

Detecting Defects in Materials using Deep Convolutional Neural Networks

Q. Boyadjian¹[0000–0001–8824–5042], N. Vanderesse¹[0000–0001–8698–8707], M. Toews¹[0000–0002–7567–4283], and P. Bocher¹[0000–0002–3698–7025]

¹ École de technologie supérieure, 1100, rue Notre-Dame Ouest, Montreal, QC H3C 1K3, Canada

² `quentin.boyadjian.1@ens.etsmtl.ca`

Abstract. This paper proposes representing and detecting manufacturing defects at the micrometer scale using deep convolutional neural networks. The information theoretic notion of entropy is used to quantify the information gain or mutual information of filters throughout the network, where the deepest network layers are generally shown to exhibit the highest mutual information between filter responses and defects, and thus serve as the most discriminative features. Quantitative detection experiments based on the AlexNet architecture investigate a variety of design parameters pertaining data preprocessing and network architecture, where the optimal architectures achieve an average accuracy of 98.54%. CNNs are relatively easy to perform and give impressive achievements in classification tasks. However, the informational complexity coming from the depth of networks represents a limit to improve their capabilities.

Keywords: Convolutional Neural Network · AlexNet · Ti-6Al-4V · Texture classification.

1 Introduction

The spatial organization of materials at the microscopic scale is currently referred to as *microstructure*. The control of microstructure is of outmost importance for a wide array of engineering applications where materials play a structural or a functional role. The resistance of most metals to an externally applied deformation, for example, is tightly linked to the average size of microstructural entities, called grains, by the well established Hall-Petch relationship [1, 2]. Other macroscopic mechanical, thermal, or electrical... properties of a material depend on many microstructural characteristics at scales ranging from the nanometer to the millimeter. The analysis and quantification of microstructures rely on many complementary experimental techniques that make use of different probing signals: optical light, laser, X-rays, electron beam, ultrasonic waves, to name a few. The output signals are usually collected to form a two-or three-dimensional image of a portion of the specimen, documenting its local properties such as chemical composition, crystallographic orientation, density, deformation, occurrence of defects, etc.

In the field of metallic materials, the typical technique for acquiring images of the surface of a specimen is optical microscopy. It can convey a large amount of information about the material, its history and its expected in-service behavior. It is no surprise, then, that metallography has been holding a close relationship with image analysis over the last decades [3–6]. Most of the quantitative studies in this field are oriented toward the segmentation and statistical quantification of features of interest, especially the grains (i.e. elementary crystalline units of the material) and second-phase particles, inclusions or defects [7–13]. The analysis of the textures found in these images has been a matter of interest for some applications, yet to a lesser degree³. Most efforts in that direction are focused on texture quantification by statistical, structural and transform-based methods [14–18]. The application of convolutional neural networks appears limited, with a few recent exceptions [19–23]. One major limitation to the penetration of modern image analysis techniques in materials science, as discussed in [24], may be the financial and time cost for the production of microstructural images (micrographs), as it requires careful specimen preparation, expertise, and expensive equipment [25]. Analysis procedures must then be adapted to size-limited datasets.

Among metallic materials of industrial importance, Titanium alloys, especially the most common Ti-6Al-4V (Titanium alloyed with 6% Aluminum and 4% Vanadium), are used in aerospace, chemistry and biomedical applications, due to their strength over weight ratio, corrosion resistance, and bio-compatibility [26]. Their microstructures are usually complex and vary in many aspects depending on the parameters of the manufacturing process (temperature, deformation, cooling rate...). This has motivated several studies aimed at assessing them by image analysis [27–30, 15, 31]. Ti-6Al-4V at room temperature presents two phases with different characteristics named α and β [26, 32, 33]. A critical requirement for many Ti-6Al-4V industrial parts is that they should present a bimodal microstructure constituted by globular α grains surrounded by a fine mixture of β and α grains. This is illustrated in Figure 1, where the surface of a specimen is imaged in optical microscopy. The α grains appear with bright shades and the mixture of fine α - β grains with dark shades. The goal of the manufacturing process is to ensure that the α grains are roughly spheroidal (i.e. circular when observed in 2D), homogeneously distributed throughout the volume, and surrounded by a fine layer of α - β mixture.

Difficulties arise when some α grains are not correctly spheroidized by the process. An example is presented in Figure 1(a): an elongated feature extending over ca. 100 μm can be observed at the center of the image. It corresponds to an insufficiently fragmented α grain. Such defects are referred to as α_{GB} (Grain Boundary). They must be detected as they can compromise the mechanical

³ It should be noted that the term *texture* in the context of materials science, almost always refers to crystallographic texture, i.e., the dominant crystallographic orientation of a set of grains, and in some cases to the morphological texture. Consequently, this confusion may not help the diffusion of texture analysis in the materials science and engineering community.

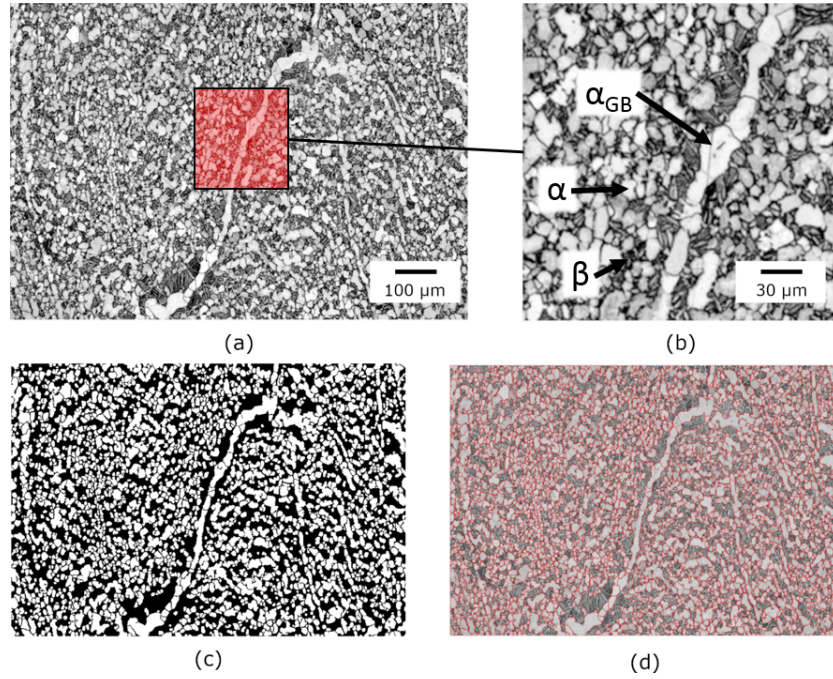


Fig. 1: (a): Micrograph of a Ti-6Al-4V specimen showing a defect; (b): Close-up on α_{GB} defect; (c): Segmented micrograph; (d): Segmented image overlaid on the initial micrograph.

performance of the finished product by initiating cracks under cycling loading, eventually leading to early fracture [34]. In the industrial context, the detection of these features is expert-based, with trained technicians visually evaluating a series of digital images acquired on semi-finished parts and measuring the amount of defect within each, if any. As this is a tedious and time-consuming task, a need exists for automatizing it. It is generally possible to segment the original images without supervision using classical image analysis techniques (Figure 1c), but the subsequent discrimination of the α_{GB} defects proves to be a complicated task, since they can be connected to other grains, and, conversely, be fragmented at some locations. This difficulty motivated the search for a more robust approach, based on the representation of the texture in each image and its classification.

The field of texture analysis has seen significant advances in the last 20 years, with the elaboration of the Bag-of-Words paradigm derived from linguistics for texture representation, and the onset of CNNs, made up of several layers of non-linear feature extractors [35, 36]. The latter demonstrated high performance in building top quality features learned from large datasets, by contrast with hand-crafted approaches based on predefined filter banks. CNNs are usually trained and applied to sets of several hundreds of images (Table 3 in [36], Table 3 in

[37]). By contrast, microstructural pictures of a given material are generally in a much lower number, which possibly puts some limitation on the accuracy of CNNs in this area. However, a recent study in the medical field showed efficient results of AlexNet [38] for a binary classification problem with a dataset consisting of less than 300 images [39]. This result has motivated the choice of AlexNet in this work for the detection of microstructural defects in Ti-6Al-4V specimens with a reduced dataset. As a first approach, the CNN was trained and tested with different hyper-parameters and datasets.

Information theory [40] is a powerful tool for quantifying the statistical link between discriminative network filter responses and classes of interest, i.e. the information bottleneck method studies the mutual information between filter responses [41]. Special attention was paid to the evaluation of the information flow [42] across the layers, so as to gain insights in their role in characterizing defects.

2 Method

2.1 Design of experiments

To the best of the authors' knowledge, the study of defects in optimal microscopy images as proposed in this work is relatively novel in the field of texture analysis. This motivated an approach in which several image parameters and CNN hyper-parameters were allowed to take several values, to define the optimal combination for detecting potential defects in the images. This work proposes a Design of Experiments (DoE) to evaluate the factors suspected to limit the final accuracy of CNN for the classification. The DoE focuses on three parameters referring to the dataset preparation, on the one hand, and three hyper-parameters related to the CNN architecture, on the other hand. They are summed up in Table 1 and exposed in the next sections.

Table 1: Design of experiments.

Parameters	Levels	Unit
Field of view	[293×293 - 586×586]	μm^2
Overlap	[1 - 50]	%
Minimal defect area	[500 - 1500]	μm^2
Kernel size of first layer	[7×7 - 15×15]	pixels
Dropout	[10 - 50]	%
Validation split	[10 - 30]	%

The use of a fractional factorial design of type 2^{6-2} reduced the required number of experiments from 64 to 16 and to keep a resolution of IV. A center point is added to take into account a potential non-linearity in the modeling. For each experiment with a given dataset and a set of hyper-parameters, the

validation accuracy was recorded at each iteration. The value after 50 iterations was considered as the final accuracy of the model, evaluating its overall performance for classifying the images and providing a metric for comparison to the other models.

2.2 Dataset generation

A set of 380 optical micrographs of Ti-6Al-4V specimens was acquired in industrial conditions using different magnifications and cameras. The field of view could vary between $900 \times 630 \mu\text{m}^2$ and $1775 \times 1330 \mu\text{m}^2$. 216 images of this dataset contain a visible defect. Since the field of view, illumination and numerical resolution could vary between the images, they were preprocessed to produce as a "clean" dataset as possible. The open-source program Fiji was used for that purpose [43, 44]. The images were first re-sampled to the same resolution, $1.29 \mu\text{m}/\text{pixel}$, and corrected for uneven illumination with a pseudo-flatfield algorithm. Their intensity histogram was also normalized. By contrast with the original AlexNet architecture [38], which processes RGB images, the images in this work were 8-bit.

The preprocessed images were subdivided into smaller images (tiles) to produce different datasets with various overlap, field of view, and minimal amount of defect present in the image. The overlap ranged from 0 to 50% ("Overlap" in Table 1), allowing for the possibility to increase the data with repeated portions of images. Since the defects can vary in size and shape, the field of view covered by each tile could vary from 293 to $586 \mu\text{m}$ ("Field of view" in Table 1). For supervised learning, the tiles needed to be dispatched in two classes, i.e. "containing a defect" or not. This task was accomplished by segmenting the initial grayscale images (e.g. Figure 1c) with morphological operators (alternate sequential filtering for denoising and convexity controlled watershed for separating touching features [45]) and Otsu binarization. The defects were then manually labeled with a flood fill tool. The labeled images were subdivided with the same parameters as the grayscale images and the area of each defect was measured in each tile. The minimum defect area necessary to consider the tile as containing a defect or not could take values 500 and $1500 \mu\text{m}^2$ ("Minimal defect area" in Table 1).

After cropping, all the tiles had numerical dimensions 227×227 pixels to fit with the AlexNet-CNN recommendations. Some non-defect images were randomly removed to generate a balanced dataset with the same amount of pictures, ca. 400, in both classes. The dataset was augmented by applying horizontal and vertical mirror transformations, as well as 90° rotations.

Figure 2 presents some pictures from a dataset with a field of view of $293 \times 293 \mu\text{m}^2$ and a minimum defect area of $1500 \mu\text{m}^2$. Informally described, the images without defects present similar textures characterized by some degree of granularity, due to the aforementioned process of α phase fragmentation. The overall contrast is constant among the images, as the initial images have been normalized, but the coarseness can vary depending on the amount of α/β phase mixture

(dark areas). By contrast, the images with defects present a discontinuity in texture, materialized by a more or less thin and tortuous shape. Closer inspection reveals that several textures, mostly differing in isotropy, can coexist in the same image (e.g. in the second image without defect starting from the upper left-hand corner). Therefore, an efficient model has to consider the variety of morphology as a unique class.

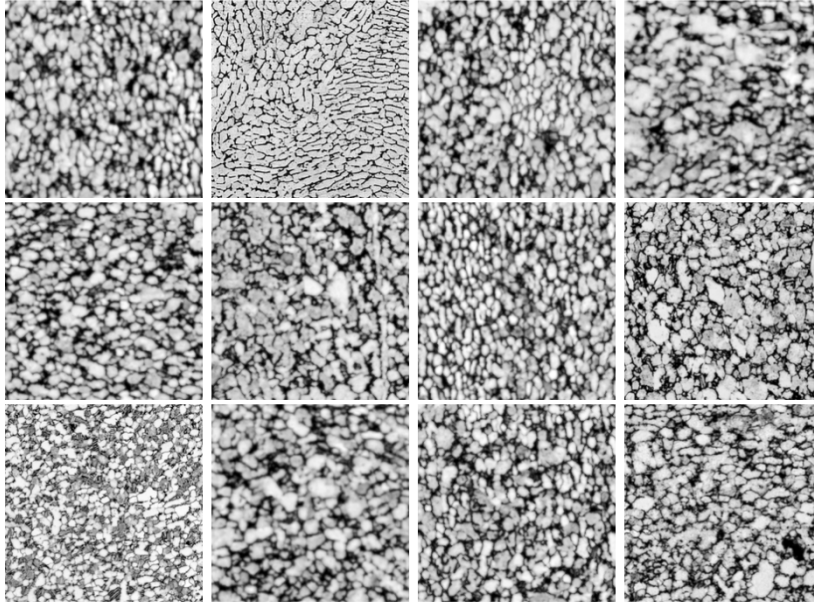
2.3 CNN architecture

The method presented here is based on the AlexNet architecture [38]. The programming language is Python 3.5 extended with the TensorFlow 2.0 machine learning platform [46]. Keras library is also used to define the layer hyper-parameters [47]. The parameters used in this study, very close to the general AlexNet settings, are presented in Table 2. The activation function for the 5 convolutional layers is the Rectification Linear Unit (ReLU). The last three fully connected layers (6, 7, 8) are regularized with a dropout whose value depending on the experiment (Table 1), and the last layer is a softmax binary classifier.

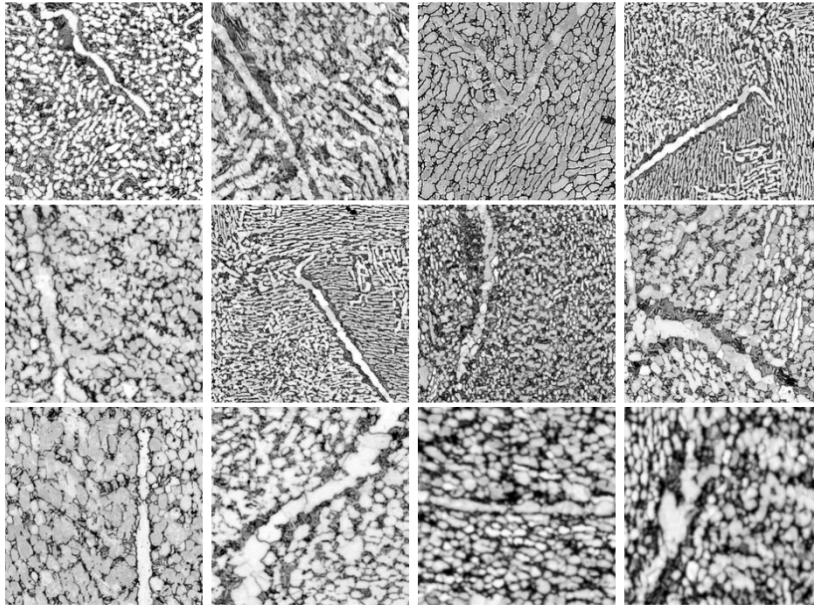
Table 2: Architecture of the studied AlexNet CNN. The parameters allowed to vary are described in the text.

Layer	Operation	Kernel size	Stride	Output dimension
0	Input			$227 \times 227 \times 1$
1	Convolution	[7,11,15]	4	$55 \times 55 \times 96$
	Max pooling	3×3	2	$27 \times 27 \times 96$
2	Convolution	5×5	1	$27 \times 27 \times 256$
	Max pooling	3×3	2	$13 \times 13 \times 256$
3	Convolution	3×3	1	$13 \times 13 \times 384$
4	Convolution	3×3	1	$13 \times 13 \times 384$
5	Convolution	3×3	1	$13 \times 13 \times 256$
	Max pooling	3×3	2	$6 \times 6 \times 256$
6	Fully connected			9216
7	Fully connected			4096
8	Fully connected			4096
9	Activation (softmax)			2

Three hyper-parameters of the CNN were allowed to vary. The kernel size of the first layer could take values 7×7 or 15×15 pixels in order to evaluate how this characteristic was related to the dimensions of the defects. The dropout value in the fully connected layers was comprised between 0.1 and 0.5. The ratio of initial images used for fitting was automatically chosen for statistical purposes and varied between 0.1 to 0.3.



(a) Images without defects.



(b) Images with defects.

Fig. 2: Typical dataset images extracted from the initial micrographs. Field of view: $293 \times 293 \mu\text{m}^2$, size: 227×227 pixels.

2.4 Performance evolution

After performing the seventeen experiments from the DoE, the model with the higher accuracy was selected for a further investigation aimed at measuring the contribution of each layer to the final binary classification. Previous authors derived various metrics from information theory and Shannon entropy for evaluating the performance of a classifier, e.g. Confusion Entropy (CEN) [48], Modified CEN [49] or mutual information score. In this work, the concept of mutual information was considered. Indeed, as discussed in [42], it provides a mathematically grounded metric for quantifying the information flow through the CNN layers. Given two discrete random variables F and C (e.g. feature and class of an object), the mutual information can be expressed by the difference between the marginal entropies of F and C , and their joint entropy:

$$I(F, C) = H(F) + H(C) - H(F, C) \quad (1)$$

where $I(F, C)$ is the quantity of information obtained on one variable by the observation of the other. $H(F)$ and $H(C)$ are the marginal entropies, and $H(F, C)$ is the joint entropy. A perfect classifier is characterized by mutual information of 1, while the worst one, in the case of two classes and two labels, is 0.

For each layer of the model, the performance of each filter as a classifier was measured by calculating its mutual information score. Figure 3 sums up the principle of the mutual information analysis for the best model.

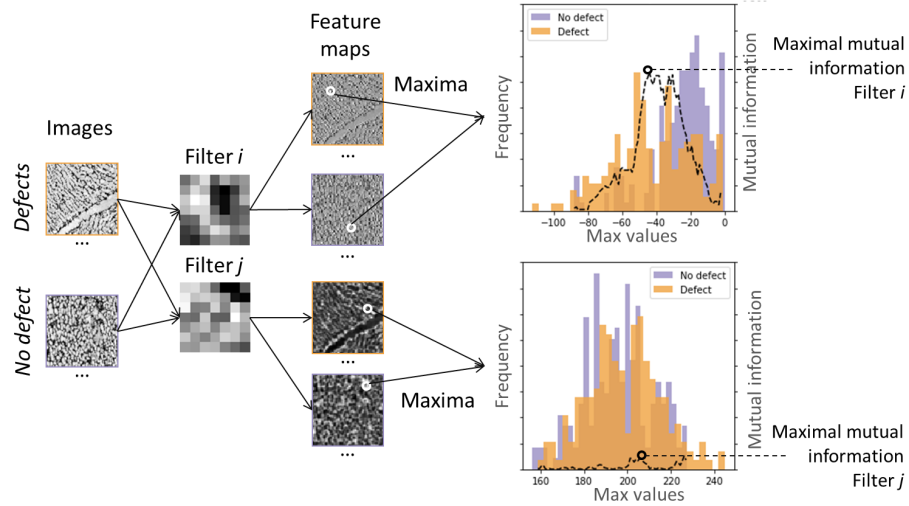


Fig. 3: Principle of the CNN mutual information analysis for two filters i and j . The output maximal mutual information, depicted on the right, shows that filter i has a higher discriminating power.

The calculation was based on the confusion matrix summing up the relative amounts of true positives (defect correctly detected), true negatives (absence of defect correctly identified), and false positives and false negatives (presence/absence of defect wrongfully detected).

A set of randomly chosen 200 images, 100 with defects, 100 without, were analyzed by the trained CNN. For each filter, the maximum values of the feature maps are recorded and reported in two histograms, one for each class with/without defect. Depending on the capacity of the classifier, the histograms can be more or less distinct. The mutual information is calculated for discrimination threshold values spanning the whole range of the histograms, in the same manner as for the derivation of a Receiver operating characteristic (ROC) curve. The performance of the filter for separating both classes is then considered as the maximal mutual information value found, corresponding to the optimal threshold value. This operation is repeated for all the filters of each layer. The maximal mutual information values are then recorded and reported in a series of 5 histograms, one for each convolutional layer, summing up the abilities of its filters to correctly label the images.

3 Results

The learning progression of a model is represented by the evolution of the accuracy at each iteration. The Figure 4 presents the learning curves of the 16 experiments.

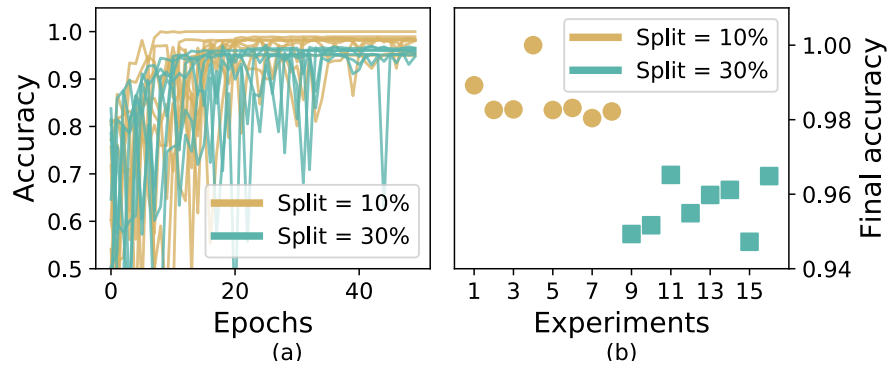


Fig. 4: (a): Bundle of learning curves obtained from the DoE; (b): Final accuracy values showing two distinct groups according to the validation split value 0.1 or 0.3.

Even though there are six varying parameters, the behavior of learning curves diverges into two different ways. The two groups revealed on the Figure 4b are differentiated by the validation split ratio. In the case of a validation split of

30%, the mean accuracy of the model after 50 iterations reaches 95.67%. In the case of a validation split of 10%, this value of mean accuracy increase to 98.54%. It appears that all models are stabilized at the end of the experiment, i.e. after 50 iterations.

To evaluate the classification ability of the model layer by layer, the mutual information is assessed for each filter, as mentioned earlier (Figure 3). It is then possible to observe the successive convolution of two images of defect and no-defect by the best filters of each layer through the network. The intensity distribution of each feature map is reported in histograms in Figure 5.

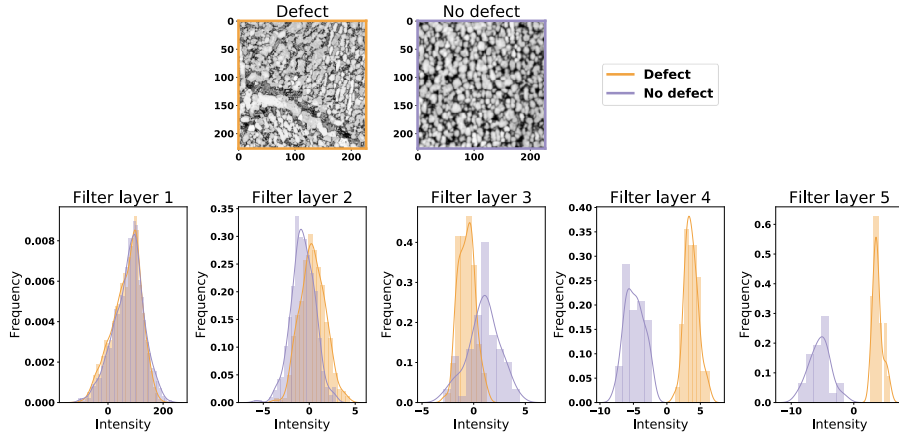


Fig. 5: Intensity histogram of the feature maps obtained with the best filter of each layer. Successful classification after the third layer.

While both histograms are overlaid in the two first layers, they start to separate at the third layer. The fourth and fifth layers exhibit two distinct distributions of intensity that reveal the successful classification of the initial images.

To evaluate the effect of the sixth factors studied in DoE, a diagram of the main effects is presented in Figure 6.

As mentioned in Figure 4, the validation split is of first importance in the final accuracy of the model. By comparison, the other parameters are of minor impact. The kernel size of the first convolution layer has a slight influence on the ultimate accuracy and has to be set to 7x7 pixels to increase the model performance. The overlap between successive images, which can induce redundant information, does not seem to affect the response of the models.

The mutual information histograms of all filters in each convolutional layer (model 6 in Figure 4) are plotted in Figure 7a. The histogram for the first convolutional layer is bimodal, showing that a significant amount of high-level filters induce different responses from both classes of images. The overall distribution for layer 2 is gathered towards low values. The distributions for layers 3, 4 and

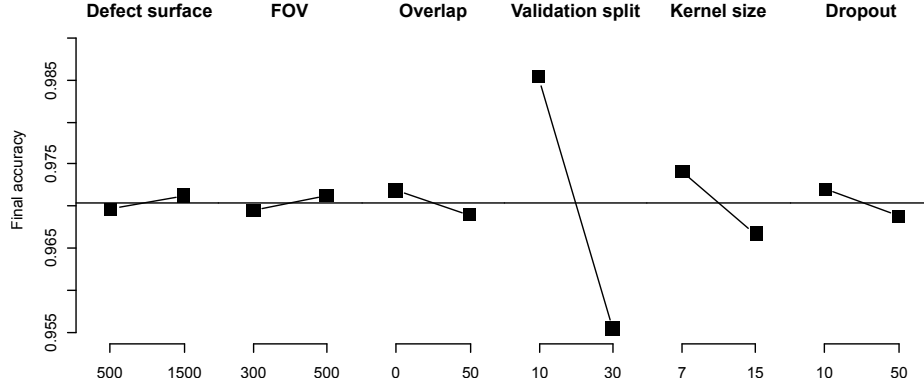


Fig. 6: Results of Design of Experiments. Main effects diagram.

5 appear similar, yet with increasing extreme values. This suggests that some low-level/long range filters perform better and better with the CNN depth. This behavior is estimated in Figure 7b, where the mutual information value separating each histogram in a 90 vs 10% frequency ratio (highest decile) is plotted against the layer number.

The last decile of the mutual information first decreases between the first and second layers, then linearly increases up to the fifth layer. This expresses the cumulative information gain brought by each deep layer and suggests that these are necessary to capture the low level features that distinguish the images with and without defects.

At the end of the model, the fifth layer is connected to three successive dense layers. A softmax function finally classifies the images. The mutual information can also be calculated at this step. The result is presented in Figure 8. The mutual information reaches 0.74 and succeeds in classifying 200 images with 93 true positives, 98 true negatives, 2 false positives and 7 false negatives.

4 Discussion

The results presented in Figure 4 indicate that all the models quickly reach an excellent accuracy, higher than 95%. As a matter of fact, one of the models achieves a theoretical accuracy of 100% (experiment 4 on Figure 4b). However, the accuracy metric is highly correlated to the number of pictures randomly selected to validate the model at each iteration. For that reason, the final accuracy can take values exceptionally high and must be put into perspective.

Accordingly, the consideration of the mutual information brings some additional knowledge about the network operation. The analysis of the mutual information through the network shows that the classification efficiency is non-linear. First of all, the bi-modality of the mutual information distribution observed in

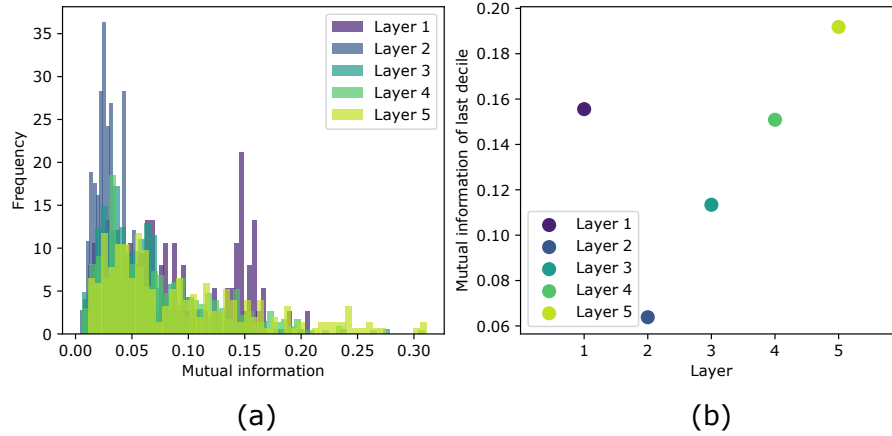


Fig. 7: (a): Distribution of mutual information in successive layers of model 6;
(b): Highest decile values of histograms shown in (a).

the first layer can be attributed to the heterogeneity of the textures at a small scale (as evidenced by Figure 2a). Indeed, a recent study on Texture-CNN (T-CNN) [50] showed that the three first layers of AlexNet are efficient in textures recognition, whereas the fourth and fifth layers are more suitable to describe shapes. In the present work, the classification efficiency increases from the third layer and becomes significant starting from the fourth layer. In the two last layers, the mutual information is high enough to separate the α_{GB} defects from the background texture. Based on this observation, a typical α_{GB} defect seems to more recognizable by the texture discontinuity that it causes than by its texture itself.

The consideration of 227×227 pixels tiles instead of large initial images has been made in the first instance to increase the dataset size and so to improve the final model accuracy. Additionally, this method can be reversed to return the localization of the defects in the initial images. For example, Figure 9 shows a map in which the colors correspond to the average occurrence probability of a defect. It was obtained by dividing the initial image in 227×227 pixels tiles separated by 5 pixels which were processed by the trained model. The final classification score of each tile was then summed and averaged at each pixel to produce the colormap overlaid on the initial image. The central defect is clearly detected (values close to 1). The upper left zone of the image appears with slightly lower values, which may be caused by the directional texture observed at this location, different from the rest of the image.

Further work shall concentrate in identifying the images that respond well to the classification, especially at the first layer. Conversely, visualizing the best performing filters of this layer will bring insights in the most significant features that make up the texture representation of the material.

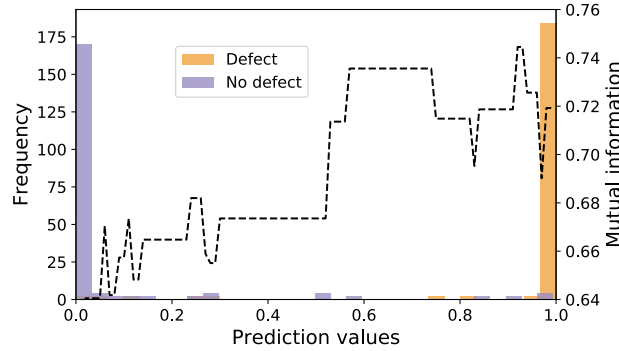


Fig. 8: Mutual information after the last activation layer (softmax).

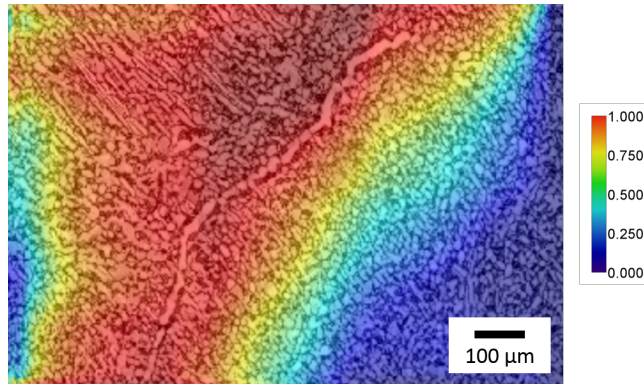


Fig. 9: Prediction map for a typical micrograph with overlaid tones corresponding to the detection probability of a defect.

5 Conclusion

In this paper, the use of AlexNet architecture was found to be a promising solution for the detection of specific geometric features in materials images (α_{GB} in Ti-6Al-4V micrographs). In a first approach, a design of experiments including six factors with two levels, related to the dataset and to the CNN, revealed an optimal model with an ultimate classification accuracy of 98.54%. Among these six parameters, the validation split ratio in the training step is of major importance. The ultimate accuracy increases when the validation split ratio decrease in the range of 30% to 10%.

In addition to the statistical analysis of the main effects, the mutual information provided by the entropy theory has been used. This approach provides a powerful tool to understand the evolution of the information gain through the network. More precisely, it allows to evaluate the benefits of each filter in each

layer. In the general context of texture classification, the convolution Neural Networks based on AlexNet are efficient in the first layers. But in the specific case of α_{GB} detection, the use of mutual information theory pointed out the benefits brought by deep layers. In this way, and contrary to the initial approach, it has been stated that the real challenge in the α_{GB} detection analysis lies in the microstructure discontinuity characterization rather than in texture recognition. In addition to that statement, there is strong evidence that adding deeper layers to the AlexNet architecture could increase the efficiency of the classification. Further efforts will focus on this opportunity to improve the efficiency of AlexNet in this context.

References

1. E O Hall. The deformation and ageing of mild steel: III discussion of results. *Proceedings of the Physical Society. Section B*, 64(9):747–753, 1951.
2. NJ Petch. The cleavage strength of polycrystals. *Journal of the Iron and Steel Institute*, 174:25–28, 1953.
3. Pascal Louis and Arun M. Gokhale. Application of image analysis for characterization of spatial arrangements of features in microstructure. *Metallurgical and Materials Transactions A*, 26(6):1449–1456, 1995.
4. Leszek Wojnar. *Image Analysis: Applications in Materials Engineering (Materials Science Technology)*. CRC Press, October 1998.
5. K Geels, DB Fowler, W-U Kopp, and M Ruckert. Quantitative metallography/materialography—an introduction. In *Metallographic and Materialographic Specimen Preparation, Light Microscopy, Image Analysis and Hardness Testing*, pages 565–565–12. ASTM International, January 2007.
6. Coll. *Stereology and image analysis in materials science : selected, peer reviewed papers from the IX International Conference on Stereology and Image Analysis in Materials Science (STERMAT 2012) 3-6 September, 2012, Zakopane, Poland*. Solid state phenomena. Trans Tech Publications Ltd, Durnten-Zurich, Switzerland, [2013].
7. Gervais Gauthier, Michel Coster, Liliane Chermant, and Jean-Louis Chermant. Morphological segmentation of cutting tools. *Microscopy Microanalysis Microstructures*, 7(5-6):339–344, September 1996.
8. Tomasz Wejrzanowski, Wojciech Spychalski, Krzysztof Różniatowski, and Krzysztof Kurzydłowski. Image based analysis of complex microstructures of engineering materials. *International Journal of Applied Mathematics and Computer Science*, 18(1):33–39, 2008.
9. S.G. Lee, Y. Mao, A.M. Gokhale, J. Harris, and M.F. Horstemeyer. Application of digital image processing for automatic detection and characterization of cracked constituent particles/inclusions in wrought aluminum alloys. *Materials Characterization*, 60(9):964–970, 2009.
10. Siddhartha Banerjee, Swarup Kumar Ghosh, Shubhabrata Datta, and Sanjoy Kumar Saha. Segmentation of dual phase steel micrograph: An automated approach. *Measurement*, 46(8):2435–2440, 2013.
11. N. Vanderesse, M. Anderson, F. Bridier, and P. Bocher. Inter- and intragranular delta phase quantitative characterization in inconel 718 by means of image analysis. *Journal of Microscopy*, 261(1):79–87, January 2016.

12. Shilan Meimandi, Nicolas Vanderesse, Denis Thibault, Philippe Bocher, and Martin Viens. Macro-defects characterization in cast CA-6nm martensitic stainless steel. *Materials Characterization*, 124:31–39, 2017.
13. Tanusree Dutta, Debdulal Das, Siddhartha Banerjee, Sanjoy Kumar Saha, and Shubhabrata Datta. An automated morphological classification of ferrite-martensite dual-phase microstructures. *Measurement*, 137:595–603, April 2019.
14. A. Cord, F. Bach, and D. Jeulin. Texture classification by statistical learning from morphological image processing: application to metallic surfaces. *Journal of Microscopy*, March 2010.
15. Antonino Ducato, Livan Fratini, Marco La Cascia, and Giuseppe Mazzola. An automated visual inspection system for the classification of the phases of ti-6al-4v titanium alloy. In *Computer Analysis of Images and Patterns*, pages 362–369. Springer Berlin Heidelberg, August 2013.
16. Brian L. DeCost and Elizabeth A. Holm. A computer vision approach for automated analysis and classification of microstructural image data. *Computational Materials Science*, 110:126–133, 2015.
17. Dominique Jeulin. Morphological probabilistic hierarchies for texture segmentation. *Mathematical Morphology - Theory and Applications*, 1(1), January 2016.
18. Subir Gupta, Jit Sarkar, M. Kundu, N.R. Bandyopadhyay, and Subhas Ganguly. Automatic recognition of SEM microstructure and phases of steel using LBP and random decision forest operator. *Measurement*, 151:107224, 2020.
19. Victor Hugo C. de Albuquerque, Auzuir Ripardo de Alexandria, Paulo César Cortez, and João Manuel R.S. Tavares. Evaluation of multilayer perceptron and self-organizing map neural network topologies applied on microstructure segmentation from metallographic images. *NDT & E International*, 42(7):644–651, 2009.
20. Aritra Chowdhury, Elizabeth Kautz, Bülent Yener, and Daniel Lewis. Image driven machine learning methods for microstructure recognition. *Computational Materials Science*, 123:176–187, 2016.
21. Brian L. DeCost, Toby Francis, and Elizabeth A. Holm. Exploring the microstructure manifold: Image texture representations applied to ultrahigh carbon steel microstructures. *Acta Materialia*, 133:30–40, 2017.
22. Jinhua Lin, Lin Ma, and Yu Yao. Segmentation of casting defect regions for the extraction of microstructural properties. *Engineering Applications of Artificial Intelligence*, 85:150–163, 2019.
23. Tiberiu Stan, Zachary T. Thompson, and Peter W. Voorhees. Optimizing convolutional neural networks to perform semantic segmentation on large materials imaging datasets: X-ray tomography and serial sectioning. *Materials Characterization*, 160:110119, 2020.
24. Dennis M. Dimiduk, Elizabeth A. Holm, and Stephen R. Niezgoda. Perspectives on the impact of machine learning, deep learning, and artificial intelligence on materials, processes, and structures engineering. *Integrating Materials and Manufacturing Innovation*, 7(3):157–172, August 2018.
25. George F. Vander Voort. *ASM Handbook, Volume 9: Metallography And Microstructures (ASM Handbook) (ASM Handbook)*. ASM International, December 2004.
26. Gerd Lütjering and James C Williams. *Titanium*. Springer, New York, 2007.
27. Jacek Chrapoński and Wojciech Szkliniarz. Quantitative metallography of two-phase titanium alloys. *Materials Characterization*, 46(2-3):149–154, 2001.
28. J Tiley, T Searles, E Lee, S Kar, R Banerjee, J.C Russ, and H.L Fraser. Quantification of microstructural features in / titanium alloys. *Materials Science and Engineering: A*, 372(1-2):191–198, May 2004.

29. N. Vanderesse, E. Maire, M. Darrieulat, F. Montheillet, M. Moreaud, and D. Jeulin. Three-dimensional microtomographic study of widmanstätten microstructures in an alpha/beta titanium alloy. *Scripta Materialia*, 58(6):512–515, 2008.
30. Hongwei Li, Zhe Ji, and He Yang. Quantitative characterization of lamellar and equiaxed alpha phases of () titanium alloy using a robust approach for touching features splitting. *Materials Characterization*, 76:6–20, 2013.
31. Hao Zhao, Alistair Ho, Alec Davis, Alphons Antonysamy, and Philip Prangnell. Automated image mapping and quantification of microstructure heterogeneity in additive manufactured ti6al4v. *Materials Characterization*, 147:131–145, January 2019.
32. Hemant Sharma, Stefan M.C. van Bohemen, Roumen H. Petrov, and Jilt Sietsma. Three-dimensional analysis of microstructures in titanium. *Acta Materialia*, 58(7):2399–2407, april 2010.
33. John W. Foltz, Brian Welk, Peter C. Collins, Hamish L. Fraser, and James C. Williams. Formation of grain boundary α in β Ti alloys: Its role in deformation and fracture behavior of these alloys. In *Metallurgical and Materials Transactions A: Physical Metallurgy and Materials Science*, volume 42, pages 645–650, March 2011.
34. R. Shi and Y. Wang. Variant selection during α precipitation in Ti-6Al-4V under the influence of local stress - A simulation study. *Acta Materialia*, 61(16):6006–6024, septembre 2013.
35. Ken Chatfield, Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Return of the devil in the details: Delving deep into convolutional nets, 2014.
36. Li Liu, Jie Chen, Paul Fieguth, Guoying Zhao, Rama Chellappa, and Matti Pietikäinen. From BoW to CNN: Two decades of texture representation for texture classification. *International Journal of Computer Vision*, 127(1):74–109, 2018.
37. Li Liu, Wanli Ouyang, Xiaogang Wang, Paul W. Fieguth, Jie Chen, Xinwang Liu, and Matti Pietikäinen. Deep learning for generic object detection: A survey. *CoRR*, abs/1809.02165, 2018.
38. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.
39. Shui Hua Wang, Yi Ding Lv, Yuxiu Sui, Shuai Liu, Su Jing Wang, and Yu Dong Zhang. Alcoholism detection by data augmentation and convolutional neural network with stochastic pooling. *Journal of Medical Systems*, 42(1), jan 2018.
40. Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, April 2012.
41. Naftali Tishby and Noga Zaslavsky. Deep learning and the information bottleneck principle. In *2015 IEEE Information Theory Workshop (ITW)*, pages 1–5. IEEE, 2015.
42. Ahmad Chaddad, Behnaz Naisiri, Marco Pedersoli, Eric Granger, Christian Desrosiers, and Matthew Toews. Modeling Information Flow Through Deep Neural Networks. *arXiv e-prints*, page arXiv:1712.00003, November 2017.
43. Caroline A Schneider, Wayne S Rasband, and Kevin W Eliceiri. NIH image to ImageJ: 25 years of image analysis. *Nature Methods*, 9(7):671–675, 2012.
44. Johannes Schindelin, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch, Curtis Rueden, Stephan Saalfeld, Benjamin Schmid, Jean-Yves Tinevez, Daniel James White, Volker Hartenstein, Kevin Eliceiri, Pavel Tomancak, and Albert Cardona. Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9(7):676–682, 2012.

45. J Brocher. The biovoxxel image processing and analysis toolbox. In *European BioImage Analysis Symposium. Paris, France*, 2015.
46. Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
47. François Chollet et al. Keras. <https://keras.io>, 2015.
48. Jin-Mao Wei, Xiao-Jie Yuan, Qing-Hua Hu, and Shu-Qin Wang. A novel measure for evaluating classifiers. *Expert Systems with Applications*, 37(5):3799–3809, May 2010.
49. Rosario Delgado and J. David Núñez-González. Enhancing confusion entropy as measure for evaluating classifiers. In *Advances in Intelligent Systems and Computing*, pages 79–89. Springer International Publishing, 2018.
50. Vincent Andrearczyk and Paul F. Whelan. Using filter banks in convolutional neural networks for texture classification. *Pattern Recognition Letters*, 84:63–69, 2016.