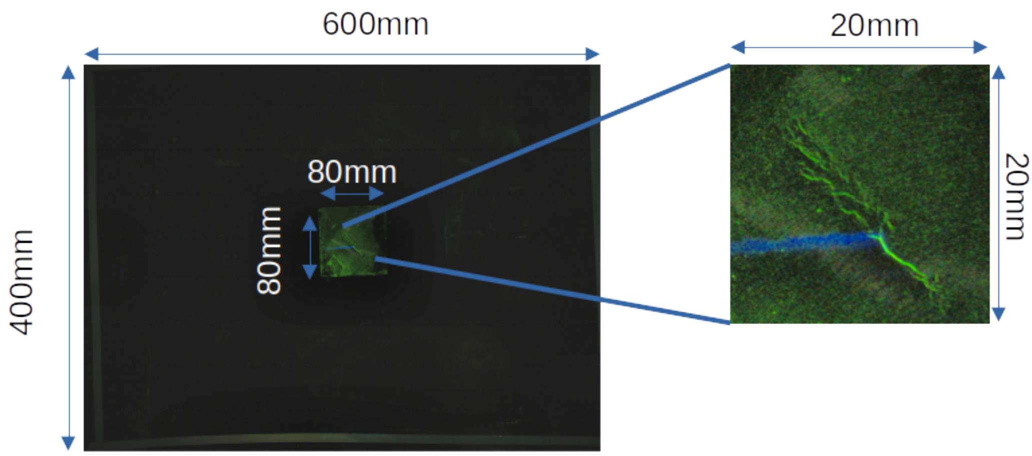


Figure 10. An example MPI result acquired with PARADES machine.

4.1. MPI Custom Dataset

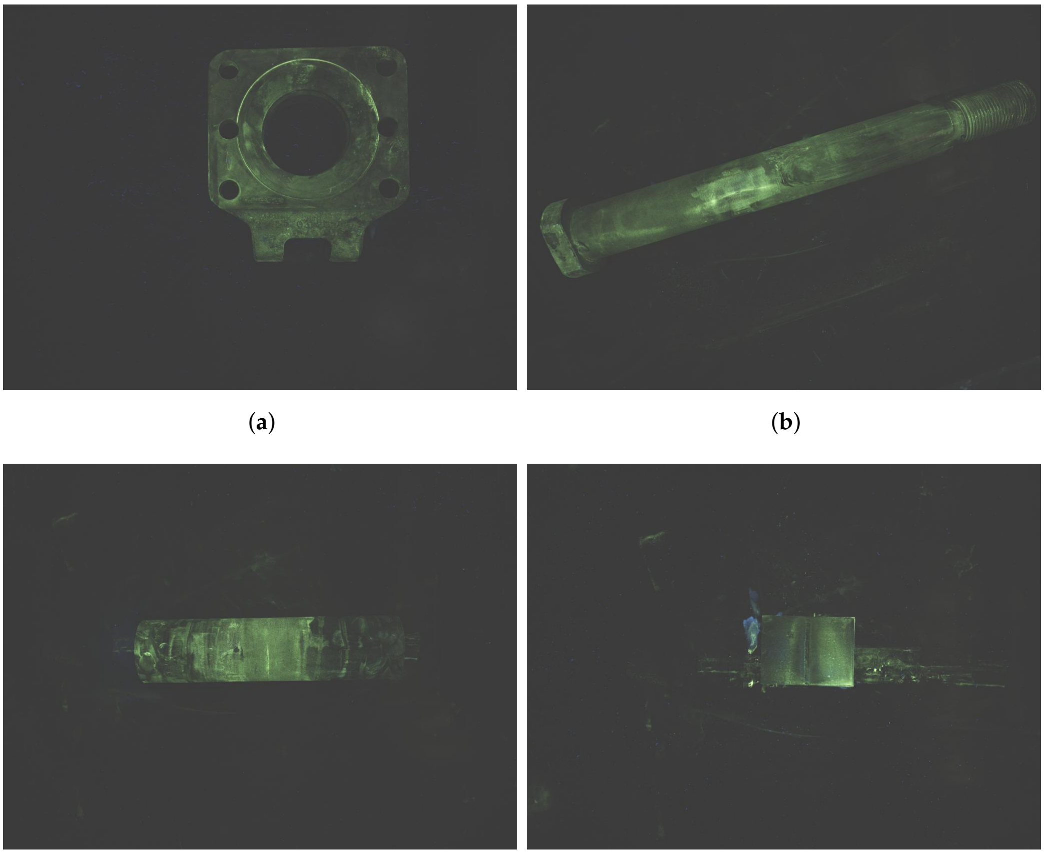
A dataset of images similar to those acquired via the PARADES system is mandatory to train our model. The images in this dataset must match the distribution of images that will be acquired by controllers using the PARADES machine in the production line. Therefore, the images must be green fluorescent (linked to the MPI process) with defects (linear scratches) ranging from 2 to 25 mm length in a 600 × 400 mm field of view. Those images must show mechanical steel parts with granulated and smooth textures in order to represent forged, milled and foundry parts. In the literature, previous similar works in deep learning in MPI have been found [9,11–13]; authors had to build their own dataset because there is no such dataset available publicly on the market. Unfortunately, those authors did not give access to their datasets. On the other hand, similar dataset in other fields could have been used, datasets of cracks on concrete [17,18] representing scratches, but unfortunately, images in those datasets are not green fluorescent and furthermore cracks appear bigger in the images than in our case. Another approach could have been to use the NEU-DET dataset [19] or GC10-DET dataset [20], dedicated to steel surface defects. However the images are not fluorescent; images are small compared to our required 600 × 400 mm field of view. In addition, the images in those datasets show defects other than scratches (crazing, inclusions, patches, etc.) which are useless in our case of MPI, containing only linear scratch defects. Regarding those considerations, we decided to build our own dataset, with the PARADES machine. Existing customer (forgery, foundry, etc.) parts have been used, along with manually created defects on metal plates or tubes in order to increase the number of total defective parts.

Some image examples of our built dataset are given in Figure 11.

The complete dataset consists of 233 train images and 29 test images. Some continuous, discontinuous, curved or straight line defects are presents. Defect length varies from small (3 mm) to longer defects (>25 mm). Several types of manufactured parts compose the dataset, such as welded, milled and raw foundry parts. Defect types can be foundry defects (Figure 12) or defects induced by mechanical stress (Figure 13).

An example of a discontinuous foundry defect is shown in Figure 12.

For the trainset, some smartphone-based (not PARADES machine) images acquired by a human controller during control sessions have been also included. The testset only includes images from the PARADES machine with all the diversity explained previously. This dataset is still under construction and will continuously be completed. It contains some proprietary manufacturer parts and so cannot be publicly available at the time of writing.



(**a**) (**b**) (**c**) (**d**)

Figure 11. Customer's part used in our dataset (1st row) and manually created parts (2nd row). (**a**) Forgery defect on customer's part. (**b**) Stress defect on a customer's screw. (**c**) Manually created defect on a metal tube. (**d**) Manually created defect on welded metal sheet.

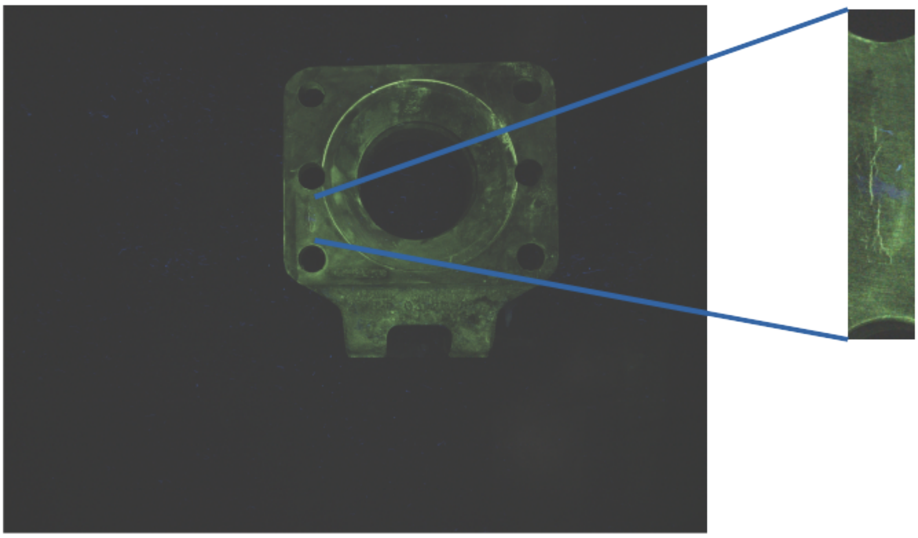


Figure 12. Example of a foundry defect on image acquired with the PARADES machine.

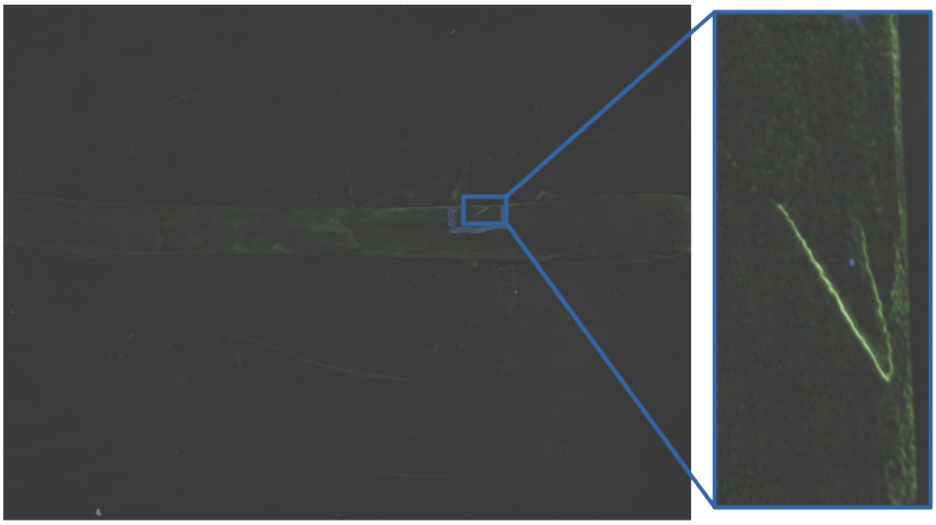


Figure 13. Example of stress-induced defect on a rod, acquired with the PARADES machine.

4.2. Defect Detection

4.2.1. Faster-RCNN Architecture

The usual available AI-based architectures for object detection are:

- Yolo [21];
- Faster R-CNN [22];
- Transformer-based (Deformable-DETR [23]).

Real-time detection is not a requirement for our project (the maximum allowed detection time is set to 60 s), but having a high PoD (which is similar to high recall) is mandatory. This eliminates Yolo-based architecture which is faster than the accurate sort due to its single stage architecture, contrary to Faster-RCNN [19,24] built around a two-stage architecture. Transformers (*-DETR) using ViT cannot be used because of the non-uniformity of the lighting. Indeed, the attention mechanism will detect non-uniformity as possible defect zones, leading to a high number of false detections.

Faster-RCNN is based on feature extraction (backbone), then a Region Proposal Network (RPN) followed at the end by the classification. Backbone based on convolution layers provides features to the RPN. RPN selects the best possible regions of interest (RoIs), to feed the classifier (Figure 14). The RPN part is relevant to our defect detection. Indeed, this defect represents 1 to 5 % of the whole image. Therefore, selecting RoI is more efficient. In this study, we used the official Pytorch implementation of Faster R-CNN.

During training, trainset images were compressed to be able to train with the CCUB (Centre de Calcul de l'Université de Bourgogne) cluster. The validation on the testset was carried out with the original image size. Data augmentation [9,11,13,20] was performed on the trainset only, with rotation as well as horizontal and vertical flips. This is motivated by the fact that during machine utilization defects can appear rotated and horizontally or vertically flipped. Data augmentation is implemented using a Python library called 'Albumentations'.

Transfer learning [11,25] using argument 'weights' set to 'IMAGENET1K_V2' was used with Pytorch implementation of Faster R-CNN to obtain the best possible recall with a limited number of training epochs. Indeed, only the fully connected layers used for classification are trained; the backbone remains frozen. This is useful for re-using the feature extraction capability of the backbone trained with a huge dataset, while specializing our network for our specific detections.

The PARADES machine is a commercial product; thus, the source code and implementation details cannot be released.

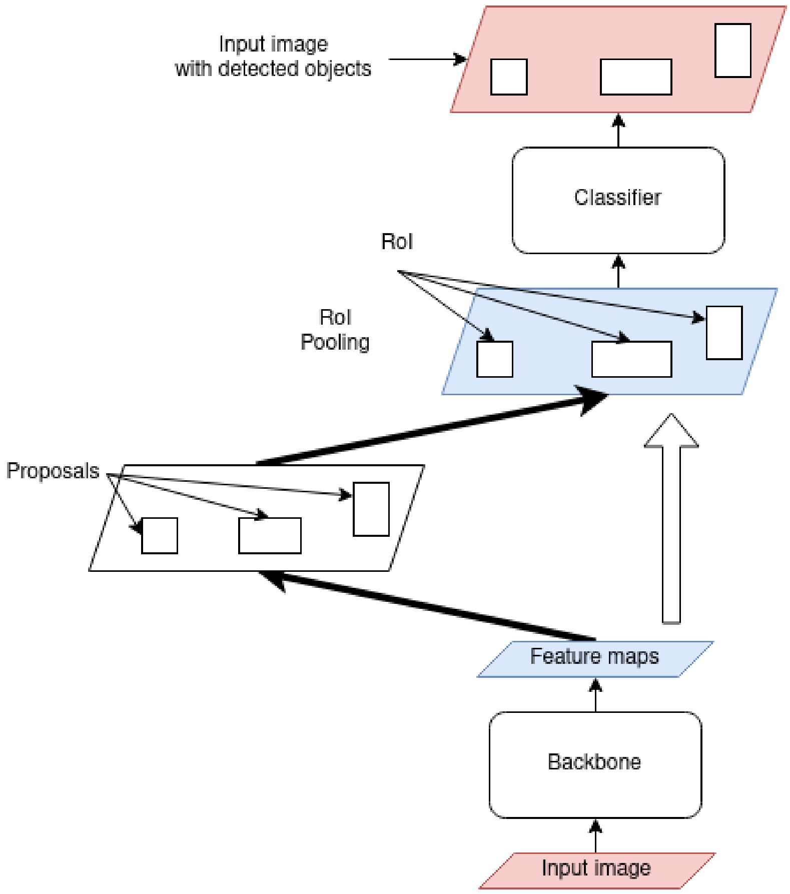


Figure 14. Faster R-CNN architecture.

4.3. Defect Length Measurement

Semi-autonomous length measurement is carried out by human controllers using a specifically developed software tool upon parts of acquired images from the PARADES system. Using a drawing tool coupled with depth information, the operator draws lines or splines on a merged 2D/3D image from a depth camera as shown in Figure 15. The 2D image from the depth camera is relatively close to the high-definition camera image used for detection: the field of view is quite similar and both are positioned close to each other and oriented the same way.

The defect length measurement tool provides two ways of measuring to the operator:

- The bird's eye distance between two points;
- the path length measurement (using pencil drawing).

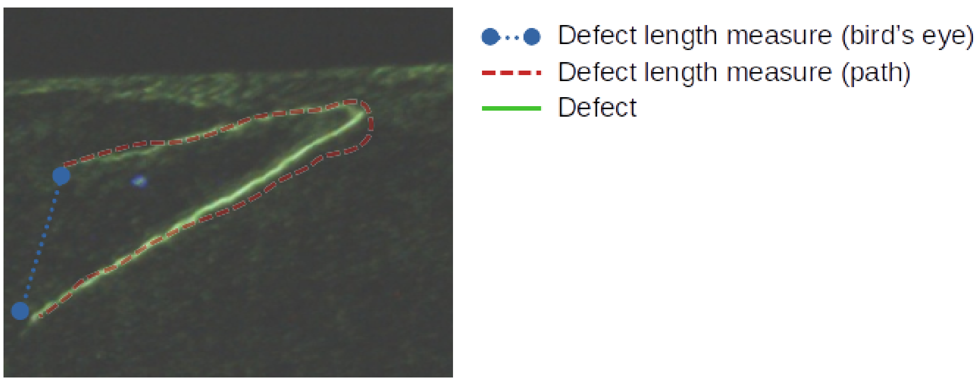


Figure 15. Two types of defect length measurement: straight line (bird's eye) and path.

Thanks to the depth camera system, pixels drawn by operators are converted into 3D voxels, using the manufacturer's SDK. The 3D coordinates of voxels in real space are extracted from depth array using SDK. To compute the path length (red line in Figure 15) we discretize the path drawn by the operator using the 3D voxels. The bird's eye distance (blue line in Figure 15) is computed using only the first and the last point. Both lengths are computed using euclidean distance.

5. Results

The goal of the PARADES machine is to help human operators detect defects between 3 and 25 mm on MPI processed parts. The first measuring performance criterion is PoD (same as recall in classification), with a requirement of >90%. Indeed, the PARADES system is used as a semi-autonomous machine to provide help for operators detecting defects. In this case, the machine shall not miss any defect but false detections will be easy and quicker for an operator to manually delete. So, precision is less important and is converted into average false detection per image to be more understandable by customers and human controllers. Average false detection per image shall not be more than 1. For defect length measurement, the Mean Absolute Accuracy (MAE) is chosen and must be less than 3 mm.

5.1. Defect Detection

A common performance metric in object detection is mean Average Precision (mAP). This represents the area under the recall/precision curve. In this study, according to the customer's requirements, the Intersection over Union (IoU) threshold is set to 0.02 for mAP computation. This can be seen as very loose but it makes sense for this use case. PARADES detection is then analyzed by a human operator who will adjust the surrounding box of detections to perfectly fit the defect.

The recall/precision curve of the PARADES machine, at IoU threshold = 0.02 is given in Figure 16. The mAP is 0.9043 with f1_max equal to 0.8434. To reach the minimal requirement of recall = 0.9, the chosen confidence score threshold is 0.31579, leading to:

- Recall = 0.9012 (>90%);
- Mean false positive per image = 0.9 (<1).

An example of defect detection on a test customer's part is given in Figure 17. As we can see, a defect is not detected (FN), there is a detection that does not match an actual defect (FP). Three other present defects (GT) are successfully detected (TP).

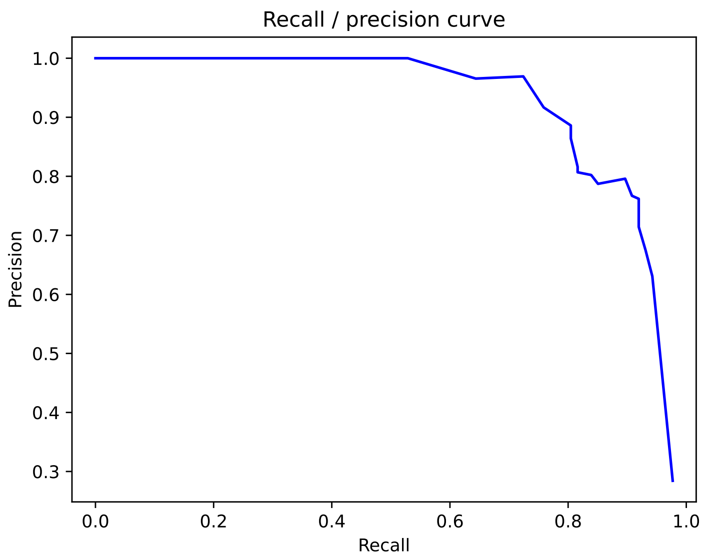


Figure 16. mAP@0.02 of the PARADES defect detection AI model.

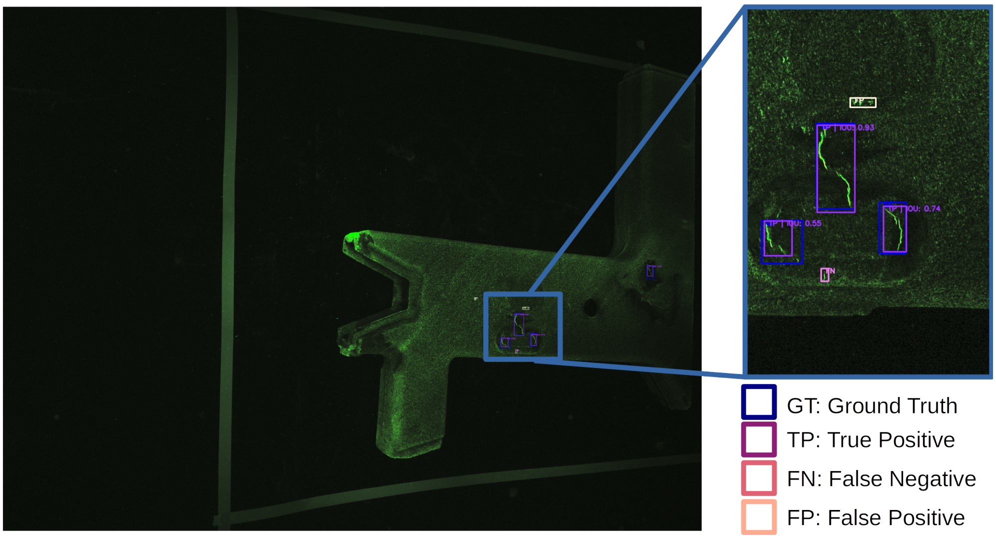


Figure 17. Example of detections on testset with correct detection (TP), missing detection (FN) and false detection (FP).

At this stage, the defect-detection model reaches the minimal recall requirement (i.e., 90%) without exceeding the average false detection per image limit (i.e., 1 FP per image). Increasing the number of parts in our dataset may increase those performances, as well as ongoing fine-tuning operations, enabling a possible fully autonomous machine.

5.2. Defect Length Measurement

Length measurement accuracy is defined as MAE on four artificially created ground truth straight paths representing defects. Those paths are straight lines ranging from 20 to 60 mm drawn on stickers which are applied on the curved surface of a real customer's part, as depicted in Figure 18. The results are summarized in Table 5. The computed MAE is 0.925 mm. We can see that MAE is not proportional to defect length. Indeed, the operator draws the defect line on an unzoomable image. This leads to operator-dependant position errors.

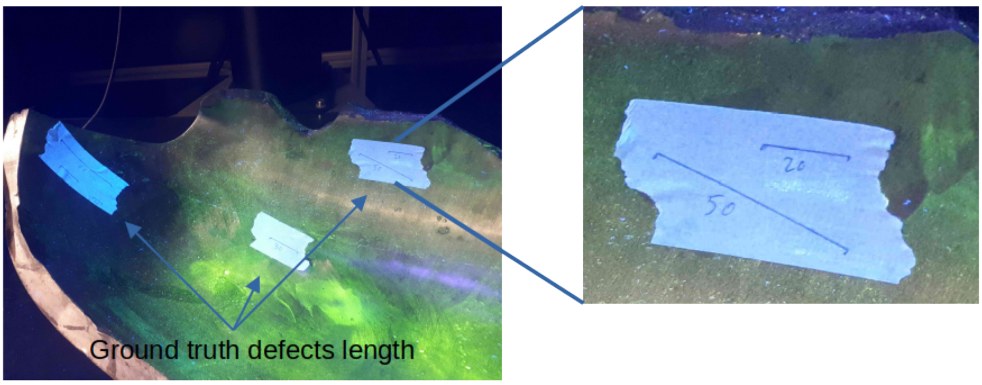


Figure 18. An example part with ground truth defect length (20 to 60 mm).

Table 5. Defect length measurement performance.

Ground Truth (mm)	Measured Value (mm)	Relative Error (%)	Absolute Error (mm)
50.0	51.8	3.6	1.8
20.0	20.0	0.0	0.0
60.0	60.1	0.2	0.1
30.0	28.2	6.0	1.8

Table 6 summarized the PARADES machine performances compared to previous work. The paper [7] shows an industrial machine used to detect flaws longer than 10 mm and wider than 0.3 mm on steel wire rod and bar. The system is designed for a specific production line at Kobe Work, with a dedicated eight camera system. The work carried out in [8] shows an experimental setup to detect flaws of 15 mm length with widths of 0.2 mm, 0.4 mm, 0.6 mm and 0.8 mm. The machine is dedicated for small parts, for example a bearing of size 20 mm $\times$ 20 mm $\times$ 20 mm. Authors in [9] present a robotic MPI defect-detection system with external PC as computing unit. The machine is used to detect small defects on big mechanical parts (a defect represents 1/4000 image size). The paper [10] presents an automatic device for MPI defect detection dedicated to crankshaft. An experimental setup is presented with one UV light bulb and a camera, plugged into an external PC as a computing unit. In the paper [11], the authors present another way of detecting defects on crankshaft using the MPI technique. Authors in [13] present a software block to detect defects with MPI on bearing rings. The hardware side is not detailed.

As we explained in the previous paragraph, only some of the related works are industrial-grade products [9,11]; the others are experimental setups that cannot be used in industry. In those works, the detection performance (i.e recall) is either not given or is below our expectations. Only [9,11,13] perform better than our needs. In addition, some presented machines are dedicated to a chosen mechanical part; the system cannot be used for another part. Only [10,11] can be used with several mechanical parts. None of the presented related works offer a tool to measure defect length nor allow for wide mechanical part image acquisition (600 $\times$ 400 mm).

Based on this comparative study, we show that none of the previous works had success building a versatile and automatic defect-detection machine in MPI which can be used in production lines. Some works present advanced mechanical construction, but specifically dedicated to a kind of mechanical part. Other works show improvement in terms of defect-detection performance, but with an experimental and non-industrial machine that cannot be used in production lines. The PARADES machine gathers the following two important features: a versatile production-ready mechanical machine and high-performance defect-detection software. In addition, the PARADES machine provides a non-existing tool to accurately measure the defect length whatever the distance between the part and the machine.

Table 6. Comparison of automatic defect detection machines for MPI.

Machine	Industrial Grade	Defect Length Measure	Versatile Applications	>90% Recall	Field View 600 $\times$ 400 mm
[7]	?	No	No	?	No
[8]	?	No	No	No	No
[9]	Yes	No	No	Yes	No
[10]	No	No	Yes	?	No
[11]	Yes	No	Yes	Yes	No
[13]	?	No	?	Yes	No
PARADES	Yes	Yes	Yes	Yes	Yes

?: This property is not provided in the cited paper.

6. Conclusions

In conclusion, we have proposed the first semi-autonomous defect detection for MPI with a novel hardware block, AI-ready, standalone, performing UV-A lighting, image acquisition and detection for several kinds of visual NDT tasks. The versatile hardware built around a lighting system, a camera and the computation module in an industrial way should be used for any tasks requiring vision computer. This hardware is perfectly characterized, in particular the lighting system, enabling customization for smaller or wider part checking. Electronics is designed to be able to use human-spectrum UV spots for usual visual inspection. Features and the purpose of the machine can be easily extended with dedicated software development. Indeed, the Arm-based, AI-ready computer runs Ubuntu to make new application developments easy.

The PARADES machine is operated from outside the darkroom, so controllers are able to work in a clean environment with natural lighting, without UV light only, thus improving operators' working conditions and health. The PARADES machine reaches the minimal required specifications in terms of detection performance (PoD) and defect length measurement, so it is ready for production.

In the future, we will implement a zoomable depth image to lower the position error during defect length measurement. Ongoing works are focused on improving PoD by increasing dataset size and testing the SOTA transformer-based model. The PARADES machine is going to be integrated into a customer production line. Then, finetuning and using TensorRT will be implemented. An FPT images dataset is under construction for detecting FPT defects. An upcoming stitching feature will allow one to cover an area of 1200×800 mm using four 600×400 mm image acquisitions.

On the other hand, a new smaller handheld PARADES machine is being designed, in order to fulfill small parts manufacturers' needs as well as bigger complex and inaccessible parts such as cumbersome nuclear or hydraulic installation parts. A human controller handles the machine or fixes it with a dedicated articulated mechanical arm.

Author Contributions: J.B.: conceptualization, software, methodology and manuscript drafting, H.C.: supervision, A.P.: resources and supervision, P.M.: funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by agence Nationale de la Recherche (ANR-21-PRRD-0047 01).

Data Availability Statement: The data presented in this study are available on request from the corresponding author. Dataset and AI software are not yet publicly available due to customer's restrictions.

Acknowledgments: This work was performed using HPC resources from GENCI–IDRIS (Grant 2023-AD011014417). Calculations were performed using HPC resources from DNUM CCUB (Centre de Calcul de l'Université de Bourgogne).

Conflicts of Interest: Authors Julien Baumeyer and Arnaud Pelletier were employed by the company CMPhy. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CCUB	Centre de Calcul de l'Université de Bourgogne
FP	False Positive
FPT	Fluorescent Penetrant Testing
IoU	Intersection over Union
MAE	Mean Absolute Error
mAP	mean Average Precision
MPI	Magnetic Particle Inspection
NDT	Non Destructive Testing
PARADES	Projet d'Automatisation de Reconnaissance et d'Analyse de DEfauts Surfaciques
PoD	Percentage of Detection
RMSE	Root Mean Square Error
SDK	Software Development Kit

References

1. Dwivedi, S.K.; Vishwakarma, M.; Soni, P.A. Advances and Researches on Non Destructive Testing: A Review. *Mater. Today Proc.* **2018**, *5*, 3690–3698. https://doi.org/10.1016/j.matpr.2017.11.620.
2. Smallbone, C. Welding and NDT—Enabling technologies to improve the global quality of life. In Proceedings of the 18th World Conference on Nondestructive Testing, Durban, South Africa, 16–20 April 2012.
3. Pelletier, A.; Crepeau, A.; Lemond, J.; Marquié, P.; Baumeyer, J.; Chatoux, H. Robotisation et Numérisation pour la Magnétoscopie et le Ressuage. *e-J. Nondestruct. Test.* **2023**, *28*. https://doi.org/10.58286/28517.

4. Ma, T.; Sun, Z.; Zhang, W.; Chen, Q. A machine vision assisted system for fluorescent magnetic particle inspection of railway wheelsets. *Proc. AIP Conf. Proc.* **2016**, *1706*, 150003. https://doi.org/10.1063/1.4940615.
5. Willcox, M. *Automatic Defect Recognition in Magnetic Particle Inspection Applications*; Technical Report; Insight NDT Equipment Ltd.: Bedford, VA, USA, 2000.
6. Liu, J.; Yu, H.; Mei, L.; Han, B. Development of a new type of automatic magnetic particle inspection wall-climbing robot. *Adv. Mech. Eng.* **2021**, *13*, 16878140211047339. https://doi.org/10.1177/16878140211047339.
7. Nakata, H.; Hirata, M.; Tada, T. Fully automatic magnetic-particle inspection system for square billets. *IFAC Proc. Vol.* **1993**, *26*, 1–4. https://doi.org/10.1016/S1474-6670(17)48210-3.
8. Chen, Y.; Kang, Y.; Feng, B.; Li, Y.; Cai, X.; Wang, S. Automatic defect identification in magnetic particle testing using a digital model aided De-noising method. *Measurement* **2022**, *198*, 111427. https://doi.org/10.1016/j.measurement.2022.111427.
9. Wu, Q.; Qin, X.; Dong, K.; Shi, A.; Hu, Z. A learning-based crack defect detection and 3D localization framework for automated fluorescent magnetic particle inspection. *Expert Syst. Appl.* **2023**, *214*, 118966. https://doi.org/10.1016/j.eswa.2022.118966.
10. Luo, J.; Tian, Z.; Yang, J. Fluorescent magnetic particle inspection device based on digital image processing. In Proceedings of the 11th World Congress on Intelligent Control and Automation, Shenyang, China, 29 June–4 July 2014, pp. 5677–5681. https://doi.org/10.1109/WCICA.2014.7053688.
11. Tout, K.; Meguenani, A.; Urban, J.P.; Cudel, C. Automated vision system for magnetic particle inspection of crankshafts using convolutional neural networks. *Int. J. Adv. Manuf. Technol.* **2021**, *112*, 3307–3326. https://doi.org/10.1007/s00170-020-06467-4.
12. Ueda, A.; Lu, H.; Kamiya, T. Deep-Learning Based Segmentation Algorithm for Defect Detection in Magnetic Particle Testing Images*. *Proc. Int. Conf. Artif. Life Robot.* **2021**, *26*, 235–238. https://doi.org/10.5954/ICAROB.2021.GS3-1.
13. Yang, Y.; Yang, Y.; Li, L.; Chen, C.; Min, Z. Automatic Defect Identification Method for Magnetic Particle Inspection of Bearing Rings Based on Visual Characteristics and High-Level Features. *Appl. Sci.* **2022**, *12*, 1293. https://doi.org/10.3390/app12031293.
14. ISO 3059:2012; Non-destructive testing, Penetrant testing and magnetic particle testing. ISO: Geneva, Switzerland, 2015.
15. ISO 9934-1:2015; Non-destructive testing — Magnetic particle testing — Part 1: General principles. ISO: Geneva, Switzerland, 2012.
16. Ozgenel, C.F. *Concrete Crack Images for Classification*; Mendeley Data: Burlingame, CA, USA, 2019; Volume 2. https://doi.org/10.17632/5y9wdsg2zt.2.
17. Silva, W.R.L.d.; Lucena, D.S.d. Concrete Cracks Detection Based on Deep Learning Image Classification. *Proceedings* **2018**, *2*, 489. https://doi.org/10.3390/ICEM18-05387.
18. Akgul, I. Mobile-DenseNet: Detection of building concrete surface cracks using a new fusion technique based on deep learning. *Heliyon* **2023**, *9*, e21097. https://doi.org/10.1016/j.heliyon.2023.e21097.
19. Zhao, W.; Chen, F.; Huang, H.; Li, D.; Cheng, W. A New Steel Defect Detection Algorithm Based on Deep Learning. *Comput. Intell. Neurosci.* **2021**, *2021*, e5592878. https://doi.org/10.1155/2021/5592878.
20. Zhou, C.; Lu, Z.; Lv, Z.; Meng, M.; Tan, Y.; Xia, K.; Liu, K.; Zuo, H. Metal surface defect detection based on improved YOLOv5. *Sci. Rep.* **2023**, *13*, 20803. https://doi.org/10.1038/s41598-023-47716-2.
21. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *arXiv* **2016**, arXiv:1506.02640. https://doi.org/10.48550/arXiv.1506.02640.
22. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv* **2016**, arXiv:1506.01497.
23. Zhu, X.; Su, W.; Lu, L.; Li, B.; Wang, X.; Dai, J. Deformable DETR: Deformable Transformers for End-to-End Object Detection. *arXiv* **2021**, arXiv:2010.04159. https://doi.org/10.48550/arXiv.2010.04159.
24. Fang, Z.; Ren, J.; Sun, H.; Marshall, S.; Han, J.; Zhao, H. SAFDet: A Semi-Anchor-Free Detector for Effective Detection of Oriented Objects in Aerial Images. *Remote Sens.* **2020**, *12*, 3225. https://doi.org/10.3390/rs12193225.
25. Mery, D. Aluminum Casting Inspection using Deep Object Detection Methods and Simulated Ellipsoidal Defects. *Mach. Vis. Appl.* **2021**, *32*, 72. https://doi.org/10.1007/s00138-021-01195-5.