




Deep learning enhanced Watershed for microstructural analysis using a boundary class semantic segmentation

G. Fotos^{1,*} , A. Campbell², P. Murray², and E. Yakushina¹

¹ Department of Design Manufacture and Engineering Management, University of Strathclyde, Glasgow, UK

² Department of Electronics and Electrical Engineering, University of Strathclyde, Glasgow, UK

Received: 14 July 2023

Accepted: 25 August 2023

© The Author(s), 2023

ABSTRACT

The mechanical properties of the materials are determined by the size and morphology of fine microscopic features. Quantitative microstructural analysis is a key factor to establish the correlation between the mechanical properties and the thermomechanical treatment under which material condition has been achieved. As such, microstructural analysis is a very important and complex task within the manufacturing sector. Published standards are used for metallographic analysis but typically involve extensive manual interpretation of grain boundaries, resulting in measurements that are slow to produce, difficult to repeat and highly subjective. Computer vision and the evolution of artificial intelligence in the past decade can offer solutions to such problems. Deep learning and digital image processing techniques allow digital microstructural analysis to be automated using a fast and repeatable method. This paper proposes a novel boundary class semantic segmentation approach (BCSS) to identify each phase of the microstructure and additionally estimate the location of the grain boundaries. The BCSS is then combined with more traditional segmentation techniques based on the Watershed Transform to improve the identification and measurement of each feature within the microstructure using a new, hybrid automated digital microstructure analysis approach. The new method is validated on a published dataset of two-phase titanium alloy microstructure pictures captured using a scanning electron microscope. Measurements match the level of accuracy of accepted manual standards, and the method is demonstrated to be more reliable than other automated approaches. The influence of the subjective nature of manual labelling, required to train the proposed network, is also evaluated.

Handling Editor: Ghanshyam Pilania.

Address correspondence to E-mail: georgios.fotos@strath.ac.uk

<https://doi.org/10.1007/s10853-023-08901-w>

Published online: 19 September 2023

Introduction

Microstructural analysis is a key factor to quality assurance of manufactured parts, as well as understanding and designing of the optimal manufacturing processes. By studying the microstructural morphology, useful information about the mechanical and physicochemical properties of the material can be extracted and correlated with prior thermomechanical treatment of the material. In this study, micrographs of bimodal Ti6Al4V microstructures were used. Bimodal microstructure is an important microstructural condition of titanium alloy where finer grains contribute towards increased strength whereas coarser grains facilitate increased ductility [1]. Accurate quantitative microstructural analysis might influence targeted manufacturing processes, so each part produced has a required microstructure and mechanical properties. This is likely to lead to fewer parts failing quality assurance checks, thus improving sustainability by minimising waste, reducing costs and time effort for the manufacturing sector.

Microstructural analysis can be divided into two groups, automated and manual. There are some advantages and disadvantages for both methods. Manual microstructural analysis is typically performed by expert material scientists using global microstructural analysis standards such as [2, 3] to measure microstructural features. This approach is versatile and uses intuitive tools, but is prone to human error and subjectivity, making measurements challenging to repeat consistently. If complex micrographs are analysed by multiple material scientists, it is expected that there may be significant deviations in the measurements [2]. Manual methods also require the most time from expert scientists and are harder to scale up to measuring large datasets.

Automated microstructural analysis is typically performed with the use of bespoke computer vision software. The benefit of automated microstructural analysis is that it avoids human error, significantly faster and can be repeated consistently. This kind of software usually utilises conventional hand-crafted algorithms designed for digital microstructural analysis. Some of the basic techniques that the traditional algorithms utilise to segment object in an image are based on mathematical morphology [4–6]. Conventional hand-designed algorithms have proven more popular in this domain and can achieve good segmentation results even with small dataset [7–10]. Nevertheless, accurate

result often depends upon non-trivial parameterisation of different methods within the analysis pipeline. For example a technique that exists to segment grains using mathematical morphology but achieving optimal results requires different filter parameters for each microscope type and sensitivity parameter, set for each dataset [7]. This adds additional steps and risks reintroducing user bias when compared to non-parameterisable methods. Phase segmentation is also required in the pre-processing stage which is challenging for complex microstructures. Although the above techniques have been used for a long time, they do not always achieve good results in all cases. In the past decade, a subset of AI that is called deep learning has revolutionised automated image processing task in a range of different fields, such as autonomous driving [11] and medicine [12]. Recent object detection and semantic segmentation approaches have detected complex features with a level of accuracy seemingly unachievable via conventional algorithms [13, 14]. Networks are trained algorithmically, with the designer only required to provide the training algorithm with examples of input data and the correct solution which the network would be expected to achieve. Such techniques are widely used for semantic segmentation of real-world data and have applications in driverless cars where it is used to identify people, cars, bikes and other potential obstacles [11]. The power of deep learning in such cases is that these features are often easily recognised by a human but difficult to write a descriptor for in terms of what precise combination of contrast and colour indicates their presence. Therefore, it can be easier to train an accurate deep learning method than it is to design an effective conventional algorithm to perform the same task. For microstructural analysis, this property of semantic segmentation is also beneficial as many microstructural features can be identified by a trained eye but are difficult to define a precise description. This makes conventional algorithms hard to design. For example material scientists could not be confused by the noise in the image, like dust or scratches for grains and platelets, yet both can be described as circular or elliptical shapes that differ in intensity from their surrounding pixels. The experts can recognise this as they learnt by example, not by strict definitions. It is that same method of learning that allows deep learning to perform well in such cases. However, the use of this technology within many aspects of materials science and manufacturing has been more limited, particularly in the automated

analysis of metallic microstructures. Recent reviews of deep learning for medical image analysis easily find hundreds of relevant works with access to a huge amount of medical data [12] within the past years [13, 14], yet a similar review for metallographic investigation, especially on the microstructural analysis found a distinct lack of relevant publicly available data [15]. Reasons for this include limited public microstructural datasets, at least partially due to some industrial data being considered IP sensitive, as well as the substantial variations between different microstructural images, even when images are of the same material. The methods that have been proposed typically focus on global classification of the microstructure (i.e. globular or lamellar), rather than a quantitative measure of individual features, such as grain size [16].

Previous research

A few researchers tried to implement deep learning for phase classification for materials microstructure such as iron and titanium alloys with quite good results achieved. Chowdhury et al. [17] presented a method where deep learning was used to train a network to make two distinctions—does the microstructure contain dendrites (tree-/branch-shaped structures), and if it does is the cross-section studied longitudinal or transverse. Azimi et al. [18] proposed a deep learning method for microstructural classification in the example of certain microstructural constituents of low carbon steel with the use of fully convolutional neural networks (FCNN). Jang et al. [19] also used the residual neural networks (ResNets) to measure the volume fraction over acicular ferrite of carbon steel and distinguish it from the rest of the microstructural features. Very recently, Baskaran et al. [20] also used machine learning to classify different morphologies of the titanium microstructures. The network was trained to classify microstructural images as either lamellar, duplex or acicular. These authors [17–20] also published a large dataset that was used for training and testing of their approach. This is extremely useful as the lack of public datasets is recognised problem with the development of machine learning techniques in this area [15]. This contrasts greatly in comparison with the medical imaging field which has far more datasets and wider use of the modern ones learning approaches [13]. The above techniques while they accurately detect the targeted phase, fail to distinguish the grain boundaries among grains without being able

to extract useful feature information like the average grain size.

New approach

In this paper, we propose a novel, hybrid automated digital microstructural analysis (*HADMA*) approach where the abstract understanding of image features provided by deep learning, is combined with the precise segmentation rules of traditional hand-crafted algorithms. Specifically, we design and train a semantic segmentation network and combine this with Watershed Algorithm-based segmentation techniques [7]. To our knowledge, the Watershed-based method used here has not been combined with semantic segmentation approaches in prior. In addition to this, we propose a new strategy for training and deploying the semantic segmentation part of the algorithm we call boundary class semantic segmentation (*BCSS*). This approach incorporates additional classes in the network to identify boundaries between the instances of each class rather than using the typical approach of identifying only the original classes themselves. For microstructure analysis, this helps separate clusters of, for example alpha grains into the individual grains by capturing information of probable boundary locations, when traditional semantic segmentation strategies would identify the entire region as consistent of alpha grains without this alpha/alpha boundary information. The result from *BCSS* was then used as the input to a technique which is based on the Watershed Algorithm and the method [7], which improves the segmentation over the phases by completing the external–internal boundaries.

Technical background: deep learning, semantic segmentation and marker-based Watershed

Accurate, statistical analysis of the individual features within images, such as individual grains of complex microstructures, can be achieved via image segmentation. Image segmentation is the process of classifying each pixel in the image such that we know either to which particular object or which class it belongs. For segmentation of objects, pixels within the same object should be given the same level and thus form a single connected component (CC), from which a variety of measurements can be achieved using generalised

methods [21]. While measurement and classification of microstructures can be performed without segmentation [22], a segmentation facilitates statistical description on per grain basis and allows to produce maps that subsequently enable visual confirmation over the identified grains.

Semantic segmentation

Measurements of a segmented image are typically trivial; however, the process of obtaining the segmentation is not. In fact, it is widely considered among the most challenging in image processing, as generalised algorithms do not work in all domains [23–25]. Semantic segmentation is a deep learning method where the network is trained to label each pixel in an image according to which class of features it belongs, as illustrated in Fig. 1. Convolutional neural networks (CNN) are used for this task, and various network architectures have been proposed in recent years, such as ResNet-18, ResNet-50 U-net [26, 27]. The network is designed to initially iteratively downsample the image, using convolutional and pooling layers. During the process, the network uses a combination of weights adjust the value of each pixel/neuron. These weights are learnt algorithmically using machine learning methods such as Stochastic Gradient Descent or Adam Optimiser [28]; thus, the only input required from the user is raw data and an expected result.

The intention is that early layers in the network capture low-level details in the image such as edges, textures and simple shapes, with deeper layers capturing large-scale information like complex and finer details [29, 30]. For microstructures, this could mean

early layers detect changes in image intensity, subsequent layers identify a combination of these changes as a boundary and latter layers detect a combination of these as a grain of a particular type. Once a low-resolution classification is reached the network will then upsample this result to achieve a 1-to-1 pixel map with the input image. This is illustrated in Fig. 2.

Marker-based Watershed Algorithm

While semantic segmentation is proven to be effective in a range of applications (automotive driving [11], physics [31]), it only segments an image class-wise. This means that it will not necessarily segment each instance of a class, particularly when these instances overlap, as is often the case with globular alpha grains. Techniques from the area of Mathematical Morphology [4] provide methods to segment a class into separate instances, even when this overlap. A method called the Watershed Algorithm is particularly effective at this and has been recently demonstrated for material science applications [7, 10]. The Watershed Algorithm requires a transformation of an image such that the maximal value typically exists at boundary locations. The Watershed Algorithm then considers this as a topographic surface and floods from low to high values, with each flood only stopping when it overlaps with another, as shown in Fig. 3. Each flood forms a catchment basin representing an object, and the location where this occurs is marked as a boundary. The process can be controlled by selecting markers from which each flood should begin [5]. We computed the markers for the Watershed Transform with the use of local maxima [32] over the distance transform [33]

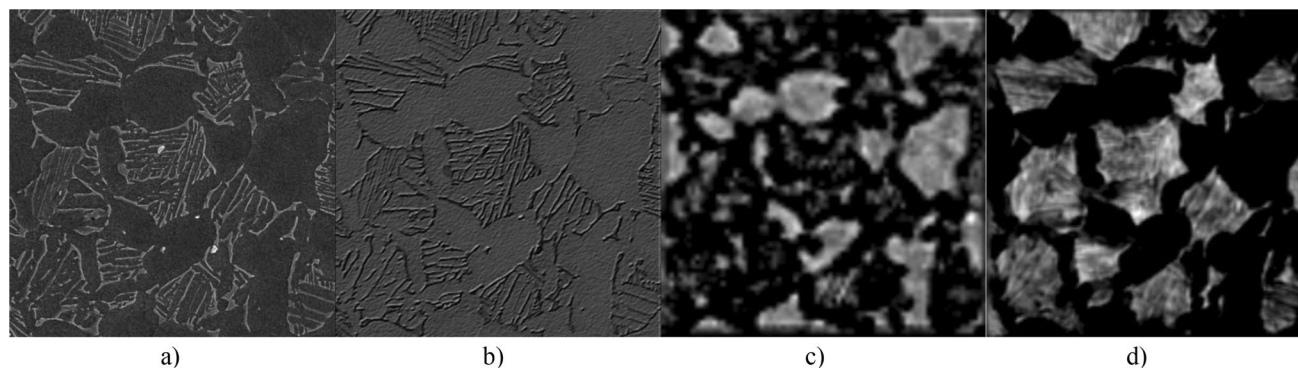


Figure 1 Different feature detections between early and latter layers of the ResNet-50. **a** An original Ti6Al4V bimodal image. **b** The 4th layer which is ReLU activation layer and the **c** 75th

ReLU activation layer and **d** 201st ReLU activation layer. All three images show the strongest activation channels of each layer.

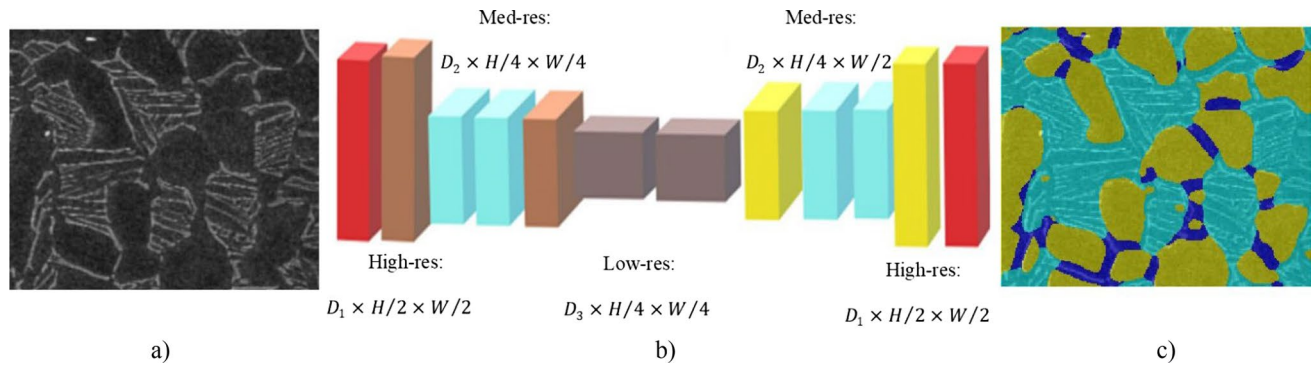


Figure 2 Illustration of the semantic segmentation, **a** input image of bimodal microstructure of Ti6Al4V [20], **b** the architecture of ResNet (features extraction), **c** phase detection after semantic segmentation (yellow colour shows the alpha equiaxed grains, light blue colour shows the lath colonies and the dark blue colour shows the grain boundaries).

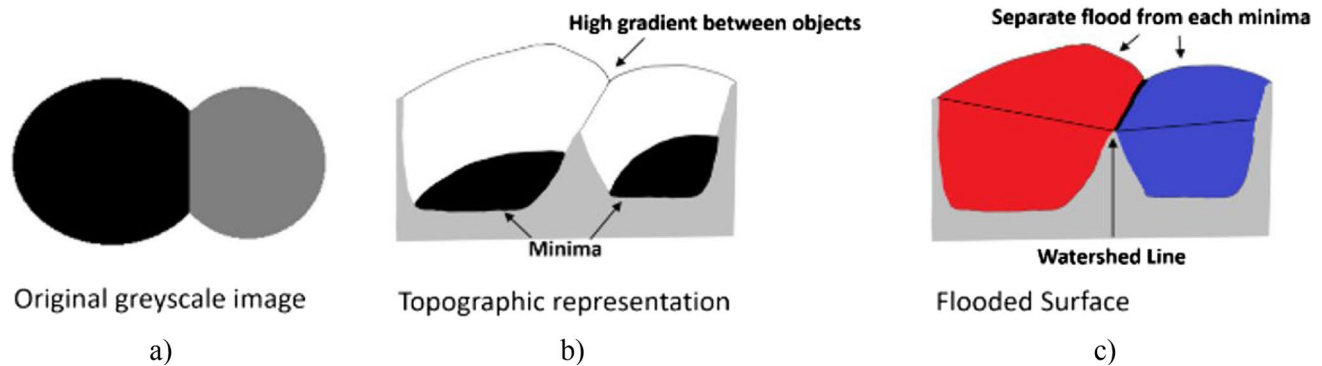


Figure 3 Illustration of Watershed Algorithm, **a** touching objects, **b** minima–maxima and the boundaries and **c** separation of each minima area.

of the inverse of the estimated boundaries, as in [7]. This produces one marker in approximate centres of each apparent grain, which is used to compute the segmentation. The h-maxima transform [32] is used to suppress noise and reduce over-segmentation problems [34].

The advantage of this algorithm is that it always provides a complete segmentation and always contains one region per selected marker.

Although using the Watershed to segment objects in an image with well-defined boundaries is usually an easy task, segmenting the morphology of the microstructure of Ti6Al4V makes it quite challenging. There are conditions/microstructures, where the colour intensity between the two phases is very similar or when the alpha phase of titanium categorised into different classes such as globular or equiaxed alpha grains and alpha platelets. For instance, lath colonies in bimodal microstructure of Ti6Al4V consist of the alpha and beta phase too.

Distinguishing the globular alpha grains from alpha platelets of the lath colonies is one of the biggest challenges of this research as share similar colour intensity. These conditions tend to create over- and under-segmentation problems as well as it is difficult to distinguish the alpha phase classes. As a result, the conventional approach is not always effective in the most complex cases like the identification of titanium’s phases in bimodal microstructure. In the cases, where the Watershed Algorithm does not segment accurately the titanium phases make the individual image editing important. Developing a new algorithm or improve the Watershed Algorithm as a second-stage analysis for each challenging image makes the whole process a labour consuming task while accurate and repeatable results are not granted.

Method: hybrid automated digital microstructural analysis

A hybrid automated digital microstructural analysis (HADMA) pipeline is proposed and combines a novel semantic segmentation approach (BCSS) and grain segmentation (Watershed Algorithm). The semantic segmentation identifies each phase of material, with the additional boundary class introduced in this work also providing some separation between each grain. The Watershed Algorithm is used in post-processing, to ensure that all grains are completely separated. In this way, we combine the abstract understanding learnt by the network, with the robustness of the deterministic rules of conventional algorithms. This is illustrated in Fig. 4.

Phase separation with semantic segmentation: BCSS

We proposed a new semantic segmentation approach called the boundary class semantic segmentation technique (BCSS). This consists of a multi-class semantic segmentation network that can provide greater separation than the two-class semantic segmentation methodology from prior studies [17, 20] of each instance (grains/laths) by considering the boundary between grains as a unique class as a single-pixel error would be enough to erroneously merge instances. By introducing the BCSS, we managed to maximise the information obtained at the semantic segmentation stage. When a network is trained to classify image features, and individual

measurements of instances of each feature are required. This network is also trained to identify boundary pixels as additional class. So, for a 2-phase microstructure, then three classes would be identified by the network: phase 1, phase 2 and boundaries. In this way, it is possible for a semantic network to learn how to partially separate touching objects such as grains without additional object detection and post-processing. Existing training approaches typically only train the network to identify real classes over the features that exist in the data. Consequently, boundary pixels must be classified in either as the same class as the object they separate creating a single homogeneous region, or as part of another existing class in the image, so as to create the designed separation. This latter approach creates fundamental issues when training for microstructural analysis as the boundaries between touching objects of one class do not necessarily share properties with any other class. For example Fig. 5a shows that a bimodal microstructure consists of two classes of features, alpha and beta phase. While the alpha phase can be divided further into other two classes globular alpha and alpha platelets.

If the network was trained to identify only two classes, then it would be difficult for that network to separate individual touching alpha grains, the boundaries form the beta phase and the lath colonies, as shown in Fig. 4b. However, it is clear from this figure that these boundaries do not share properties with beta phase and thus attempting to train the network to identify them as such is problematic and might lead to other classification errors.

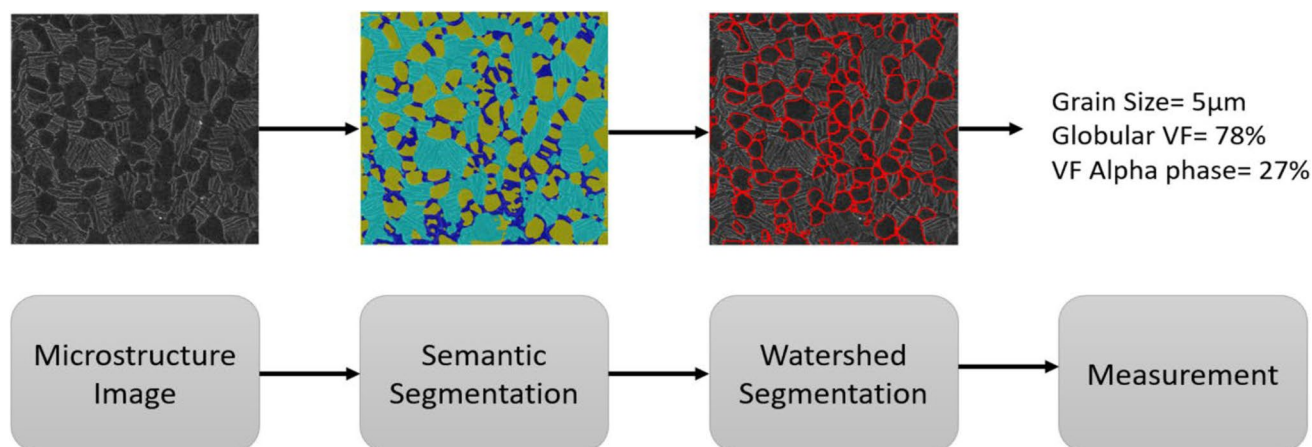


Figure 4 Pipeline for hybrid automated microstructural analysis (HADMA).

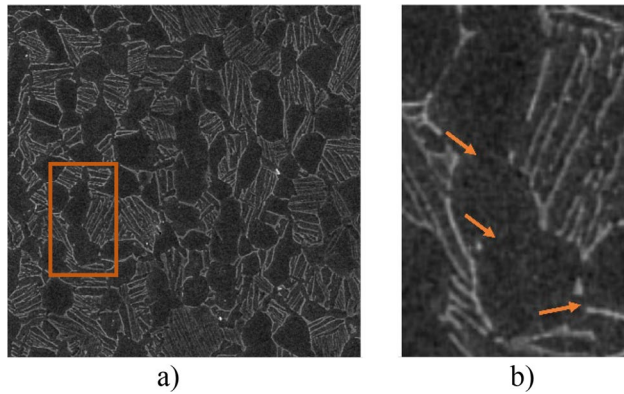


Figure 5 **a** Bimodal microstructure image from SEM and **b** cropped image highlighting with orange arrows the inner boundaries between touching globular grains.

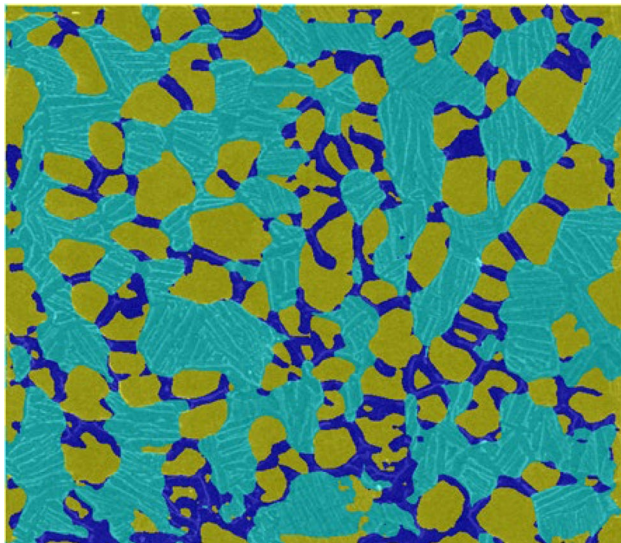


Figure 6 Phase classification with the use of semantic segmentation of the micrograph from 4a. The yellow, green and blue colours depict the alpha grains, lath colonies and grain boundaries, respectively.

We propose, therefore, to use a separate class to identify these boundaries. This decreases the inter-class variance while also ensuring that any measurement that considers the volume fraction of a particular phase is not distorted by these boundary regions being included in their class. The results of applying the final semantic segmentation network to the microstructure from Fig. 5a are shown in Fig. 6.

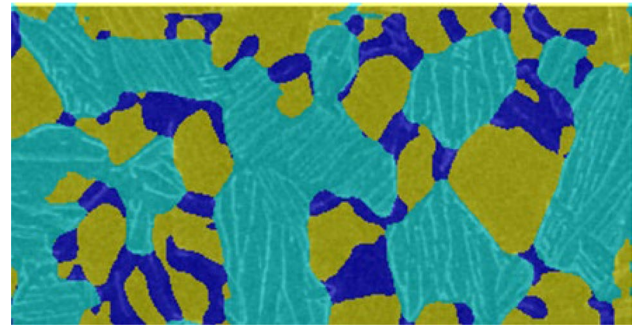


Figure 7 Oversized boundaries which cut area from alpha phase.

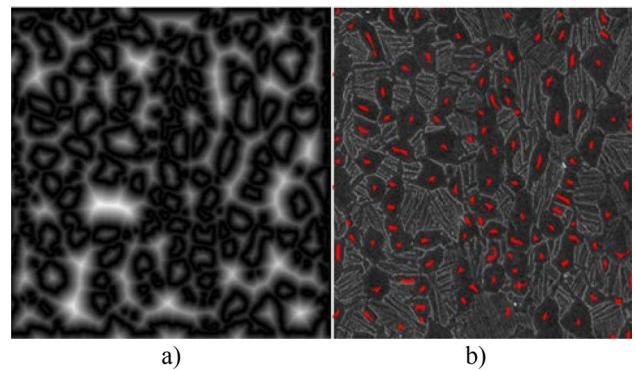


Figure 8 **a** Distance transform of the micrograph and **b** markers of each grain from semantic segmentation.

Grain segmentation (Watershed Algorithm)

Our proposed semantic segmentation method featuring a boundary class can provide some separation of the grains; however, an imperfect result, such as a few pixels of the alpha phase being misclassified as boundary or when the boundary pixels are classified as alpha, will risk under-segmentation and erroneous measurements, Fig. 7.

A Watershed Algorithm can be deployed, as the suggested technique from [7], to correct segmentation errors from possible misclassified pixels as well as to complete the inner boundaries. For the topographic function, the distance transform [33, 34] of the globular alpha class is used, as illustrated in Fig. 8a, and floods the image from each of these markers to find boundaries between grains. Ordinarily, this would erroneously over-segment the platelets; however, the prior semantic segmentation in the HADMA prevents this. By restraining the Watershed to only include globular alpha phase grains, a

detailed and accurate segmentation identifying each grain is produced, as shown in Fig. 9.

Neural network training and evaluation

Network selection–training

The novel BCSS training and deployment strategy in the proposed algorithm requires the use of a deep CNN that must be trained to recognise phases and boundaries of whichever material this is to be applied to existing network structures can be used in unison with our innovation. In this section, we describe how the network was chosen and trained for the experiment presented in this study.

In this study, ResNet-50 network was used for semantic segmentation, Fig. 10. This decision was based on the previous comparative studies and the fact that ResNet 50 is a popular neural network used

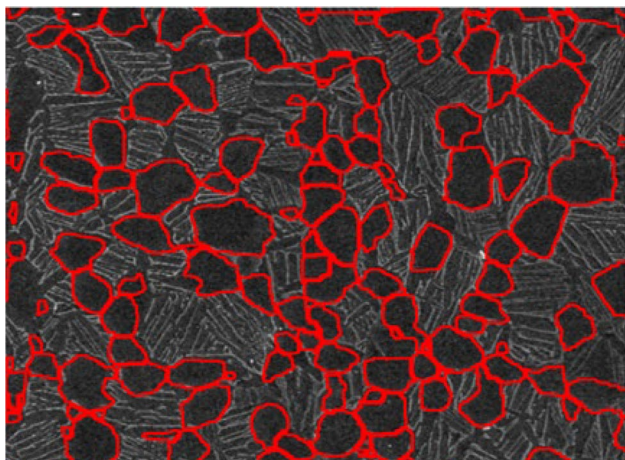


Figure 9 Final grain segmentation with proposed method HADMA.

for semantic segmentation [19] and object classification tasks [26]. We strongly believe that a wide range of neural networks can be employed for semantic segmentation. Moreover, a key contribution of our work is to introduce the third “boundary” class and enhance grain boundary detection. We believe that other neural networks can benefit from this approach due to additional separation created between overlapping instances of the same class. As there is no comprehensive research defining the optical neural network architecture for semantic on metallurgical images, we utilised the knowledge other disciplines, such as biology, to select the network. For instance, according to [35], U-Net–ResNet surpasses the Jaccard index and Dice coefficient of U-Net. Additionally, ResNet efficiently utilises GPU resources, and skip connections allow it to effectively capture context and improve its accuracy in image classification and semantic segmentation tasks [35]. ResNet-50 uses shortcut connections to address the issues of low accuracy and vanishing gradients [30]. This allows the network to decrease the training error and converge faster compared to other networks [30, 35] and has shown incredible capability in other domains [15, 19, 35, 36]. The network’s final layer (classification layer) is fully connected layer modified to have three channels, to match the three classes we wish to segment.

Training such a CNN to perform semantic segmentation requires training images and corresponding per-pixel labels (ground-truth image). In this work, SEM images of Ti6Al4V from a public dataset [20] were used with ground-truth data created manually for this study. Training deep networks from scratch often requires a vast amount of data that can be impractical to obtain for many materials analysis problems. To resolve this, we use a technique called transfer learning [34] to initialise the network with

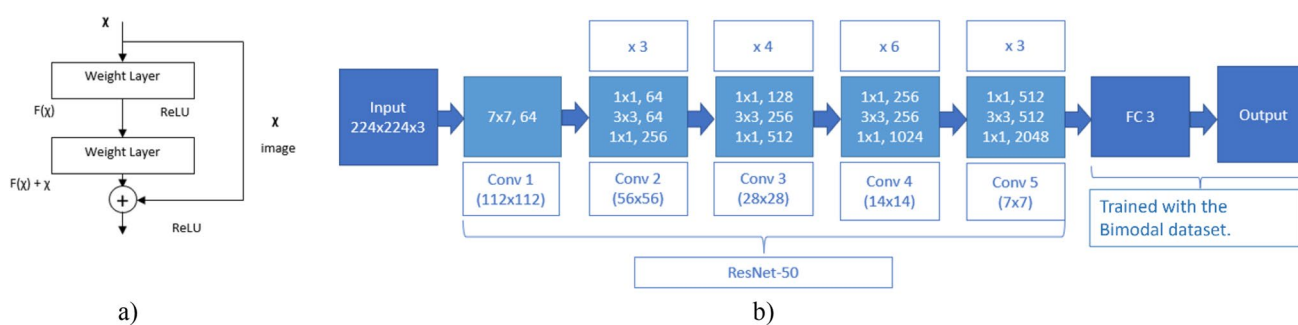


Figure 10 **a** Shortcut connections in the ResNet-50 and **b** architecture of the ResNet-50.

weights learnt from prior training in other applications. This can improve segmentation results with limited datasets and significantly reduces training time. A pre-trained ResNet-50 (ImageNet) network was used, with weights trained on 1,000,000 images recognising 1000 different objects [37]. While the pre-trained weights are unlikely to contribute to the training, the early layers, which often recognise the contrast and the edges, are often transferrable, reducing the amount of training required by up to 50% [38].

Our research was based on a published dataset by Baskaran et al. [20], which consisted of 1000 images displaying different microstructures of Ti6Al4V. For the purposes of our work, we focused on the bimodal micrographs, of which there were 329 in the original dataset, each with a size of 600 × 600 pixels. This dataset contains images of different microstructural morphologies, not all are suitable for measuring globular grains or volume fraction. For the purposes of this study, we selected 172 suitable microstructures which are manually analysed by a materials scientist to produce a ground-truth segmentation. For our training and validation, we selected 122 images, while we reserved 50 for testing. As our network has an input layer of 224 × 224, which is smaller than the original images, we must either crop or resize images for training. Cropping was selected so as to avoid the loss of fine spatial features. During the cropping process, we randomly selected regions of the original images to crop, allowing partial overlap between crops to prevent the loss of information that lay on the boundaries between crops. Based on our tests, which are shown in Table 1, we found that cropping the original images 50 times resulted in the highest class accuracy score for the alpha grains while also maintaining a reasonable training time. The resulting test–validation set contains 6100 images. To augment the dataset, we included original images from various magnification levels and rotated the cropped images by -10–10 degrees, promoting scale and rotation in variance in the trained network.

Table 1 Parameters that we used for the trail–error approach (We run all the possible combination for hyperparameters)

Learning rate	$[10^{-4}-10^{-1}]$	Logarithmically increase by 1 unit
Number of epochs	[20–50]	Integer increase by 5 units
Momentum	[0.9–0.99]	Decimal increase by 0.01 unit
Regularisation	$[10^{-4}-10^{-1}]$	Logarithmically increase by 1 unit
“Alpha” class weight factor	[1–10]	Integer increase by 1 unit

The network was trained using the solving method known as stochastic gradient descent with momentum (SGDM) with cross-entropy loss function [39]. The stochastic gradient with momentum is given by Eq. (1):

$$\theta_l = \theta_l - \alpha \nabla E(\theta_l) + \gamma(\theta_l + \theta_{l-1}), \tag{1}$$

where ℓ is the iteration number, $\alpha > 0$ is the learning rate, θ is the parameter vector and $E(\theta)$ is the loss function. In the standard gradient descent algorithm, the gradient of the loss function, $\nabla E(\theta)$, is evaluated using the entire training set, and the standard gradient descent algorithm uses the entire dataset at once. The parameter γ (momentum) determines the contribution of the previous gradient step to the current iteration. Two major problems for the training are overfitting and underfitting. Underfitting occurs when the neural network cannot capture the characteristics of training sets and cannot fit the target mappings well, leading to low accuracy rates and high loss values for all three datasets [40]. While underfitting happens when the NNs are overly dependent on training sets and learn wrong mappings which work well in training sets but perform poorly in validation or test sets [40]. Adding the regularisation term for the weights to the loss function $E(\theta)$ is one way to reduce overfitting [41].

The loss function with the regularisation term takes the form (2):

$$ER(\theta) = E(\theta) + \lambda Z(w), \tag{2}$$

where w is the weight vector, and λ is the regularisation factor (coefficient), and the regularisation function $Z(w)$ is (3):

$$Z(w) = \frac{1}{2} w^T w. \tag{3}$$

There are a variety of hyperparameters that affect the training performance using this method. These include the proportion between the data used for training and validation, the number of epochs, the importance of each class, the momentum and regularisation. All of these are explained in Li Yang et al.

[28], with more details on hyperparameter tuning provided in to achieve better segmentation results. The method we followed in tuning the hyperparameters of the neural network was based on a trial-and-error approach. We fine-tuned the network based on the momentum, learning rate, regularisation, number of epochs and validation patience. For instance, the way the learning rate affects the training is that when it has a large value, the training can finish much faster, but at the same time, the gradient fluctuates around the different values or even cannot converge. On the other hand, a small learning rate will increase the training time but will converge better. So, an ideal learning rate will activate the loss function to converge at a global minimum for an affordable training time. Moreover, an extensive analysis of the trade-offs between the learning rate and the convergence can be found at the study of Gupta et al. [42].

Another way to improve the training of a neural network is the manual calibration of the class weights. The manual calibration of weights allows the user to specify each class's relative importance and helps improve the overall performance of the final segmentation analysis. Using different weights for different classes in phase classification can help to address the class imbalance, where some classes have more samples than others. In our study, we aimed to calibrate the alpha phase weights to improve the semantic segmentation results as it is most critical class for the intended final measurement. For our intended purpose of measuring alpha grain size and volume fraction misclassification between beta phase and alpha grain boundaries, it will have minimal impact on the final measurement, while misclassifications of the alpha class would be more significant. In order to determine the optimal model for the semantic segmentation analysis, we compared the experimental results with the ground-truth dataset, assessing the accuracy of each class using the true-positive rate score metric. We tested weight factors ranging from 1 to 10 and weights of 50, 100, 200 and 400 to observe their impact on class accuracy, Table 2. As predicted, increasing the importance of one class resulted in a notable improvement in that class's accuracy, while the accuracy of the other two classes decreased. Our experiments concluded at a weight of 400, where the beta class had a true-positive rate of 0.0001, and the alpha class had a true-positive rate of 0.9987, Fig. 10. After testing, we concluded that the best model is the one with alpha weights multiplied by 5. This weight returned a high

Table 2 The accuracy metrics of the class with the corresponding weights

Alpha class weights	Alpha class accuracy	Boundary class accuracy	Lath class accuracy
1	0.8592	0.4642	0.8893
2	0.866	0.4157	0.8876
3	0.8848	0.4394	0.8587
4	0.8854	0.4279	0.8692
5	0.912	0.4144	0.8416
6	0.8875	0.4286	0.8558
7	0.9079	0.4098	0.8455
8	0.8998	0.3908	0.8476
9	0.9172	0.3873	0.8398
10	0.9182	0.3954	0.8269
50	0.9688	0.2394	0.7181
100	0.9869	0.1199	0.6399
200	0.9969	0.0281	0.4779
300	0.9972	0.0074	0.4525
400	0.9987	0.0001	0.4268

level of accuracy in the alpha class, while also maintaining strong accuracy in both the boundary and lath classes. When using weights of 5, the true-positive rate of the boundary class is 0.4223. However, this accuracy is affected by the fact that the boundaries of the ground-truth images are only 3 pixels wide. Hence, even a small change in the position of the boundary or its size in the predicted images can significantly impact their accuracy. Despite the low boundary accuracy, the predicted boundaries provide enough information for the Watershed to enhance or complete the grain boundaries. On the other hand, using unbalanced weights resulted in the highest boundary and lath class accuracy, but at the expense of alpha class accuracy, which was the most important for our experiment. Therefore, we decided to exclude unbalanced weights from our experiments. While the best alpha class accuracy was achieved with weights of 10, this weight resulted in the lowest accuracy in the other two classes. As a result, we had to limit our options and choose the most balanced weight distribution for our experiment. In order to determine the optimal model for the semantic segmentation analysis, we compared the experimental results with the ground-truth dataset. The true-positive rate score is given by the following equation: true-positive rate = $TP / (TP + FN)$, TP is the number of true-positives pixels, and FN is the number of false-negative pixels.

The values of the parameters for our experiments are given below. We set the initial learning rate for the model to 0.001, learning rate drop fact was set to 0.3 every 10 epochs, the validation patience was set to 20, the number of epochs was set to 25, the momentum to 0.99 and the regularisation to 0.001. All the values that we used for the hyperparameters in our experiments are shown in Table 1. While these settings gave optimal results in our trial, it is expected that other networks and training regimes will perform similarly well.

Data analysis and subjectivity

Manual labelling of training data in deep learning is often a trivial task. For instance, in applications, of self-driving cars, there is rarely ambiguity over what is and is not a car. However, microstructures are often far more subjective. Consider the example in Fig. 11 which shows manual labels produced by different scientists, both with similar training. There is an observable difference in the level of segmentation with no certainty of which is correct. The arrows show an example of subjectivity in the labelling process, while in the red squared area, some grains are missing from the second analysis. This can be an issue as the neural network may learn this subjective opinion during the training, which consequently may affect our ability to quantify accuracy.

Differences between manual and automatic digital segmentations can be measured with BF score (boundary F1-measure) [41], shown in Eq. (4).

$$F1 = \frac{2TP}{2TP + FN + FP} \quad (4)$$

where TP = true-positive prediction, FP = false-positive prediction and FN = false negative. A true-positive prediction is when a specific pixel of the predicted image belongs to the segmented grain of the ground-truth image.

BF score is a relation between precision [41] and recall [41], which are accuracy metrics of detection or segmentation. We measure BF score by overlapping the segmented image over the ground-truth image and identify the matching area of the grains in the two images measuring this way the accuracy of the grain segmentation.

Precision and recall give as information about how many grains have been segmented correctly compared to the predicted grains and the total true grains, respectively.

Results

The novel *HADMA* approach is validated using the public dataset recently published by [20]. We selected 172 images from this dataset that shown bi-modal

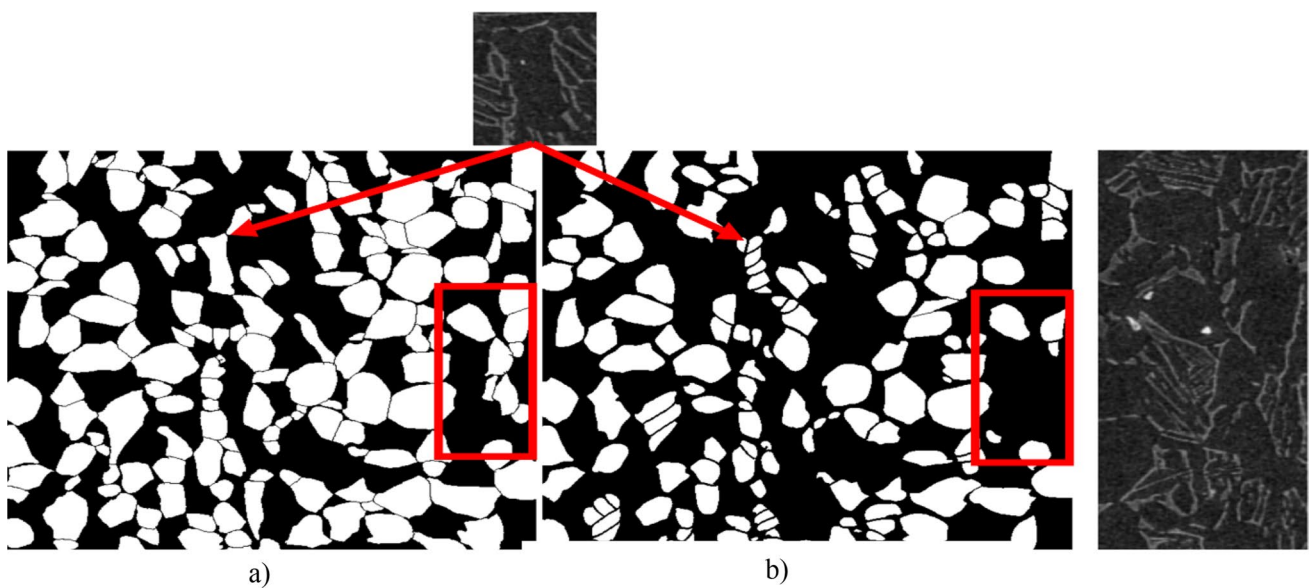


Figure 11 Subjectivity in manual ground-truth labelling where ground-truth images are produced by different scientists.

microstructures with delineable grains. A total of 122 are used for training and 50 for validation. Also, in our ground-truth data, every grain is measured to ensure accurate results; however, the inter-operator variability of 16% within ASTM standard is based on statistics and recommends counting at least 10 grains per linear intercept on a random section or at least 100 grains per field of view when using the planimetric. Thus, while we know that our results are accurate, this variability between users is higher as any uncertain grain boundaries and regions become more statistically significant. An example of images of this dataset is shown in Fig. 12a and d.

The processing methods used by the BCSS results in single segmentation and can be used to create visualisation of the microstructural analysis, as shown in Fig. 12b and e. While the greatest benefit of this approach is the quantitative measurement, this enables that these visualisations are also helpful for further visual inspection of the microstructure like the complete detection of the grains. Those methods

can also increase the confidence in the quantitative measures by demonstrating how each measurement is derived. Two key microstructural features were measured in our experiments: mean grain size and the volume fraction of the primary alpha phase. These measurements were compared to ground-truth measurements produced by expert materials scientists, following the ASTM E112 [2] and E562 [3] standards. A key consideration during this study is that no method is expected to achieve 100% accuracy as the ground-truth is uncertain. Therefore, any difference between measurements using the HADMA and manual results from the user less than 16% is within the expected variation [2], and either result may be correct. Any result with a difference exceeding the 16% must be considered as erroneous from the automated method. This approach has also been used in previously published methods [7, 43]. A similar threshold of 10% is set for comparison of the volume fraction, as per E562 standards [3].

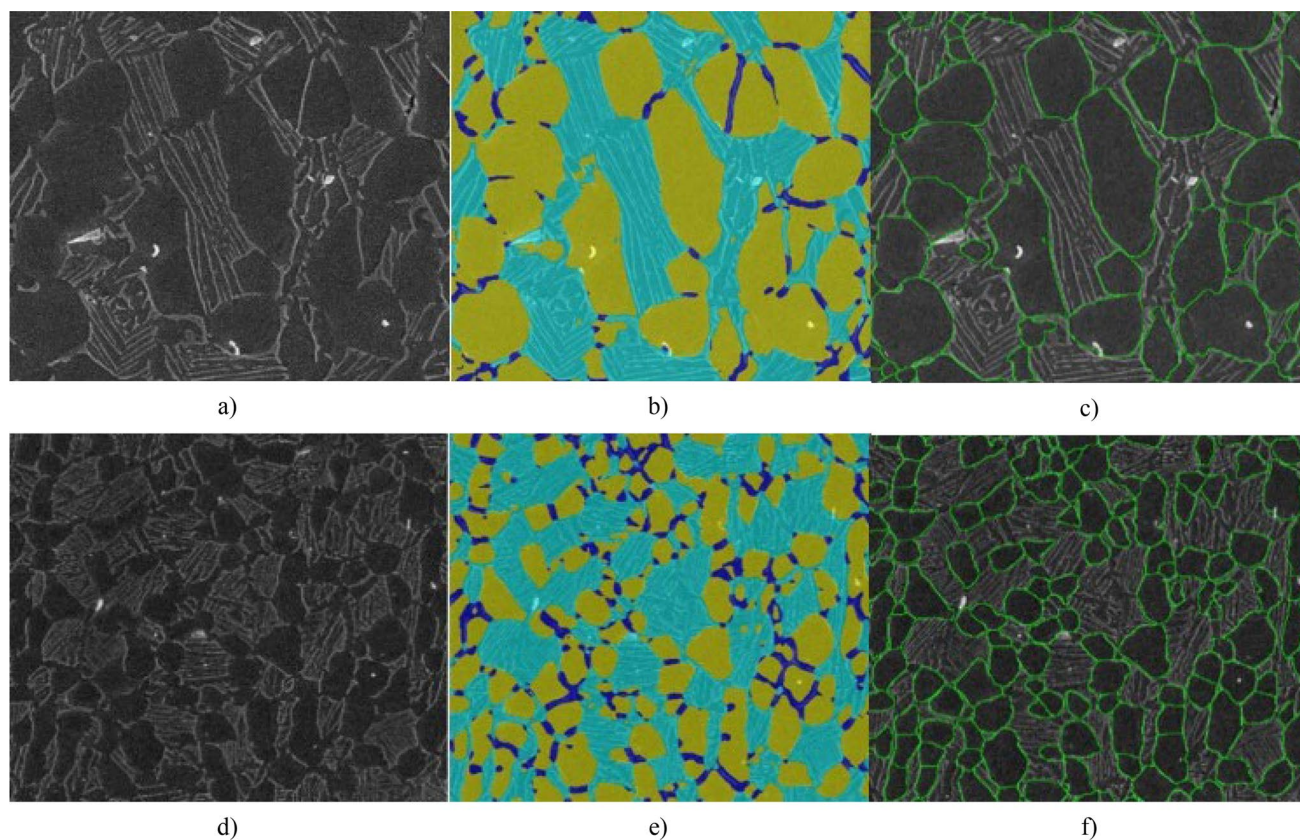


Figure 12 Examples of images from the validation dataset and visualisation of results where **a** and **d** original SEM images, **b** and **e** phase separation after semantic segmentation while **c** and **f** a segmentation of each grain measured by the algorithm.

We also compare the produced results from the HADMA and a recently published method [7] for Ti6Al4V bimodal microstructure. A short study using an additional data is also included to evaluate the generalisability of the proposed technique.

Grain size and globular volume fraction measurements

The difference between the mean grain sizes measured by the HADMA and from existing standards, for each microstructure in the test dataset, is plotted in Fig. 13.

Results from the HADMA typically fall comfortably below the 16% threshold of the inter-operator repeatability, meeting the standard of accuracy set in the E112 standards. There are only two micrographs (14 and 16) that the deviation between the HADMA versus the scientist exceeded the acceptable threshold. The smaller grains that the HADMA detects reducing the measurement of the average grain size compared to the scientist. It is clear that the predicted image captured more details. This is an outcome of the subjectivity that explained on “Neural Network Training and Evaluation” Section. Also, in the training set, there are images which show fewer coarse grains altering the number of the grains per image.

For ASTM E112, a minimum of 40–50 grains is required but 500 is recommended [2]. Our dataset allows the minimum conditions to be met but not

always the recommended. Thus, the fact that the algorithm still performs well is more impressive. For instance, Fig. 12a and d illustrates the aforementioned differences.

To measure the average grain size of the tested images, we used the below function:

$$AGV = \frac{\sum_{i=1}^n (L(i) + l(i))}{2 \times n}, \tag{6}$$

where the AGV = average grain size, L = major axis length which is the longest straight line that fits inside the grains, l = minor axis length which is the maximum length inside the grain that is perpendicular to the major axis length and n = number of the detected grains. Both minor and major axes are crossing the centre of the grains as both have normalised second central moments as the grain.

The units we used to measure distance are pixel units.

Similar results were achieved for the globular volume fraction when we compared it to the ground-truth image, as shown in Fig. 14. To compute the VF (volume fraction), we used the follow equation:

$$VF = \frac{TPG}{TP}, \tag{7}$$

where VF is the volume fraction of the grains, TPG = total number of pixels that belong to the alpha

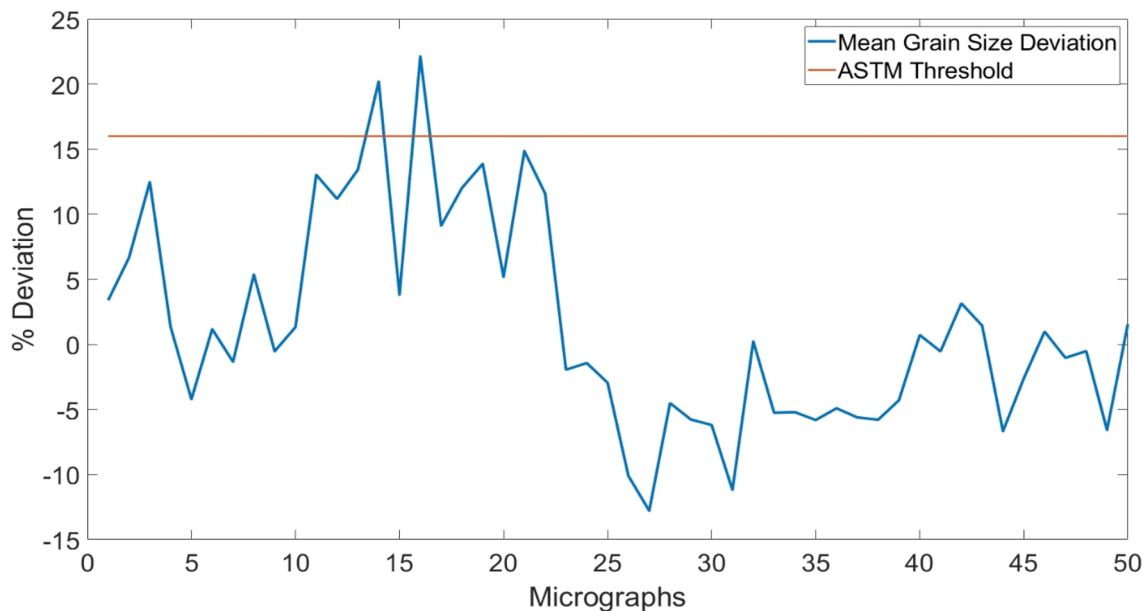


Figure 13 Chart of % deviation of ground-truth and HADMA for the tested micrographs.

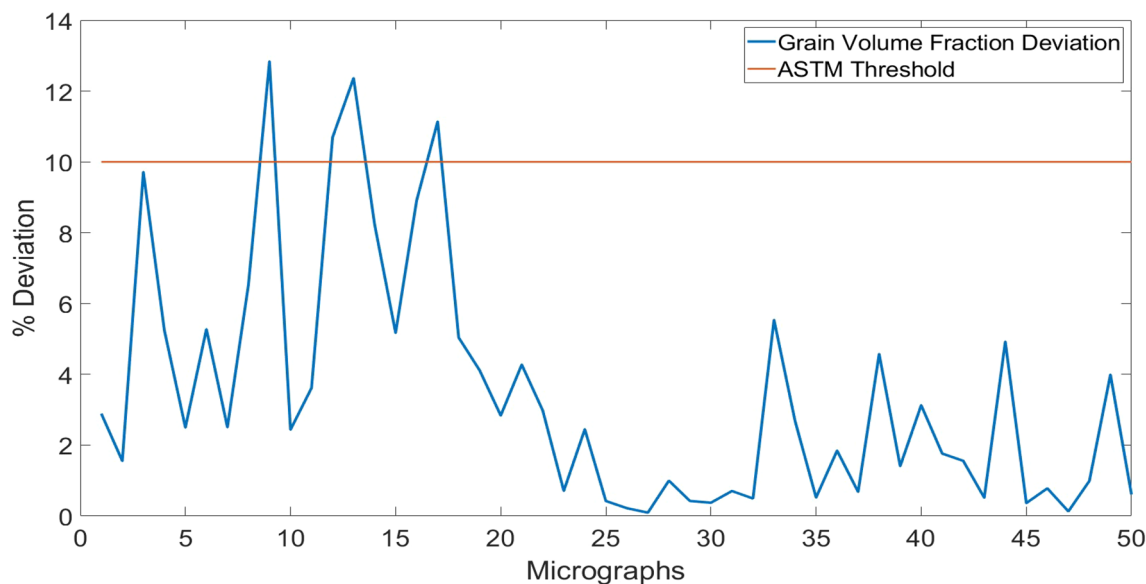


Figure 14 Chart of volume fraction % deviation of the ground-truth and HADMA for the tested micrographs.

Table 3 Accuracy metrics between HADMA and ground-truth

Original image	BF score (%)	Precision (%)	Recall (%)
Micrograph 13	82.43	76.89	88.83
Micrograph 16	81.05	74.71	88.57
Micrograph 17	76.96	72.06	82.58
Micrograph 21	82.49	77.47	88.21

class and TP = total image pixels. Only four of the images that we used for testing lie above the ASTM threshold of 10%. With the worst-case result reaching only 2.8% above the expected threshold, the reason for these differences is due to subjectivity and the ability to perceive finer grains. In the manual labelling process, if a few pixels were included or excluded when the ground-truth image was produced could alter the less than 3% deviation. For instance, the differences in the perimeter size of the grains have been identified from the scientist versus the algorithm. This is reflected at the accuracy metrics as shown in Table 3, where recall is higher than BF score, this means that model's true-positive rate and false-negative rate are higher and lower, respectively, compared to its precision (which measures the proportion of predicted positive cases that are actually positive) and false-positive rate (which measures the proportion of predicted negative cases that are actually positive).

Comparison with existing method

Furthermore, the *HADMA* is compared with the method [7] as is illustrated in Fig. 16. The suggested hybrid analysis *HADMA* improved the method [7] significantly. Firstly, the existing method required some manual parameterisation, which limits its robustness to variations in the dataset. These optimal parameters were selected empirically using the first image in the dataset (as suggested in prior work [7]), but the larger and more diverse dataset meant that the results used remained inconsistent compared to the hybrid approach. So, compared to the *HADMA*, the previously developed method has a more limited ability to interpret the challenging morphology of bimodal microstructure of Ti6Al4V. This leads to large errors in certain images, like in Fig. 16e, some areas were considered as a single grain instead of four. While there are areas where the lath colonies were erroneously measured as primary alpha phase, Fig. 16b. The benefit of *HADMA* is that it requires no previous parameterisation and had a good ability to distinguish phases. Moreover, the algorithm managed to classify accurately areas that were considered as laths, Fig. 15c. Other than these evident differences it also appeared that the use of deep learning to perform an initial estimation of grain boundaries typically led to more accurate separation of the grains. This is illustrated in Fig. 16f,

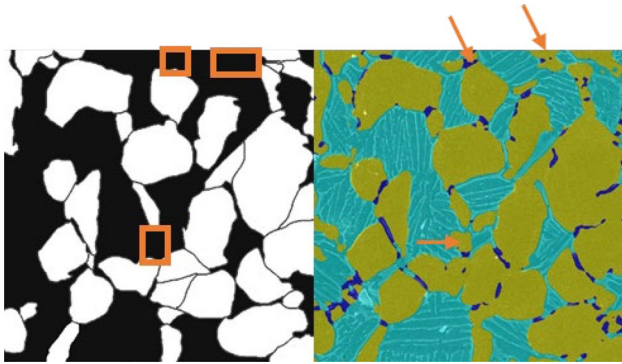


Figure 15 Differences between ground-truth and predicted image from BCSS. The yellow arrows show the finer grains that are detected in the predicted image and the rectangles the missing grains in the ground-truth image.

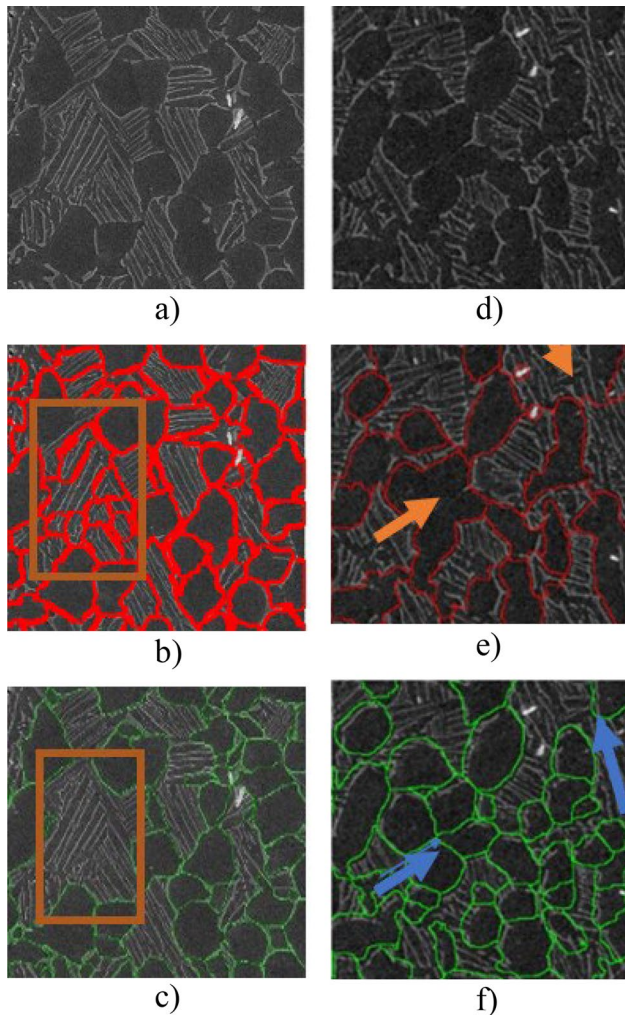


Figure 16 a, d the original images, b, e the Watershed [7] and c, f HADMA.

Table 4 Mean grain size and equiaxed alpha volume fraction comparison of the aforementioned techniques

Original image	Ground-truth (μm)	HADMA (μm)	Watershed-based method [7] (μm)
Micrograph 01	9.37	9.05	7.32
Micrograph 03	9.91	8.67	7.42
Micrograph 31	4.89	5.43	5.99

Table 5 Primary alpha phase volume fraction

Original image	Ground-truth (%)	HDMA (%)	Watershed-based method [7] (%)
Micrograph 01	58.78	57.02	71.9
Micrograph 03	44.39	48.64	70.87
Micrograph 31	43.78	43.54	79.41

Table 6 Error illustration of the mean, std and min and max for the “bimodal” test set

	Mean error	Std error	Min error	Max error
Grain size	0.208 μm	0.423 μm	0.012 μm	2.027 μm
Volume fraction	0.0124%	0.004%	0.00042%	0.0543%

which illustrates an example where touching alpha grains were separated significantly better adopting the new approach. This arises as previous approach [7] requires some section of the grain border to be delineable, while the HADMA was able to interpret this utilising the entire morphology of the grains predicting parts of the grains that are not well defined in the image. The differences between the two methods are reflected in Table 4 which shows the mean grain size comparison for the ground-truth, HADMA and the Watershed approach [7].

Table 5 shows the deviation of the volume fraction for the two methods and the ground-truth image. Again, the same micrographs have been used.

Both tables show that the HADMA has similar results with the ground-truth images compared to the method [7]. The maximum deviation for the mean grain size is 12.51% for HADMA and 25% for the existed method [7]. While the maximum deviation for the volume fraction for HADMA and method [7] is 8.73% and 44.86%, respectively. Table 6 shows the mean, standard deviation and min and max error for the “bimodal” test set.

Applicability to different microstructural morphologies of Ti6Al4V

It has been demonstrated that the HADMA approach is successful for datasets similar to those for which the method was developed. However, the microstructure of materials and the resulting micrographs are known to exhibit significant variations [15, 44], particularly in the case of Ti6Al4V. This has been previously cited as a key reason for deep learning methods of image identification being less efficiently applied in materials science, as compared to fields such as medicine [13].

To validate the approach on other microstructures of Ti6Al4V, we acquired microstructural images of two other microstructural types. The types of microstructures we used consisted of elongated and equiaxed alpha grains, illustrated in Fig. 17a and b, respectively. Due to morphological differences between the Ti6Al4V microstructures, the first model (M1) that we trained on the bimodal dataset did not perform well when tested it on the new starkly different data. To resolve this issue, we trained two different models, one for each of the new datasets,

to analyse microstructures with elongated and equiaxed grains (M1, M2 and M3 refer to the trained models on the “bimodal”, “elongated” and equiaxed datasets, respectively). For each microstructure, an area was imaged containing approximately 2583 elongated and 2000 equiaxed alpha grains.

However, for the “elongated” dataset, this was achieved by capturing 43 images with approximately 63 grains per images, whereas with the “equiaxed”, this was achieved using only two images, but with 1000 grains per image. The original size of the images for both datasets is 2048×1887 pixels, which resulted in a significant resolution difference in terms of the number of pixels per grain. For the data population, we cropped the images following the same approach as described in 4.1. For the “elongated” dataset, we reserved two images for testing while the remaining 41 were cropped randomly 100 times creating a training–validation set of 4100 images. Whereas for the “equiaxed” dataset, we reserved one image for testing and one image for training. Thus, we selected to randomly crop a single-image 400 times. For both datasets, we tested a various number of cropping

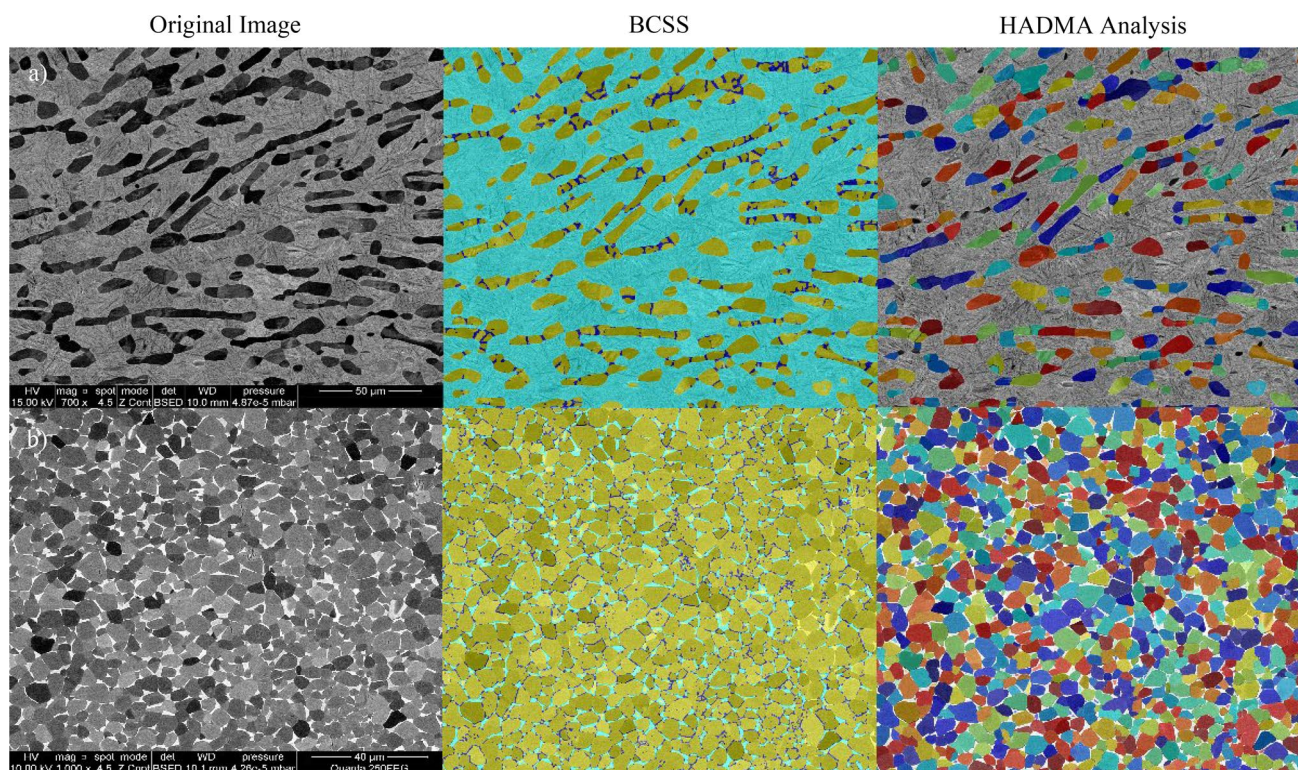


Figure 17 **a** Original image of Ti6Al4V taken from SEM, **b** boundary class semantic segmentation. The top row of the images belongs in the elongated dataset and the bottom row to

the equiaxed dataset. The three columns of the figure from left to right depict the original images, the BCSS and HADMA.

Table 7 Comparison of classes' accuracy score versus number of crops for the elongated grains

Number of crops	Alpha class accuracy	Boundary class accuracy	Beta class accuracy
Equally cropped (72)	0.9785	0.2371	0.733
100	0.9347	0.0875	0.0081
200	0.9874	0.0244	0.0033
300	0.9688	0.123	0.7122

where are found in Tables 7 and 8. We chose those numbers as they presented a slightly better accuracy score for all classes.

For the training purposes, we split the aforementioned training and validation data 95% for training and 5% for validation. The hyperparameters that we selected to train M2 and M3 were refined using the empirical approach described in 4.1. We chose to balance the weights like in the first model (M1) to make alpha class 2 times more important than the other two classes (boundaries and beta class). The training options are based on Table 1. For both M2 and M3, the optimal hyperparameters were found to be the initial learning rate for the model to 0.001, the validation patience to 20, the number of epochs to 25, the momentum to 0.99 and the regularisation to 0.001.

The numbers in the below tables for the elongated grains refer to the average values for the two test images. While the numbers for the equiaxed grains refer to the one available image.

Table 9 shows, the mean grain size results, which are similar between the *HADMA*, the previously developed method and the ground-truth image. The deviation between the results for the volume fraction of the primary alpha phase is shown in Table 10 and lie below than 10%, which follows the international standards [3].

The results have less than 1% deviation in all metrics for models M2 and M3 (Tables 11 and 12) which justify our choice for training two different neural networks for the new datasets. Also, this is a proof of how the topography of the microstructure affects the segmentation especially for those microstructures that are evaluated as "easy". Where phases create similar morphological objects, conventional algorithms work similarly well like the approach of deep learning. But, as it was mentioned before, what is the time-cost to develop or parameterise such algorithms? While with the *HADMA* adaptability, we showed that sometimes

Table 8 Comparison of classes' accuracy score versus number of crops for the equiaxed grains

Number of crops	Alpha class accuracy	Boundary class accuracy	Beta class accuracy
Equally cropped (63)	0.9785	0.2371	0.733
100	0.9347	0.0875	0.0081
200	0.9874	0.0244	0.0033
300	0.9688	0.123	0.7122
400	0.9526	0.4072	0.9164
500	0.9619	0.3949	0.4996
600	1	0.0001	0.0002

Table 9 Mean grain size, comparison between the *HADMA*, existing method and materials experts

Original image	Ground-truth (μm)	<i>HADMA</i> (μm)	Watershed-based method [7] (μm)
Elongated alpha	7.47	7.2	7.38
Equiaxed alpha	4.086	4.34	4.55

Table 10 Primary alpha phase volume fraction

Original image	Ground-truth (%)	<i>HADMA</i> (%)	Watershed-based method [7] (%)
Elongated alpha	25.51	27.8	25.6
Equiaxed alpha	86.91	84.87	87.61

even with one image and suitable processing, Fig. 17b shows enough to train a deep neural network for phase segmentation and get totally acceptable results.

Conclusion

A novel approach of hybrid automated digital microstructural analysis, *HADMA*, has been developed which combines deep learning-based networks for semantic segmentation networks with deterministic segmentation methods based on the Watershed Transform. This method incorporates a new strategy for deploying semantic segmentation networks by training the network to identify boundary class, which provides additional information to guide grain segmentation and give consistently reliable results.

Table 11 The results for BF score, precision and recall for the images in Figs. 17a and b with the HADMA approach

Original image	BF score (%)	Precision (%)	Recall (%)
Elongated alpha/HADMA(M2)	98.13	98.66	97.6
Elongated alpha/method [7]	97.28	97.53	97.02
Equiaxed alpha/HADMA(M3)	97.74	97.1	98.39
Equiaxed alpha/method [7]	97.47	97.07	97.88

Table 12 The results for BF score, precision and recall for the images in Figs. 17a and b compared to the trained models, M1 versus M2 and M1 versus M3 for elongated and equiaxed, respectively

	Elongated		Equiaxed	
	HADMA (M1) (%)	HADMA (M2) (%)	HADMA (M1) (%)	HADMA (M3) (%)
BF score	70.03	98.13	72.69	97.74
Precision	95.88	98.66	92.64	97.1
Recall	55.15	97.6	59.82	98.39

The approach is validated by multiple experimental trials to assess both measurement accuracy and the ease of adapting the technique to new datasets.

Firstly, a study with a relatively large dataset of 172 micrographs is presented with *HADMA* used to measure grain size and globular volume fraction. Results matched ground-truth measurements from published standards in 93.88% of samples for the mean grain size analysis and 91.2% for the volume fraction of the Ti6Al4V bimodal microstructure. Where this standard was not met, the deviation in measurement is small and is explainable by the subjectivity in these micrographs. Furthermore, our method produces useful visualisations of the segmentation that correlates with measurements to increase confidence and confirm that all results are realistic of what an expert materials scientist may measure.

Secondly, an additional study was performed with smaller and more varied datasets. Training and test data are considerably limited in this trial, so while this later trial is not optimal for assessing the ultimate performance of the *HADMA*, it is very useful for assessing how well the method adapts to other datasets and microstructural problems. While some re-training of the network was required, it is demonstrated that re-training on small dataset can allow the *HADMA* to be adapted to accurately measure different microstructures. This adaptability is

demonstrated by good results achieved on both equiaxed and elongated microstructures. Compared to an existing method to automate analysis of such microstructures [7], which requires manual parameterisation, it was found that the *HADMA* adapted equally well without the need for this additional step.

Based on this later study, we believe that *HADMA* can be used with other Ti or non-Ti alloys. However, our findings suggest a few factors that should be considered when attempting this.

The new material must be allotropic and present microstructural similarities in terms of shape, texture and colour. Despite this requirement, the proposed hybrid pipeline of *HADMA* can be generalised, as it is possible to train a new neural network for a completely new material and still achieve accurate segmentation results by following the proposed hybrid analysis approach. There are several other factors that could influence the microstructural analysis of new materials, such as the material type (metal or non-metal), number of phases, balanced or unbalanced classes in the dataset, types of microscopes used (LOM, SEM, etc.), image capturing (external or integrated camera in the microscope, etc.) and lighting conditions. Therefore, we suggest defining the number of phases in the new material to determine the number of classes and experimenting with a variety of weights for balancing the phases in the dataset. Hence, a transfer learning approach also it is advised especially from the pre-trained weights of a model that was trained on a similar dataset. Additionally, to improve the approach's applicability to other materials, we recommend exploring the potential of a cyclical learning rate [45]. Method [45] eliminates the need for manual experimentation to determine the best global learning rates and schedules. Rather than continuously decreasing the learning rate, this approach allows for a cyclic variation between optimal values. By utilising cyclical learning rates instead of static values, classification accuracy is improved and often in fewer training iterations, eliminating the need for constant tuning.

The foremost advantages of HADMA are the adaptability and the minimisation of manual effort needed, fast measurement times and the repeatability of measurements inherent in automated digital analysis approaches. The HADMA can typically be redeployed between different microstructural datasets using a parameterisation free approach, requiring no input beyond loading the proper image to obtain the same output every time. Also, ASTM grain standards such as E112 [2] need roughly 15 min per picture of 60–70 grains, as reported in [7] and validated when obtaining ground-truth data for the current investigation. In comparison, our approach needed a mean processing time of 0.41 s per picture for images in this investigation. This measurement is based on the average time of the two machines; the first with an i7 8700 k, 16 GB RAM with a 1080Ti GPU and the second with a Ryzen7 5800 × 8 CORE, 16 GB RAM with RTX 3070 GPU. This is a level of technology that is accessible for the most modern laboratories. Furthermore, process time consistent primarily of the time required to segment the grains; therefore, each additional measurement computed from the result is negligible. This is not typically true for existing manually driving analysis approaches. Alongside the absence of the need to interact with the measurement system, this makes the HADMA substantially more scalable than many existing approaches. Furthermore, this also gives a baseline from which to further enhance these approaches.

Acknowledgements

The authors gratefully acknowledge the help of Titanium Metals Corporation (*TIMET*) and Dr Dorothy Evans NMIS Industrial Doctorate Centre Manager.

Author contributions

GF worked in resources, data curation, methodology, original draft, writing—original draft preparation, writing—reviewing, software, validation, visualisation, conceptualisation and investigation. AC helped in conceptualisation, resources, writing—reviewing and editing, software, supervision, validation, visualisation and investigation. PM helped in conceptualisation, writing—reviewing and editing and supervision.

EY helped in conceptualisation, writing—reviewing and editing and supervision.

Data availability

The dataset that has been used in this study is publicly available from A. Baskaran and can be found here: <https://github.com/ArunBaskaran/Image-Driven-Machine-Learning-Approach-for-Microstructure-Classification-and-Segmentation-Ti-6Al-4V>. Also, the ground-truth dataset can be found here: <https://doi.org/10.15129/46955351-408b-4dc3-840e-3bc6a9f3432a>.

Declarations

Conflict of interest This manuscript has not been published and is not under consideration for publication elsewhere. We have no conflicts of interest to disclose.

Ethical approval Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- [1] Vajpai SK, Ota M, Watanabe T, Maeda R, Sekiguchi T, Kusaka T, Ameyama K (2015) The development of high performance Ti-6Al-4V alloy via a unique microstructural

- design with bimodal grain size distribution. *Metall Mater Trans A Phys Metall Mater Sci* 46:903–914
- [2] Astm Standard (2012) E112-12:Standard test methods for determining average grain size. ASTM International E112-12
- [3] E562 A (2000) ASTM E562. Refractories
- [4] Soille P, Pesaresi M, Ouzounis GK (2006) Mathematical morphology and its applications to image and signal processing
- [5] Meyer F, Beucher S (1990) Morphological segmentation. *J Vis Commun Image Represent* 1:21–46
- [6] Hušek M (1989) Categories and mathematical morphology. In: Ehrig H, Herrlich H, Kreowski H-J, Preuß G (eds) *Categorical methods in computer science with aspects from topology*. Springer, Berlin, pp 294–301
- [7] Campbell A, Murray P, Yakushina E, Marshall S, Ion W (2018) New methods for automatic quantification of microstructural features using digital image processing. *Mater Des* 141:395–406
- [8] Collins PC, Welk B, Searles T, Tiley J, Russ JC, Fraser HL (2009) Development of methods for the quantification of microstructural features in $\alpha + \beta$ -processed α/β titanium alloys. *Mater Sci Eng A* 508:174–182
- [9] Biswal SR, Sahoo T, Sahoo S et al (2009) Automatic grain size determination in microstructures using image processing. *Mater Sci Eng A* 46:1431–1438
- [10] Biswal SR, Sahoo T, Sahoo S (2021) Prediction of grain boundary of a composite microstructure using digital image processing: a comparative study. *Mater Today Proc* 41:357–362
- [11] Fujiyoshi H, Hirakawa T, Yamashita T (2019) Deep learning-based image recognition for autonomous driving. *IATSS Res* 43:244–252
- [12] Piccialli F, Di SV, Giampaolo F, Cuomo S, Fortino G (2021) A survey on deep learning in medicine: why, how and when? *Inf Fusion* 66:111–137
- [13] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B, Sánchez CI (2017) A survey on deep learning in medical image analysis. *Med Image Anal* 42:60–88
- [14] Wang J, Zhu H, Wang SH, Zhang YD (2021) A review of deep learning on medical image analysis. *Mobile Netw Appl* 26:351–380
- [15] Dimiduk DM, Holm EA, Niezgodá SR (2018) Perspectives on the impact of machine learning, deep learning, and artificial intelligence on materials, processes, and structures engineering. *Integr Mater Manuf Innov* 7:157–172
- [16] Plath N, Toussaint M, Nakajima S (2009) Multi-class image segmentation using conditional random fields and global classification. In: *Proceedings of the 26th international conference on machine learning, ICML 2009* 817–824
- [17] Chowdhury A, Kautz E, Yener B, Lewis D (2016) Image driven machine learning methods for microstructure recognition. *Comput Mater Sci* 123:176–187
- [18] Azimi SM, Britz D, Engstler M, Fritz M, Mücklich F (2018) Advanced steel microstructural classification by deep learning methods. *Sci Rep* 8:1–14
- [19] Jang J, Van D, Jang H, Baik DH, Yoo SD, Park J, Mhin S, Mazumder J, Lee SH (2020) Residual neural network-based fully convolutional network for microstructure segmentation. *Sci Technol Weld Joining* 25:282–289
- [20] Baskaran A, Kane G, Biggs K, Hull R, Lewis D (2020) Adaptive characterization of microstructure dataset using a two stage machine learning approach. *Comput Mater Sci*. <https://doi.org/10.1016/j.commatsci.2020.109593>
- [21] He L, Chao Y, Suzuki K, Wu K (2009) Fast connected-component labeling. *Pattern Recognit* 42:1977–1987
- [22] Campbell A, Murray P, Yakushina E, Borocco A, Dokladal P, Ion W Automated analysis of platelet microstructures using a feature length orientation space
- [23] Fu KS, Mui JK (1981) A survey on image segmentation. *Pattern Recognit* 13:3–16
- [24] Gonzal RCREW (1988) *Digital image processing (second edition)*. Opt Lasers Eng 8:70–71
- [25] Zaitoun NM, Aqel MJ (2015) Survey on image segmentation techniques. *Procedia Comput Sci* 65:797–806
- [26] Konovalenko I, Maruschak P, Brezinová J, Viňáš J, Brezina J (2020) Steel surface defect classification using deep residual neural network. *Metals (Basel)* 10:1–15
- [27] Olaf Ronneberger, Philipp Fischer and TB (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9351:12–20
- [28] Yang L, Shami A (2020) On hyperparameter optimization of machine learning algorithms: theory and practice. *Neurocomputing* 415:295–316
- [29] Zeiler MD, Fergus R (2014) Visualizing and understanding convolutional networks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8689 LNCS:818–833
- [30] He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: *Proceedings of the IEEE computer society conference on computer vision and pattern recognition 2016-Decem*: 770–778
- [31] Abratenko P, Alrashed M, An R et al (2021) Semantic segmentation with a sparse convolutional neural network

- for event reconstruction in MicroBooNE. *Phys Rev D* 103:52012
- [32] Arcelli C, di Baja GS (1988) Finding local maxima in a pseudo-Euclidian distance transform. *Comput Vis Graph Image Process* 43:361–367
- [33] Rosenfeld A, Pfaltz JL (1968) Distance functions on digital pictures. *Pattern Recognit* 1:33–61
- [34] Panigrahi S, Nanda A, Swarnkar T (2021) A survey on transfer learning. *Smart Innov Syst Technol* 194:781–789
- [35] Jain R, Nagrath P, Kataria G, Sirish Kaushik V, Jude Hemanth D (2020) Pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning. *Measurement (Lond)* 165:108046
- [36] Tang H, Cen X (2021) A survey of transfer learning applied in medical image recognition. In: 2021 IEEE international conference on advances in electrical engineering and computer applications, AEECA 2021 94–97
- [37] Stanford Vision Lab (2016) ImageNet Dataset. Stanford Vision Lab, Stanford University
- [38] Karimi D, Warfield SK, Gholipour A (2020) Critical assessment of transfer learning for medical image segmentation with fully convolutional neural networks. 1–11
- [39] Zhang H, Zhang L, Jiang Y (2019) Overfitting and underfitting analysis for deep learning based end-to-end communication systems. In: 2019 11th International conference on wireless communications and signal processing, WCSP 2019. <https://doi.org/10.1109/WCSP.2019.8927876>
- [40] Sokolova M, Lapalme G (2009) A systematic analysis of performance measures for classification tasks. *Inf Process Manag* 45:427–437
- [41] Sasaki Y (2007) The truth of the F-measure. *Teach Tutor mater* 1–5
- [42] Gupta S, Zhang W, Wang F (2015) Model Accuracy and runtime tradeoff in distributed deep learning: a systematic study
- [43] Sun S, Lv W (2016) Microstructure and mechanical properties of TC18 titanium alloy. *Rare Metal Mater Eng* 45:1138–1141
- [44] Danielsson PE (1980) Euclidean distance mapping. *Comput Graph Image Process* 14:227–248
- [45] Smith LN (2017) Cyclical learning rates for training neural networks. In: Proceedings—2017 IEEE winter conference on applications of computer vision, WACV 2017. <https://doi.org/10.1109/WACV.2017.58>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.