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. Deep learning is now a very efficient processing approach, 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 in comparison with the theory of Peclet number, for both simulated and experimental data. Then we present FST automated defect detection results using two different neural network architectures and also study the sensitivity of their performances to the physical theory of Peclet number. Finally, we analyse the defect detection results on both IR and visible spectrum separately using deep learning, 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 a deep learning complex remnant 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 order to conduct data fusion using deep learning, the first step is to build a realistic database using an experimental bench. In this paper, we focus mainly on the FST bench and study the settings parameters that maximise the thermogram 1D signal response to a high frequency filter. We show that the optimal settings regarding this response are consistent with the Peclet number theory. Then we use data extracted from this experimental bench to train two different deep learning architectures for automatic defect detection on IR spectrum and show that their performances are already quite satisfying, second that their performance are also consistent 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 the normalized 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.

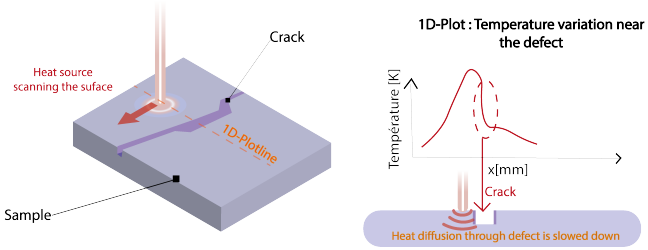


Fig. 1. *Illustration of the principle of Flying-spot thermography*

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 thermography, in order to isolate part 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 isolate structures like edges or cracks.

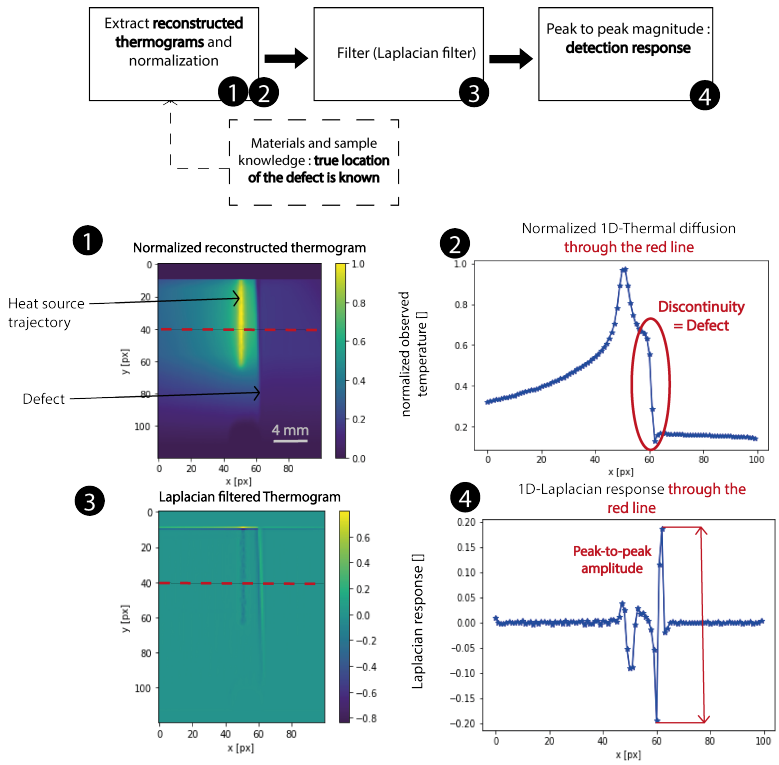
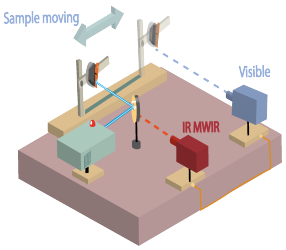


Fig. 2. *Illustration of the quantification of temperature discontinuity due to the presence of a defect*

3.2 Experimental FST bench



1	Laser heat source (λ = 532 nm)
2	Sample moving through the 3 directions of space
3a	IR-MWIR Camera (3-5µm)
3b	Visible Camera

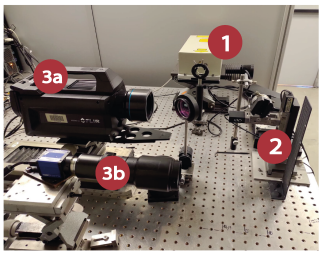


Fig. 3. *Illustration of the FST bench of ONERA*

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. These variations of structural elements like shape, defect geometry, and crack orientation are relevant.
The laser Spot length 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.

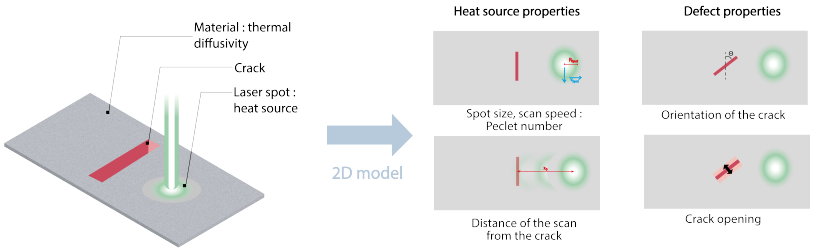


Fig. 4. *A summary of the main parameters that can be tuned using the simulation work*

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: Does the critical detection distance be influenced by the scan parameters, which are considered optimal for a Peclet close to 1 ? We will proceed to a comparison between FEM simulations and experiments with our bench.

4.1 Simulation results

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. To study the crack impact on thermal response we use 1D curve T = f(x) orthogonal with 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.
As predicted and as shown in figure 5 giving the evolution of detection amplitude with the distance of the spot from the crack for different Peclet values, the best amplitude variation is found for the Peclet number closest to 1.

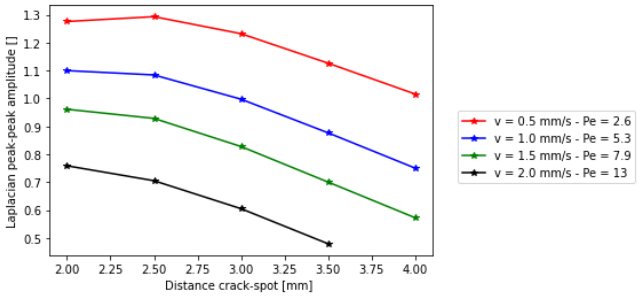


Fig. 5. *Simulation results: evolution of Laplacian peak-peak response for a large spot size (1.5mm), with associated Peclet number*

The impact of heat source power on the thermal discontinuity is explored in the figure 6, 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 seems to have a linear correlation between power and the peak-peak amplitude due to the crack.

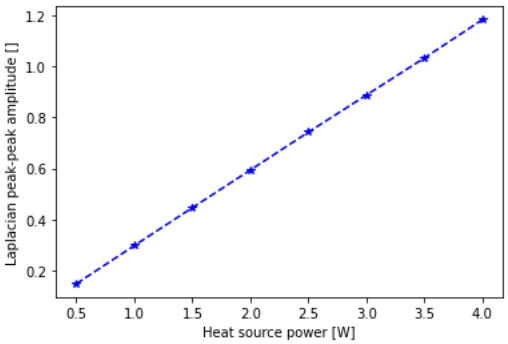


Fig. 6. *Evolution of the Laplacian response with power: the evolution seems to be linear*

The angle variation gives thermal responses, illustrated with figure 7, that we can find unusual at first sight. The figure gives the thermal image and 1D temperature simulated. For angular rotations greater than 45 degrees the gap of the thermal response is reversed. The main explanation is that thermal flux arrives on the other side of the crack before passing through it. The measurement of this kind of response might need further exploration. However generally we can still measure the discontinuity of the Gaussian thermal response to detect the crack.

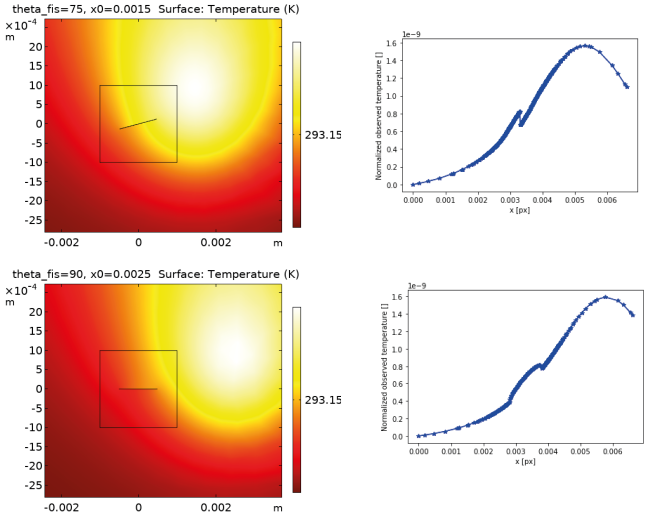


Fig. 7. *The cases of unexpected thermal diffusion illustrated. Left: thermograms. Right: 1D-temperature plot through the defect. The gap is abnormal, with higher thermal value from the side of the crack far from the laser than closer to the heat source*

The figure 8 illustrates a simulation with angle and a large defect, with thermal discontinuities associated.This simulation illustrates that our hypothesis looks relevant: the gap is not affected by the angle when the defect is far longer than before (like a side-by-side crack). 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. In our experiments we proceed to parallel scans. it allows us to avoid this issue. Angle variation added for data augmentation for artificial intelligence in section 5 is limited to an angle between 0 and 45 degrees (0 corresponding to a parallel scan).

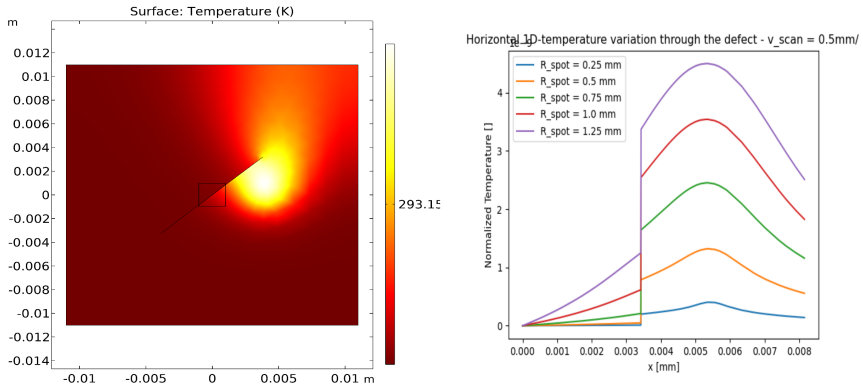


Fig. 8. *Left : random thermal response for a side-by-side defect / Right : 1D horizontal plot through the crack. The discontinuity induced by the crack on thermal diffusion is as expected by theory [2]*

4.2 Experimental results

The curves plotted in figures 9 and 10 for a large spot-size (left graph) and for a small spot-size (right graph) highlight 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 detection of the defect. However other elements have to be taken into account like sensor resolution and sample properties need to be kept in mind. As figure 10 shows for these experiments, temperature discontinuity measured by the Laplacian becomes not detectable closer to the defect 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.

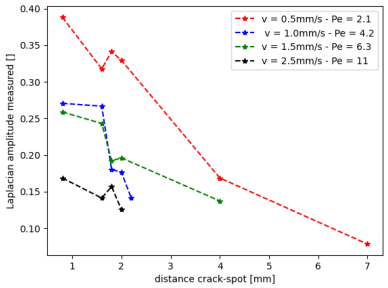


Fig. 9. *Evolution of the Laplacian peak-to-peak amplitude with distance (R-spot = 1.3mm)*

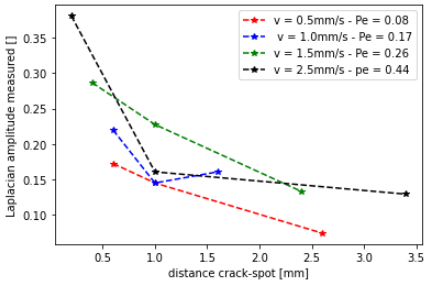


Fig. 10. *Evolution of the Laplacian peak-to-peak amplitude with distance (R-spot = 0.3mm)*

Power is another relevant parameter that influence the defect detection, and can be seen as a "thermal brightness" parameter. Our experiment results are plotted in figure 12 with an example of thermal image associated 11. 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.

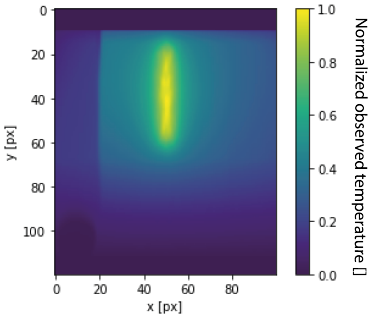


Fig. 11. *A thermogram used to evaluate the impact of power: here we have chosen optimal Peclet parameters, with scan-velocity of 0.5 mm/s and large spot-size*

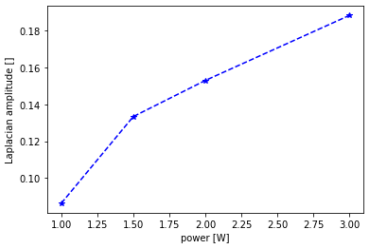


Fig. 12. *Experiment results: impact of power on Laplacian peak-to-peak amplitude. Amplitude is measured for a spot at 6.2 mm from the crack. We see an increase of the response following the increase of power*

The phenomenon of reversed discontinuity, or flat gaps, highlighted in simulations for high rotation values for the defect, is not present on real data from optimization samples. Indeed heat flux cannot pass through the defect by each side of the crack (the sample has a "from side to side" crack). The first idea coming to prevent from both phenomena 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.
These optimization results and thermograms made give us a dataset containing plenty of thermal images to train neural architectures, in order to study how a machine-learning/deep-learning system is affected by various sets of parameters.

4.3 Optimization of bench settings: conclusions

To conclude this part about optimization. The 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 make the best defect detection. Power if chosen wisely is also a big parameter to improve experiments. The amount of experiments made gives 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 spectrum can be complementary.

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

Artificial intelligence and neural networks are a big improvement in computer vision tasks, due to their ability to learn to classify images from data, due 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 XXX, 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. We will discuss this issue in this section.
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

The first neural network, summed-up in the figure 13, 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 next figure gives the general structure of our CNN. The major inspiration for this network is VGG network [14].

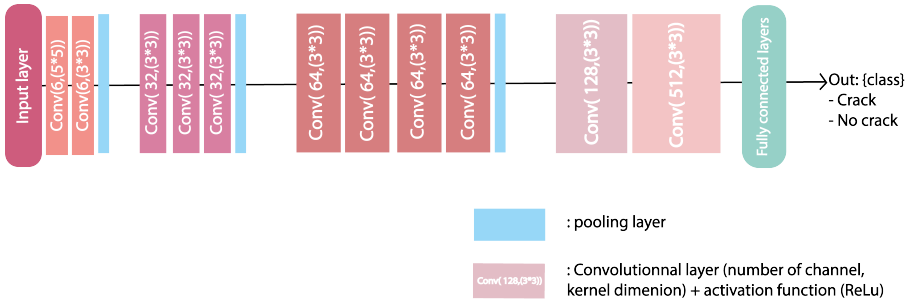


Fig. 13. *General structure of the developed CNN: each convolutionnal layer is described with its number of channel and kernel size*

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.

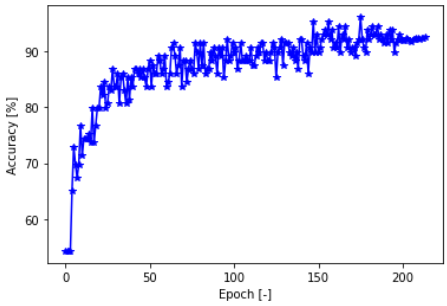
5.2 Impact of FST settings on the neural network performance

5.2.1 Dataset

The study of this subsection is focused on CNN training. 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 is made of 200 number of epoch depending of the architectures and when they converges. The optimizer -this element will drive how the different weights of our networks varies between the different epoch- chosen is ADAM with a start-learning rate of 1e-4. The cost function used here is the cross-entropy loss: this is a common choice for classification tasks. The batch-size (the number of images given simultaneously to the network during training) is 4. Then specific test sets with various couple of experiments parameters (scan velocity, distance to the defect, power ...): the main idea is to see if the number of effective detections will be physically accurate. Test-Accuracy (the ratio between good detections 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 testset built without physics consideration.

5.2.2 Training

The evolution of and general scores of this model are summed-up in the table 1 while figure 14 monitors results at each training step (called epoch). We can see that the network appears to converge 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*

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

The final value of test-accuracy (around 91 %) can indicate that non-optimal experiments may alter our dataset enough to impact learning performance. The figure 15 giving some thermograms 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. 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.

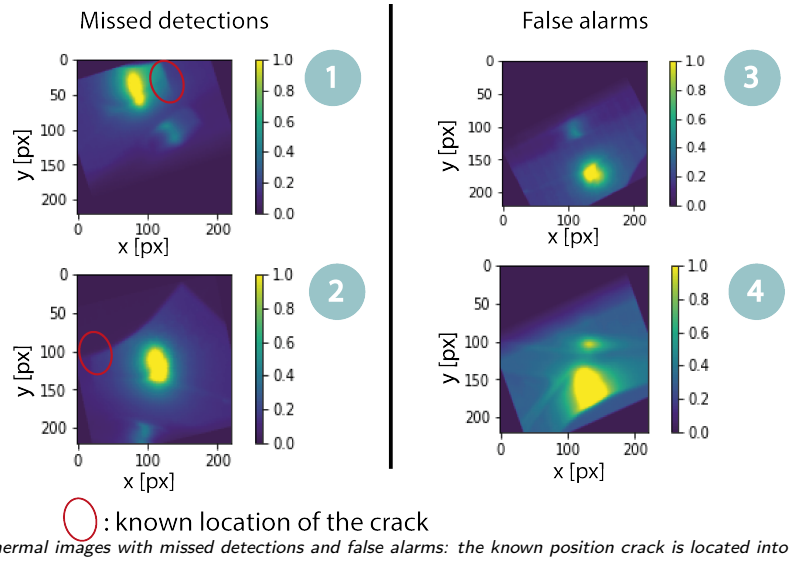


Fig. 15. *Thermal images with missed detections and false alarms: the known position crack is located into the red circle*

5.2.3 Test of the influence of experiments settings on detection performance

After network training on the whole variation of parameters, the study focus on the impact of some parameters 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 set 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: for 0.5W it contains 34 thermograms, for 1.0W it's 43 thermograms, and 17 for 1.5W. The test-accuracy obtained are respectively 0.76, 0.83 and 0.88. 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. The first set containing 50 large-spot size images: it gives an accuracy of 1.0. For the small spot-size test set containing 27 thermal images, the accuracy is around 0.82. It indicates a correlation between Peclet number and performance of the neural network.

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.

5.3 Comparison 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 probleme 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, as a preliminary work for IR/visible fusion. For IR spectrum thermograms generated on superalloy parts selected with considerations to optimization experiments are used: the previous section gives an optimized dataset giving 166 cracked thermal images and 166 uncracked thermograms.
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.
For our training we decided to use Cross-entropy loss, with ADAM as optimizer. The learning-rate is at 1.0e-4. Batch-size in visible spectrum is 16, and decreased to 4 for the training of the CNN in IR (due to the lack of images in this case). It is decreased to 12 for the transformer in IR, because it finally gives the best results on the smaller IR dataset. For both spectrum 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).
The table 2 sum-up our main results after training. 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. For each 1 is equal to the perfect score. As seen in previous section, test-accuracy is the ratio between good classifications and the total number of images, evaluated on the test images at each step.

Architecture	Convolutional neural network		Visual Transformer	
Spectrum	visible	Infrared	Visible	Infrared
Final testing accuracy	96.37 %	91.53 %	98.35 %	91.59 %
F1-score	0.964	0.90	0.979	0.906
Recall	0.977	0.949	0.980	0.954
Precision	0.951	0.852	0.977	0.863

Table 2. *Synthesis table giving score obtained with each neural networks trained on both spectra*

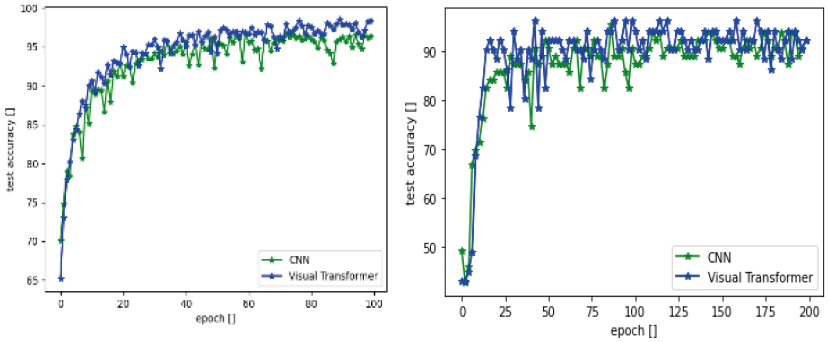


Fig. 16. *Evolution of test accuracy during training for each spectrum: the larger amount of data in visible explains why accuracy is more stable in this spectrum than in IR*

At first sight we see that both networks gives interesting results in each spectrum: metrics converge to stable values, near the top-values. A first comparison between spectrum: we can guess that the difference of accuracy and F1-score is due to the size of each dataset: there is also fluctuations of final scores in IR of 1 to 5 %, too due to the limited number of images in IR. The next point of interest is the difference of performance between architectures: the visual transformer gives better results and the difference between precision and recall is reduced: there is 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. In IR the results are correct (> 90 %) but less important than for Visible: the first major reason explaining this difference is dataset size. On the other hand false alarms due to the structural elements (in our case surface coating) continue to disturb performance (as shown in figure 19 with the example of thermal false alarm: the thermal barrier might be seen as a defect). Both neural networks have relatively unstable test-accuracy during training, presenting some accuracy peaks around 96%: adding more optimized data may be a way to avoid this issue while increasing performance. It also highlights the importance of the data amount for an effective training: the comparison between the evolution of accuracy figure 16 and 14 shows that the more data you gives to the network, the more stable neural network you have. The two architectures gives also close scores in IR, evoking that structural noise which is inherent from our dataset -thinking about structural noise due to edges or coatings on our samples- disturbs learning for both of our simple architectures. 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.

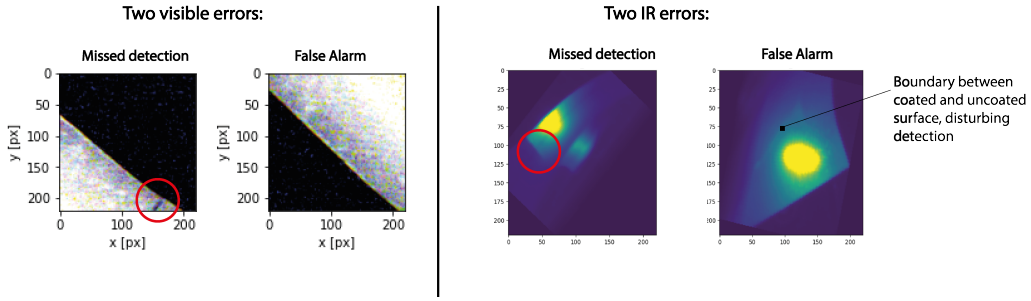


Fig. 17. *examples of errors of classification obtained for each spectrum. For missed detections the defect is circled in red*

To conclude this section, neural networks are promising for defect detection with FST: the choice of experiment settings seems to have an impact on detection. However with larger amount of data a deeper neural network may be able to also perform detection with a degraded Peclet value. Fusion between visible and IR can also be complementary, visible allowing to identify edges and coating in order to reduce false alarms with FST: 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 study 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, comparing performance for each spectrum.
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.