Towards deep learning fusion of flying spot thermography and visible inspection for surface cracks detection on metallic materials

Kevin Helvig, Ludovic Gaverina, Pauline Trouvé-Peloux, Jean-Michel Roche, Baptiste Abeloos, C Pradere, Guy Le Besnerais

Towards deep learning fusion of flying spot thermography and visible inspection for surface cracks detection on metallic materials

by K. Helvig*, L. Gaverina**, P. Trouvé-Peloux*, J-M. Roche**, B. Abeloos*, C. Pradere***, G. Le Besnerais*

* *DTIS, ONERA - Université Paris-Saclay, F-91123, Palaiseau, France*
** *DMAS, ONERA - Université Paris-Saclay, F-92322 Châtillon, France*
*** *EPSILON-ALCEN, esplanade des Arts et Métiers 33400 Talence, France*

Abstract

"Flying spot" laser infrared thermography (FST) is a non destructive testing technique able to detect small defects through scanning surfaces with a laser heat source. Defects such as cracks can indeed be detected by the disturbance of heat propagation measured by an infrared camera. However this examination method is limited to small regions of interest and the measurement might be affected by heterogeneous surface properties. Moreover, visible spectrum enables the localisation of variations of properties on the surface and an inspection within large field of view in a single snapshot. However, such inspection could miss cracks with small dimension due to resolution issue. The fusion of defect detection of these two modalities could overcome these difficulties. Regarding data processing, deep learning approaches are now very efficient, first to automatically analyse and exploit context information from data, and second to conduct data fusion. Hence, we propose here to develop a new crack inspection method fusing FST and visible spectrum using deep learning. This paper presents our preliminary work towards this fusion. We first focus on IR spectrum, starting on bench settings optimization using both simulated and experimental data, in particular in comparison with the theory of Peclet number. Then we present FST automated defect detection results using deep learning and also study the sensitivity of the detection results to the experimental settings. Finally, we analyse the automatic defect detection results on both IR and visible spectrum separately, to illustrate the potential of fusing these spectra.

1. Introduction

The FST was originally developed for cracks detection in military aircraft parts in the end of the 1960s' [1]. In [2], Krapez describes the Peclet number theory, giving a ratio between convective and conductive heat flux, and how it will influence cracks detection for constant scanning speeds. Maffren uses FST to examine cracks on AM1 superalloy samples covered with thermal barriers, revealing how difficult it can be to detect cracks on heterogeneous surfaces [3]. A lot of work has been made in order to improve the technique, like flying-line thermography using a line of laser spot as heat sources in order to accelerate scans [4]. Simulation works using finite elements models (FEM) of flying-spot and flying-line has also been made, giving a way to study the thermal physics phenomena behind this examination technique and to plan experiment set-ups [5, 6]. There are still a lot of development around the theoretical basis of this technique associated with experiments, like for characterisation of detected defects, the reference [7].
Many image processing techniques dedicated to this NDT process have been proposed however, to the best of our knowledge, only the recent publication of [8] proposes to use deep learning complex recurrent architectures able to manage temporal features In visible deep learning for cracks detection is more developed, as in civil engineering [9]. IR-Visible fusion for surveillance has been explored in some papers, based on the creation of an intermediate image combining visible and IR images [10], and works using deep learning to fuse thermal and visible images are mostly driven by pedestrian detection for autonomous navigation [11] [12]. Data fusion between IR and visible using deep learning for automatic object detection in NDT using seems not to be yet developed in the literature.
In this paper, we present several contributions towards a new method for defect detection on metallic surface based on the fusion of IR and visible spectrum using deep learning. First, regarding FST we propose to use a single laser scan on the material. This reduces the inspection duration, but leads to more challenging processing of the thermogram in order to isolate cracks from edges. Second, we study the settings parameters that maximise the thermogram 1D signal response to a high frequency filter. We show in particular that the optimal settings regarding this response are consistent with the Peclet number theory. In order to process the single scan thermogram, we also propose to use deep learning patch classification approach that can benefit from context information to isolate cracks from edges. Using data acquired with our experimental bench with a large set of experimental settings, we train a deep learning architecture for automatic defect detection. We show that its performances are quite satisfying, and that its is also sensitive to experimental settings, consistently with the Peclet number theory. Finally, we analyse jointly defect detection using neural network on both IR and visible spectra separately, in order to highlight the potential interest of fusing these techniques.

2. Paper organization

The third section of this paper sums-up the main elements of flying-spot physics. Then we study in section 4 the impact of settings parameters on the amplitude of the thermogram response for single pass scans to a high frequency filter, on both simulated and experimental data and compare obtained results with the Peclet number theory. Section 5 is dedicated on the use of deep learning approaches for automatic crack detection. We first train two defect detection networks on a experimental thermogram database obtained with our experimental bench, then we study the effect of physical parameters on their performance. In order to evaluate the potential of the fusion of both spectra, we finally compare detection scores between infrared and visible spectra for these architectures, on specific unpaired images which contain small cracks or structural noise like due to coatings.

3. Flying-spot: theoretical elements, simulations and experiments specifications

3.1 Theoretical principle of FST

The main idea of FST can be synthesised as a spot laser heat source moving on the surface of the sample, scanning it as illustrated in figure 1. The defect appears as a discontinuity of the heat diffusion on the surface of the material. This heat diffusion variation can be captured using a thermal sensor -like an IR camera. Then thermal information is post-processed and filtered. The whole process is synthesised in figure 2. Reconstructed thermograms are used for this study: they are obtained by making normalizing the mean of each image of the scan movie, corresponding to each position of the spot on the sample. In FST state of the art, the material is scanned back and forth, in order to make the subtraction between the reconstructed thermogram and its return. This pre-processings step cleans the data from structural noise and edges, that produce local gradients that could be detected as a crack. Instead, to reduce the examination duration, we propose to make a single scan of the material. The processing of this data is then more challenging, justifying the interest of high level processing such as neural network.

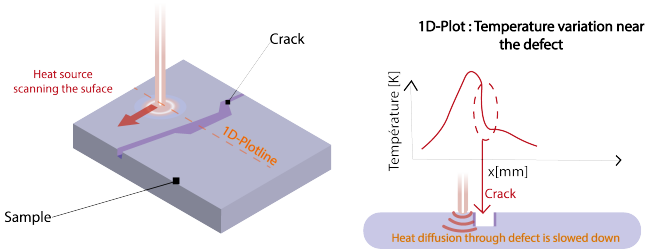

Fig. 1. *Illustration of the principle of Flying-spot thermography.*

Our main hypotheses are that the scan velocity is constant during the examination. The thermal emissivity and diffusivity of the material is supposed quasi-constant too: examinations are made in our case for small temperature variations as studied in MWIR bandwidth, with a small heat source power. The theoretical work of Krapez [2] introduces the Peclet number: this non-dimensional number, corresponds to the ratio between convective heat transfert and thermal diffusion.According to the theoretical elements developed in [2], the best detection -understood as the most important thermal discontinuity due to the crack- corresponds to a Peclet of 1.

$$Pe = \frac{\text{Convective heat flux}}{\text{Heat diffusion}} = \frac{v_{spot} \times R_{spot}}{\alpha}$$

.

with v_{spot} [mm/s]: the velocity of the heat source scanning the surface. R_{spot} [mm]: size of the spot due to heat source and α [mm^2/s]: the thermal diffusivity of the materials.

In order to quantify this discontinuity, Laplacian filters are commonly used for in order to extract edges or cracks, both inducing thermal discontinuities [3] [13]. They give a specific signal at the position of the discontinuity, with a local maximum and a local minimum : we can measure the mean of peak-to-peak magnitude through the length of the defect to quantify the quality of detection for the defects. This approach is synthesized in figure 2, giving us a way to extract structures like edges or cracks.

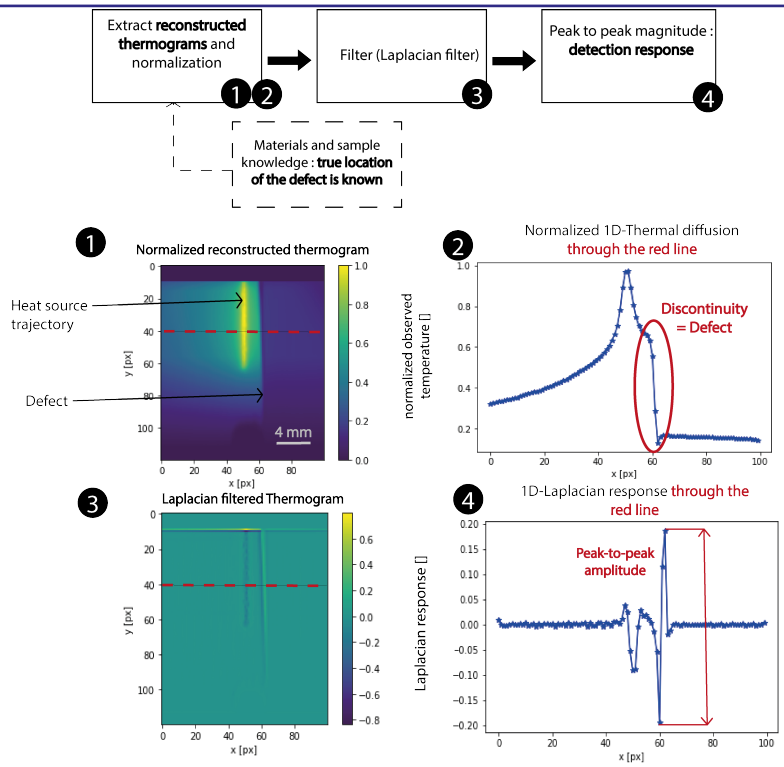

Fig. 2. *Illustration of the quantification of temperature discontinuity due to the presence of a defect.*

3.2 Experimental FST bench

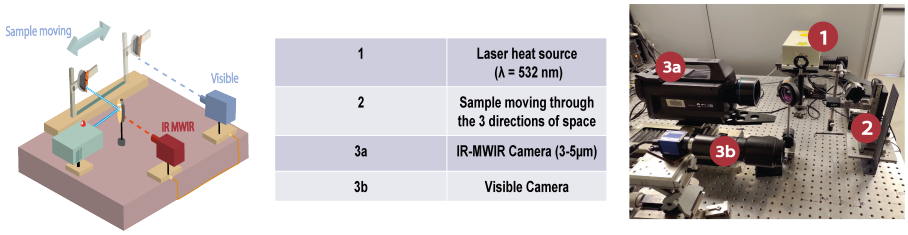

Fig. 3. *Illustration of the FST bench of ONERA.*

The FST bench of ONERA illustrated in figure 3 uses a laser heat source with a power varying from 0.5 to 3 W. The wavelength is 532 nm. A dichroic lens is added to reflect laser spot on the part and to leave IR heat flux arrive to the MW-IR camera, sensitive between 3 and 5 µm. Our samples are superalloy samples coated with a thermal barrier: 3 with cracks, and 3 without crack. We also scan 2 fatigue-test specimens made of steel with a surface crack, giving us more variability in defect opening and defect penetration through materials introducing variations in structures and an heterogeneous thermal diffusion. A last specimen of superalloy has been used for optimization, presenting a side-by-side surface crack. We do planar rotations of this defect to add crack orientation variability in our study.
The laser spot-size is chosen between two values: approximately 0.5 mm and 1.5 mm. The spot covers a distance of 4.5 mm for horizontal scans, 6 mm for vertical scans, close to the defect. Scan speeds varies between 0.5 and 2.5 mm/s. An angular rotation is applied manually to some of our parts to increase the dataset. It varies from 0°(defect is vertical) to 45°. A visible camera (MVBlueFox 124C Cmos camera) has been added to make visible inspections.

3.3 FST simulation framework

There are multiple ways to simulate FST process, and one of the most common in the literature is using FEM models [5] [6]. We use a FEM simulation framework to make predictions about the influence of the distance between the spot and the defect on detection: all our simulations are made using COMSOL simulation software. The main idea of FEM is to use simple elements like triangles to model locally the behavior of a complex physics phenomenon, making it easier to simulate and to study globally. We model a 2D surface. Our defect is defined as a simple line crack with a thermal resistance: this parameter can be chosen by giving a crack opening value. The advection-diffusion equation is used in our simulation, modelling how the heat will infuse in our material and making scan speed and spot size intervene. The boundaries of our model are made on a typical air convection, corresponding to the experiment environment. Thermal materials properties correspond to the properties of a coating material: a thermal barrier. Scan velocity and spot-size are sat to correspond to experiment set-up: speed evolves between 0.5 and 2.5 mm/s, spot size between 0.25 and 1.5 mm.
After building this model, settings of the flying-spot inspection can be tuned, in particular those which have to be chosen by the experimenters, such has scan velocity, spot-size and power. Defect variation can also be introduced such as crack opening size and crack orientation. Our simulations consist in making rectilinear scans parallel to the defect to study the impact of distance on the detection. Besides we studied the response with defect angle variation. The figure 4 gives a synthesis of parameters that can vary in our simulations: in this paper we focus on angle, Peclet parameters and power. Crack width is locked at 0.1 mm: this value is an accommodation corresponding to our samples crack sizes.

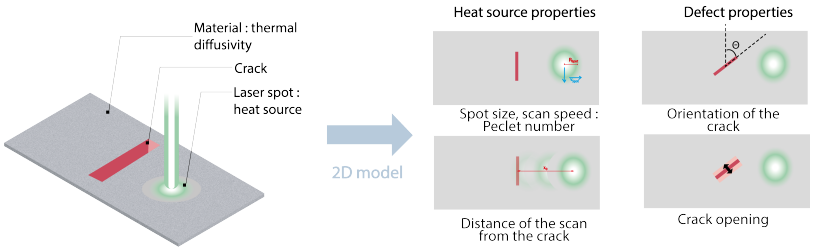

Fig. 4. *A summary of the main parameters that can be tuned using the simulation framework.*

4. Optimization of experiment settings of FST for crack detection

The Peclet number developed in subsection 4.1 gives an optimization parameter easy to use for experiment settings. Theoretical results developed in the bibliography said that best detection happen when the Peclet number tends to 1. The main purpose of this part is verifying if the Peclet number affects also the detectability of a defect through distance between the crack and the heat source. To conduct this verification, we will proceed to a comparison between FEM simulations and experiments with our bench. We will also evaluate the influence of heat source power on detection, as well as the impact of the orientation of the crack with respect to the laser spot.

4.1 Results on simulated data

Using the simulation framework described in section 3, we generate a set of simulated thermograms having variations of laser power, radius, velocity and distance to the crack, as well as angle between the spot trajectory and the crack. Figure 5 gives an example of thermal images obtained. To study the crack impact on thermal response we use an horizontal 1D curve T = f(x) orthogonal with the defect and passing through the center of the defect. We firstly focus on the influence of Peclet parameters, then studying influence of power and finally angle. As predicted we have a discontinuity on the Gaussian thermal response. In order to estimate the amplitude of this discontinuity we use the Laplacian filtering process described in figure 2, section 3.
Figure 6 gives the evolution of detection amplitude with the distance of the spot from the crack for different Peclet values, obtained by varying scan velocity. The best amplitude variation is found for the Peclet number closest to 1, as predicted by the theory.

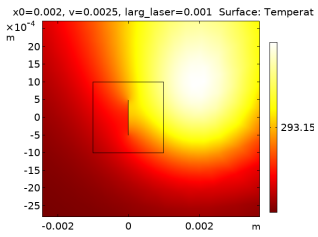

Fig. 5. *Example of simulated thermal image. The defect is the black line at the center of the image.*

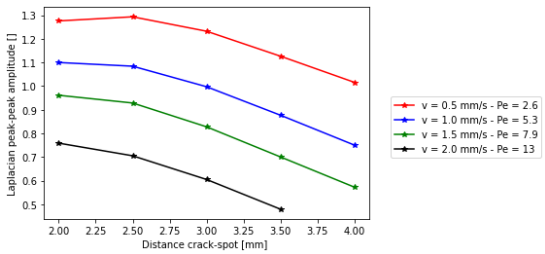

Fig. 6. *Simulation results: evolution of Laplacian peak-peak response with distance for a large spot size (1.5mm), with associated Peclet number.*

Impact of heat source power on the thermal discontinuity is explored in figure 7, describing the evolution of amplitude with power. Simulations are obtained with optimal parameters corresponding to the Peclet number closest to 1 and the measure is made for a spot at 1 mm from the crack. The influence of this parameter can be important, depending on the material studied and IR bandwidth exploited. This parameter can be seen as an illuminance parameter, enlarging the height of the gap. We can observe a linear correlation between power and the peak-peak amplitude due to the crack. Maximizing power during experiments can be beneficial for crack detection.

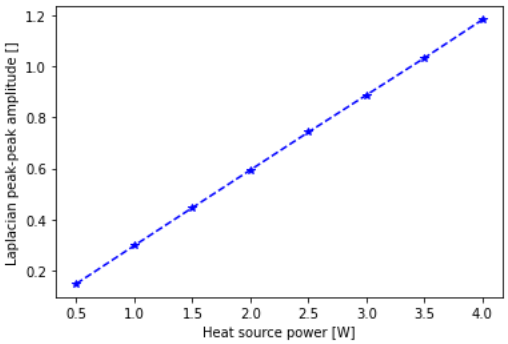

Fig. 7. *Evolution of the Laplacian response with power: the evolution seems to be linear.*

We also study the influence of crack orientation on the thermal response, illustrated with figure 8 which presents simulated heat flux and 1D thermal response associated for two extreme angles (>45 degrees), that we can find unusual at first sight. For angular rotations greater than 45 degrees the gap of the thermal response is reversed. The main explanation can be associated with 2D heat diffusion and its properties: the thermal flux arrives on the other side of the crack before passing through it. To go further we simulate also a much longer defect.

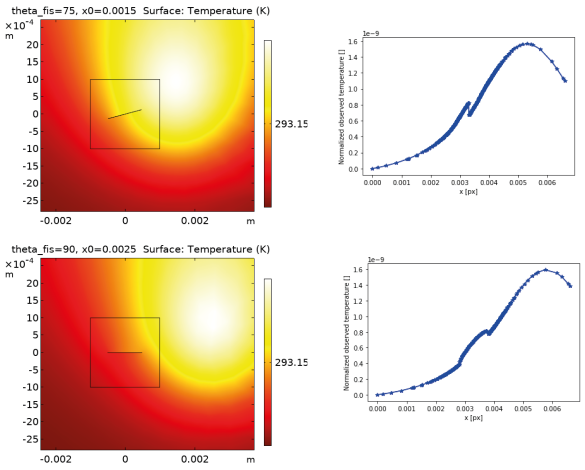

Fig. 8. *The cases of unexpected thermal diffusion illustrated. Left: thermograms. Right: 1D-temperature plot through the defect. The gap is abnormal, with higher thermal value from the side of the crack far from the laser than closer to the heat source.*

Figure 9 illustrates a simulation with an angle of 45 degrees and a much larger defect, with thermal discontinuities associated. Simulation illustrates that our hypothesis is relevant: gap is not affected by the angle when the defect is much longer than scan length. For this kind of defect orientation is not a problem. However for a local and small defect a first tool to avoid detection issues may be multiple scans with various orientations. It can also be a point promoting flying-line techniques. It allows us to avoid this issue.

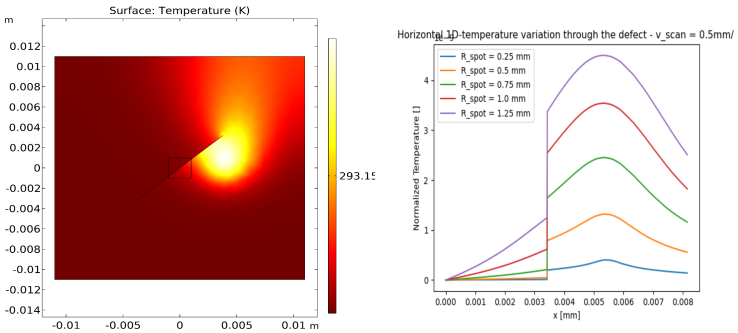

Fig. 9. *Left : random thermal response for a side-by-side defect / Right : 1D horizontal plot through the crack. The discontinuity induced by the crack on thermal diffusion is as expected by theory [2].*

4.2 Experimental results

The curves plotted in figures 10 and 11 describe for a large spot-size (left graph) and for a small spot-size (right graph) the peak-to-peak amplitude evolution with distance between defect and heat source. It highlights that the best peak-peak amplitude is found for Peclets close to 1 in each case: Peclet number theory looks to be a relevant and simple way to maximize the defects peak-to-peak amplitude. However other elements have to be taken into account like sensor resolution and sample properties. As figure 11 shows for these experiments, critical distance to perform peak-to-peak amplitude measure is smaller than for a large spot-size: The critical distance of measurement looks to also be correlated with heat flux entering the sample during the scan, which is affected by exposure time, therefore by scan duration.

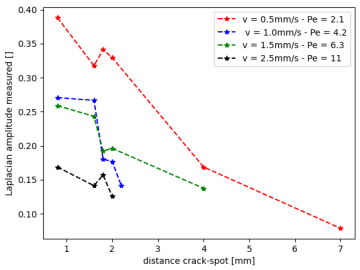

Fig. 10. *Evolution of the Laplacian peak-to-peak amplitude with distance (large spot size: R-spot = 1.3mm, Pe>1)*

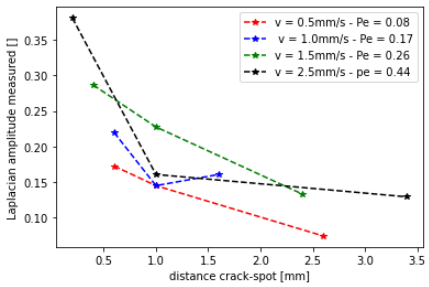

Fig. 11. *Evolution of the Laplacian peak-to-peak amplitude with distance (small spot-size: R-spot = 0.3mm, Pe<1).*

As for simulations we conduct experiments to evaluate the influence of power. Our experiment results are plotted in figure 13 giving the evolution of peak-to-peak amplitude with power. An example of thermal image exploited for this analysis is given in figure 12 . The trend of these results agrees with simulations on the general idea that increasing power seems to increase the discontinuity due to the defect. However the choice of power needs to be made carefully, thinking about materials properties, working bandwidth. Maximizing its value can be beneficial to improve easily defect detection. The major difference between simulation and experiment is that the evolution of power isn't linear and seems to tend to a maximum: the main hypothesis to explain this phenomenon may be on the sensor side, camera thermal saturation, that isn't present in simulations.

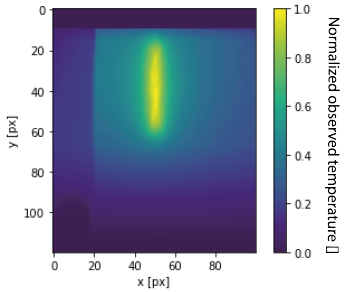

Fig. 12. *A thermogram used to evaluate the impact of power: here we have chosen optimal Peclet parameters, with scan-velocity of 0.5 mm/s and large spot-size(1.3mm).*

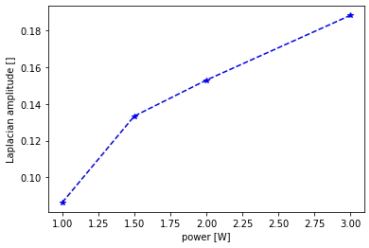

Fig. 13. *Experiment results: impact of power on Laplacian peak-to-peak amplitude. Amplitude is measured for a spot at 6.2 mm from the crack. We see an increase of the response following the increase of power.*

Angle variation has also been studied: our thermal images doesn't present reverse or flat gap phenomena. Indeed for experiments scan length is much smaller than defect length (>5mm). However if scanning smaller cracks the idea coming to prevent from unexpected gap may be making at least 2 scans with very different orientations (vertical or horizontal scans for example) but we need further works to test it.

4.3 Optimization of bench settings: conclusions

In conclusion Peclet number gives a simple and relevant way to set-up effective experiments. Depending on the bench components and sample properties, choosing scan velocity and spot-size that gives a Peclet closest to 1 allows to have the best peak-to-peak amplitude. Power if chosen wisely is also a big parameter to increase this response.
The amount of experiments made gives also datasets to feed-up deep-learning algorithms with various settings: the purpose of the next part will be to study settings influence on a neural network and to compare detection performance in IR and visible spectra, in order to see how both spectra can be complementary.

5. Towards deep learning IR and visible fusion for crack detection

Artificial intelligence (AI) and neural networks are a big improvement in computer vision tasks, due to their ability to learn to classify images from data, and to their ability to extract contextual information from image features like part edges or contrast variations due to heat source. Potential of AI in NDT is great, giving a way to make flexible and refined processing of examination data.

As mentioned in section 3, to reduce inspection duration, we propose to use a FST with a single scan, hence no background subtraction between the reconstructed thermogram and its return can be conducted - processing step usually used to clean structural noise and edges. To deal with those more challenging data, we propose to use neural network that will extract context information for the crack detection like edges. However deep learning performance can be dependant to large amounts of labeled data for training, implying the issue of how to obtain enough images from examination of a small number of sample. In this section, we first evaluate the ability of a neural network to detect crack within a patch of a single pass thermogram. To do so we study the response of different neural networks trained on thermal images in the context of binary classification task between defect-images and without defect images. Then we study the sensitivity of the network performance with respect to settings parameter, once again in comparison with the Peclet number theory. A last part is dedicated to a comparison of the response of these different architectures for visible spectrum, in order to compare it with IR results, as an opening for IR-Visible fusion.

5.1 Network architecture for image patch classification

In this paper two architectures for image classification have been used. The first neural network is a simple convolutional neural network (CNN), made of a few layers. The core principle of this architecture is to use neurons like convolution kernels scanning the input image, stacked in order to learn to perform some tasks by extracting specific features from the image. Each convolutionnal layer learns to extract features -patterns like edges- by filtering its input. The major inspiration for this network is VGG network [14].
A vision transformer (ViT) for classification has also been implemented [15]. The main idea behind transformers is to cut input images in small patches. Then the attention mechanism is used to calculate an output matrix evaluating the relative weight of each patch: Due to this attention, the network can learn to be reactive to the most salient patches. Multiple attention layers can be stacked with a final classification layers. This kind of architecture is relatively starving for data, therefore our network architecture has been sat following papers on adapting visual transformers for small datasets [16]. This quite new neural architecture is seen as an alternative for classical CNN.

5.2 Impact of FST settings on the neural network performance

5.2.1 Dataset and training settings

Using the experiment data acquired with our bench we build a first dataset of 314 IR images without defect, and 332 thermal images having a surface crack, from various superalloy parts (3 with real defect). This training step has been made without considerations about experiment optimization at first sight. Data augmentation is made of mirror-flips and orientation changes of our images (brightness or noise alteration is avoided in order to not to change physics-influenced features of the image). Our training runs during 200 epochs. The optimizer chosen is ADAM with a start-learning rate of 1e-4. The cost function used here is the cross-entropy loss. The batch-size is 4.
Evolution of test-Accuracy (the ratio between good detection and the number of images) gives a first way to monitor training of the network at each step. Test-accuracy is measured at each step on a global test-set.

5.2.2 Results

The evolution of accuracy and performance scores of this model are presented in table 1. Note that the recall score measures the ability of the classifier to find all the samples with a defect, whereas precision gives its capability not to label as positive patches that are not cracked. F1 combines both in a synthetic score. Value of 1 corresponds to the perfect score. Figure 14 monitors accuracy on our test-set at each training step. The network appears to converge during training.

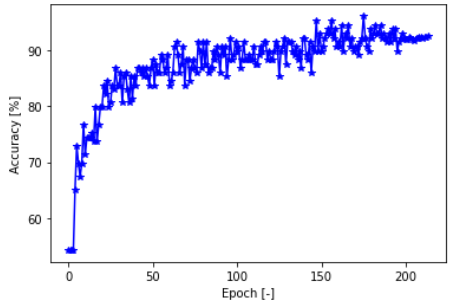

Fig. 14. *Evolution of test-accuracy during training.*

Score	value
Test-accuracy [%]	91.47
F1 score [0-1]	0.922
Recall	0.929
Precision	0.915

Table 1. *Score evaluated on the general test-set after training for IR spectrum.*

The final value of test-accuracy (91 %) and other performances scores are quite satisfying. Figure 15 gives some misclassified thermograms: two false alarms and two missed detections. They may induce first hypotheses on the behaviour of the network: thermograms 1 and 2 may evoke the influence of distance of the crack from the spot as a source of error. For false alarms illuminance of the image, highly accurate with power, may be a relevant hypothesis (patch 3 and 4). Some errors seems to be due to image artifacts like reflections of the laser spot on part structures, as highlighted on patch 4: the sample presents an heterogeneous coating surface that can be the cause of the false detection.

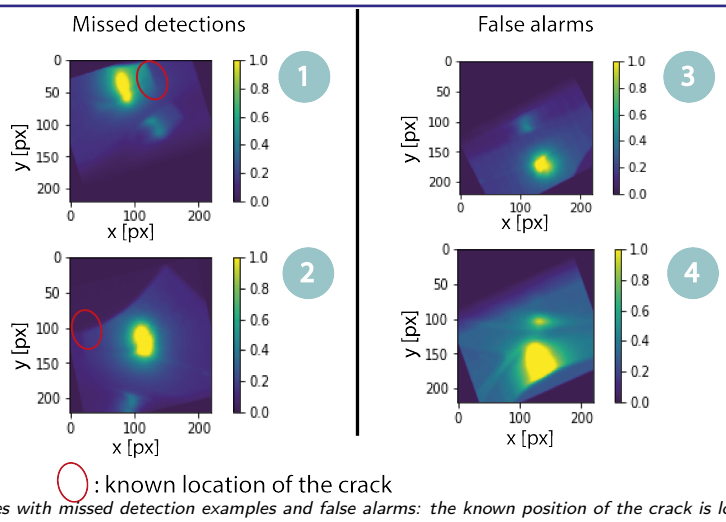

Fig. 15. *Thermal images with missed detection examples and false alarms: the known position of the crack is located into the red circle.*

5.2.3 Test of the influence of experiments settings on detection performance

After network training on the whole experiments dataset, we study here the impact of some experiment settings that we isolated by studying test-set errors. We build new test-sets sorted with our different parameters, two made with each spot-size in order to study easily the extreme values of Peclet (scan velocity fixed at 0.5 mm/s). There are other test sets made to study the influence of power on detection. All our s are built with thermograms with a defect. The main idea is to estimate if and how important is the parameters influence the number of missed detection by a trained neural network: accuracy is used here to evaluate performance. It is also important to notice that these test-set are built with a mixture of new thermograms and thermal images from the training.
To test the influence of spot power, 3 different test-sets are made: table 2 synthesizes evolution of accuracy on these test-sets with different powers and size of each set. At first sight we can see that these results are coherent with optimization results. Maximize the value of power is an easy way to improve the response of the network (while considering sample properties, and sensor specifications to avoid saturation). Scan duration may be another way to improve the measurement, not explored yet: non-detection cases for the sets studied in this subsection seem to appear for the fastest scans, and scan duration affects heat flux sent to the sample and "illumination" of the thermal image.
We evaluate Peclet impact through spot-size variation with speed fixed. Table 3 synthesizes accuracy obtained for a small spot-size (degraded Peclet) and for a large spot-size (Peclet closer to 1). It indicates a correlation between Peclet number and test-performance of the neural network.

Power[W]	0.5	1	1.5
Number of image per test-set	34	43	17
Accuracy []	0.76	0.83	0.86

Table 2. *Accuracy obtained on our test-sets with various power.*

Spot-size [mm]	1.3 (large)	0.3 (small)
Number of image per test-set	50	27
Accuracy []	1.0	0.82

Table 3. *Accuracy obtained on our test-sets with our two different spot-size (scan velocity at 0.5 mm/s).*

To conclude this subsection about neural networks behavior regarding experiment parameters, the main idea highlighted here is that bench optimization using Peclet number can improve neural network detection. Power is another way to easily improve detection performance. It can also be relevant to evoke the idea that a deeper neural network, if fed with enough data, may be less dependant to experiment settings.
Table 4 gives performance on a smaller optimised set for ViT and CNN architectures: this dataset contains 166 cracked thermal images and 166 uncracked thermograms obtained with optimized experiment settings. When training with this smaller dataset of optimal settings thermograms, we found little bit better accuracy. However during training accuracy was less stable, and precision is degraded: it needs to be explored deeper in further work but the major explanation can be the lack of data. The two architectures gives also close scores in IR, evoking that structural noise which is inherent from our dataset -thinking about edges or coating on our samples- disturbs learning for both of our simple architectures.

Architecture	CNN	ViT
Final test-accuracy	91.53 %	91.59 %
F1-score	0.90	0.906
Recall	0.949	0.954
Precision	0.852	0.863

Table 4. *Synthesis table giving score obtained with each neural networks trained on IR-optimized dataset.*

5.3 Towards a fusion between IR and visible spectrum on defect detection

The number of false alarm induced by structures on the surface of the sample like coatings can be a is a problem for defect detection in IR spectrum. However these elements can be easily detected using visible spectrum. The main purpose here is to make a global comparison between IR and visible spectrum for the different neural architectures. This is a preliminary work for IR/visible fusion, through identifying benefits and drawback of each spectrum, studying their classification errors. For visible spectrum we use the camera of our bench to generate a dataset of visible images. We go from high resolution large scale images to patches of 227px*227px manually sampled between cracked or not cracked images. This process gives us a dataset of 6000 visible images.
Training settings are same as for section 5.2. For both spectra we added some data augmentation: this process gives an easy way to increase the amount of available images. In order to have only physically relevant images, there are only augmentations on image orientation (horizontal and vertical flip, rotations).
Table 5 presents our main results after training in visible spectrum (refer to last section for scores in IR). These scores are associated with the evolution of test-accuracy during training 16

Architecture	CNN	ViT
Final test-accuracy	96.37 %	98.35 %
F1-score	0.964	0.979
Recall	0.977	0.980
Precision	0.951	0.977

Table 5. *Synthesis table giving score obtained with each neural networks trained in visible spectrum.*

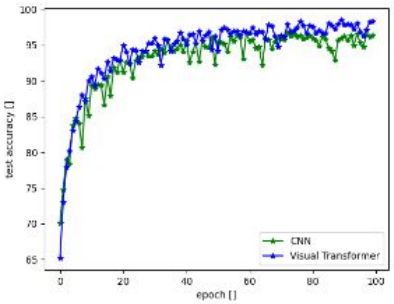

Fig. 16. *Evolution of test-accuracy during training in visible spectrum for both architectures.*

Detection scores are quite high in visible spectrum, as in IR. The difference of performance between architectures: the ViT gives better results and the difference between precision and recall is reduced. There are less false alarms than for CNN. A first explanation can be the complexity of ViT, giving them the ability to infer better than CNN. The attention process can also explain this difference of performances: The network may be able to focus on the most relevant patches of the image in order to classify better between cracked or not-cracked.
As seen in section 5.2 in IR results are correct (> 90 %) but less important than for visible: the first major reason explaining this difference is dataset size. Figure 17 gives examples of misclassifications in both spectra. There are false alarms on our thermal dataset due to the structural elements (in our case surface coating) disturbing performance. Boundary with thermal barrier might be seen as a defect by our networks: it highlights the interest of fusion with visible in order to extract easily more information about surface studied, to avoid several false alarms.
In visible spectrum, as illustrated with visible examples figure 17, for missed detection size of the defect and its location on the patch may be the major reason for non-detection. False alarms are more difficult to interpret: the example given may have an artifact near its left corner that can be classified as a defect. Smaller and thinner defects are also harder to detect: IR can help a future fusion network architecture to detect this kind of crack, harder to distinguish in visible spectrum. detection errors due to scratches or artifacts can also be avoided by adding IR inspection data.

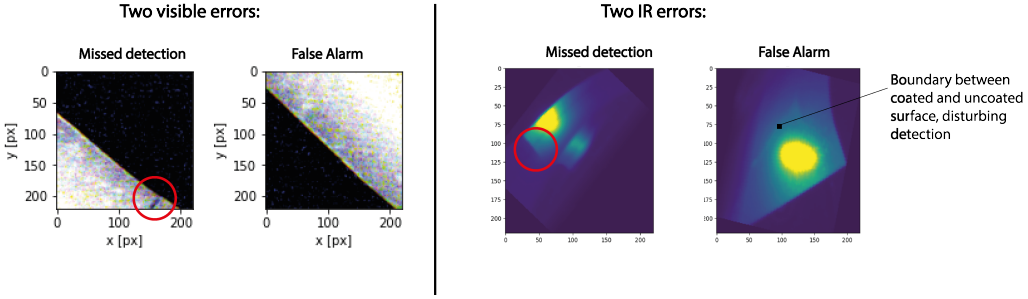

Fig. 17. *Examples of errors of classification obtained for each spectrum. For missed detections the defect is circled in red.*

To conclude this section, good detection results have been obtained with neural network associated with FST and in visible spectrum: the choice of experiment settings seems to have an impact on detection, even if deeper architectures may have enhanced robustness to degraded Peclet values. Fusion between visible and IR can also be complementary, visible allowing to identify edges and coating in order to reduce false alarms with FST while IR can help for defects hard to detect in visible. IR may also help in avoiding false alarms due to scratches. This fusion will be explored in further work with more complex neural networks.

6. Conclusion

In this paper we presented our work towards the fusion of IR Flying Spot and visible inspection. This work was focused on IR bench optimization to validate Peclet number theory on experiments and simulated data. Defect detection and how neural networks are influenced by experiment parameters, finding-out some core-principle to set-up efficient deep learning computer vision in this spectrum. Using an experimental bench having both modalities, two datasets have been built and we trained defect detection neural network, and analysis of performance for each spectrum has been conducted.
Detection results obtained using this preliminary work illustrate that neural networks associated with FST are able to perform automatic defect detection. However noise from coatings or edges disturbs performance: Visible spectrum can give information on surface properties in order to avoid these false alarms. In further work we will continue to explore both spectra with the idea of generating paired thermal and visible images. We are also working on state-of-the-art more complex neural network architectures for object detection, that would for instance be able to differentiate different regions like coated or uncoated surfaces, structural elements on the data. Architectures fusing IR and visible data will be developed to find benefit from information given by both modalities.

References

[1] Edward J. Kubiak. Infrared Detection of Fatigue Cracks and Other Near-Surface Defects. Applied Optics, 7(9):1743–1747, September 1968. Publisher: Optical Society of America.

[2] Jean-Claude Krapez. Résolution spatiale de la camera thermique source volante. International Journal of Thermal Sciences, 1999.

[3] Thierry Maffren. Détection et caractérisation de fissures dans des aubes de turbine monocristallines pour l'évaluation de leurs durées de vie résiduelles. Phd thesis, Paris, CNAM, April 2013.

[4] Y. Mokhtari et al. Comparative study of Line Scan and Flying Line Active IR Thermography operated with a 6-axis robot. In Proceedings of the 2018 International Conference on Quantitative InfraRed Thermography. QIRT Council, 2018.

[5] T Li et al. Crack imaging by pulsed laser spot thermography. Journal of Physics: Conference Series, 214:012072, March 2010.

[6] T Li et al. Crack imaging by scanning laser-line thermography and laser-spot thermography. Measurement Science and Technology, 22(3):035701, March 2011.

[7] Agustín Salazar et al. Flying spot thermography: Quantitative assessment of thermal diffusivity and crack width. Journal of Applied Physics, 127(13):131101, April 2020. Publisher: American Institute of Physics.

[8] Wenxiong Shi et al. A technique combining laser spot thermography and neural network for surface crack detection in laser engineered net shaping. Optics and Lasers in Engineering, 138:106431, March 2021.

[9] Ç.F Özgenel et al. Performance Comparison of Pretrained Convolutional Neural Networks on Crack Detection in Buildings. ISARC Proceedings, pages 693–700, July 2018. Publisher: IAARC.

[10] Amar El-Maadi et al. Visible and infrared imagery for surveillance applications: software and hardware considerations. Quantitative InfraRed Thermography Journal, 4(1):25–40, June 2007. Publisher: Taylor & Francis _eprint: https://doi.org/10.3166/qirt.4.25-40.

[11] Chengyang Li et al. Illumination-aware Faster R-CNN for Robust Multispectral Pedestrian Detection. arXiv:1803.05347 [cs], August 2018. arXiv: 1803.05347.

[12] Hui Li et al. DenseFuse: A Fusion Approach to Infrared and Visible Images. IEEE Transactions on Image Processing, 28(5):2614–2623, May 2019. arXiv: 1804.08361.

[13] Nelson W. Pech-May et al. Detection of Surface Breaking Cracks Using Flying Line Laser Thermography: A Canny-Based Algorithm. Engineering Proceedings, 8(1):22, 2021. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

[14] Karen Simonyan et al. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556 [cs], April 2015. arXiv: 1409.1556.

[15] Alexey Dosovitskiy et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv:2010.11929 [cs], June 2021. arXiv: 2010.11929.

[16] Ali Hassani et al. Escaping the Big Data Paradigm with Compact Transformers. Technical Report arXiv:2104.05704, arXiv, August 2021. arXiv:2104.05704 [cs] type: article.