

Recognition and Quantification of Dual Phase Titanium Alloy Microstructures Using Convolutional Neural Networks

Ryan Noraas, Vasisht Venkatesh, Luke Rettberg, Nagendra Somanath
Pratt & Whitney Engines Division, United Technologies Corporation, East Hartford, CT 06108, USA

Olusegun Oshin, Michael Giering
United Technologies Research Center, United Technologies Corporation, East Hartford, CT 06108, USA

ABSTRACT

Recent advances in machine learning and image recognition tools/methods are being used to address fundamental challenges in materials engineering, such as the automated extraction of statistical information from dual phase titanium alloy microstructure images to support rapid engineering decision making. Initially, this work was performed by extracting dense layer outputs from a pre-trained convolutional neural network (CNN), running the high dimensional image vectors through a principal component analysis, and fitting a logistic regression model for image classification. K-fold cross validation results reported a mean validation accuracy of 83% over 19 different material pedigrees. Furthermore, it was shown that fine-tuning the pre-trained network was able to improve image classification accuracy by nearly 10% over the baseline. These image classification models were then used to determine and justify statistically equivalent representative volume elements (SERVE). Lastly, a convolutional neural network was trained and validated to make quantitative predictions from a synthetic and real, two-phase image datasets. This paper explores the application of convolutional neural networks for microstructure analysis in the context of aerospace engineering and material quality.

INTRODUCTION

Design and control of material microstructure is important in many industries, in particular aerospace, where critical rotating hardware operate under extreme conditions for extended amounts of time. Microstructure is a complex and major contributing factor in bulk material behavior. Small local variations can sometimes have significant impact on part-level performance or durability. Over the past few decades, access to materials characterization data has grown tremendously, with sources coming from automated optical systems, high resolution scanning electron microscopy (SEM), electron-backscatter diffraction (EBSB), high energy x-ray diffraction (HEXRD), micro-computed tomography, among many others. The increasing size and rate of incoming data necessitates the development and use of automated, high-throughput tools and analysis methods. Another alarming, but often overlooked challenge is the issue of repeatability in quantitative analysis of microstructure data. For instance, a metric as commonplace as grain size: there exists multiple standardized methods of measurement, but round robin lab results across industry rarely agree. To improve material design and engineering decision making under uncertainty, there is a strong need for objective, scalable methods for statistical representations of microstructure.

CURRENT STATE

Conventional standards for microstructure analysis have largely relied on domain expertise, institutional knowledge, and visual conformity to established materials specifications. Across industry it is standard practice to report metrics like average grain and associated area fraction(s). The benefits of such methods include simplicity and interpretability and for many applications, this approach is well-suited. However, user variability in the image segmentation process and slow turnaround times due to manual inspection of each micrograph usually leads to low quantities of low confidence data. It is also easily seen that simple microstructure descriptors are unlikely to be the optimal set because different types of microstructure could be shown to produce identical measures. This is similar to the case of Anscombe's quartet where four different datasets are shown to have identical statistical properties, yet appear very different when graphed.

N-POINT STATISTICS

Another more recent, data analytic approach to microstructure quantification is n-point spatial statistics. Spatial correlations ($n=2$) describe the first-order spatial correlations between the constituent distinct local states in the internal structure of the material [1]. Unlike traditional single point measures, n-point correlations account for the statistical and spatial arrangement of features in the microstructure. This added complexity comes at the expense of interpretability, where the spatial statistics are typically run through dimensional reduction algorithms to enable visualization of images in a low-dimensional latent space (2D or 3D). Due to computational expense, the approach also generally requires upfront segmentation of images, which introduces additional bias as to what features are significant.

CONVOLUTIONAL NEURAL NETWORKS

Convolutional neural networks (CNNs) are finding widespread adoption in the image domain and have enabled great strides in applications such as object detection, image classification and image segmentation [2]. Their success derives from the ability to make use of information at various length scales for pattern recognition. Useful for both classification (discrete class output) and continuous response variable problems, CNNs have quickly become the state-of-the-art method when working with image data.

Many popular CNN architectures consist of three main types of differentiable layers: convolutional, pooling, and fully-connected (dense) [3-6]. These layers are stacked in various, often repeating sequences. During the forward pass, the convolutional layers compute a dot product between their weights and a small, local image region to which they are connected to from the prior layer (known as the receptive field)¹. The convolutional layers are followed by a layer that performs an element-wise activation, often non-linear. The activation layer can be followed by a pooling layer, which performs a down-sampling operation along the spatial width and height dimensions, resulting in an output that is typically half the size of the input layer size, depending on the pooling layer stride length. These sequences of layers are stacked and alternated to learn filters applicable at different length scales (edge, feature, object detectors). Finally, a combination of fully-connected layers may be used to convert the high dimensional representation from the lower layers to the final model output using an activation function appropriate for the learning task. The prediction error is used during backpropagation to update model weights and biases. This training process is iterated until model convergence.

CNN models automatically handle the task of feature engineering and correlation of learned features to the specified response variable; users do not need to manually specify and quantify any features in images, eliminating a significant source of bias. In cases where simple microstructure descriptors are inadequate to predict material response, it can be of exceptional value to be able to automatically identify critical features in materials microstructure. Naturally, the autonomy of the approach requires the use of training techniques to minimize risk for over-fitting, some of which are detailed in this paper.

RESULTS AND DISCUSSION

The goal of this work is to demonstrate the capacity of convolutional neural networks for microstructure image classification and property prediction in Ti-6Al-4V, a workhorse alloy in the aerospace industry. Pre-trained and fine-tuned CNN models were utilized for image classification, RVE assessment, and continuous property prediction. Transfer learning, a technique in which domain knowledge learned from one model acts as a starting point for secondary task, was leveraged to improve model training and reduce risk of overfitting. A pre-trained model like VGG16 [2] is a great starting point for new image-related tasks because it has already learned useful high frequency filters like edge, feature, and object detectors. This domain knowledge is found to be largely transferrable to new image sets, including materials microstructure even though they have little resemblance to the natural images found in the ImageNet dataset.

DATA

Optical microscopy images were collected at approximately 500X magnification for 19 different pedigrees of Ti-6Al-4V material. Pedigrees are primarily distinguished by changes in solution heat treat temperature and cooling rate. Others received additional heat treatment cycles such as stabilization and aging. Over 500 images were captured from each pedigree, collectively producing a dataset with over ten thousand images and associated pedigree class label. Images tiles were extracted from full-resolution images and resized to dimensions 224x224x3 (RGB) for input into the neural network models.

For quantitative prediction of phase volume fraction from images, two datasets were prepared. The first is a synthetic, binary two-phase dataset (5000 images) for which the exact area fractions of white pixels are known precisely. The second, labelled dataset (1100 images) was created using the results of historical image analysis jobs for which primary alpha grains have been segmented out from optical micrographs using image analysis software. The output of the image analysis is the area fraction of primary alpha grains within each image. These datasets are therefore comprised of image-alpha volume fraction pairs.

MICROSTRUCTURE CLASSIFICATION

During initial work, the pre-trained VGG16 model was used solely to extract statistics for input into a logistic regression classifier. The high dimensional dense layer output was extracted for each image in the titanium dataset producing a (N, D) matrix where N is the number of images in the dataset and D is the dimension of the model layer (D=4096 for VGG16). Principal component analysis was used to reduce dimensionality of the images while maximizing variance between components. A K-fold (K=10) cross-validation method was used to estimate validation classification accuracy as a function of the number of principal components. The training sets were scaled prior to input to PCA using the Python sklearn.preprocessing.MinMaxScaler function.

Using mean K-Fold cross-validation accuracy, the optimal number of PCA components was found to be around 20. The mean K-Fold cross-validation accuracy for the logistic regression classifier using 20 PCA components was 88% \pm 2%. Manual review of some of the misclassifications shows that the mistaken microstructure classes are virtually indistinguishable to the trained eye.

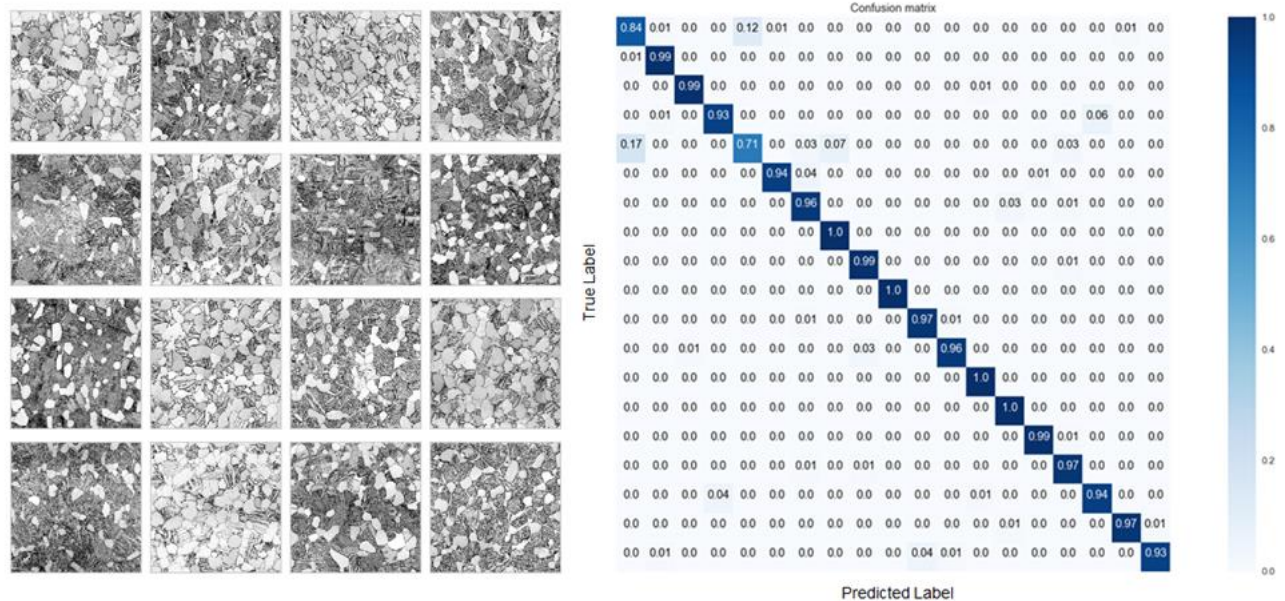


Figure 1. Sample microstructure images and normalized confusion matrix for validation samples

In an attempt to improve classification performance, VGG16 was also fine-tuned on the titanium dataset. This was accomplished by freezing all but the last two dense layers and modifying the dimension of the soft-max output node

to 19. During training, real-time image augmentation was used to randomly flip, rotate, mirror, and modify contrast, brightness, and sharpness in order to reduce risk of overfitting. The learning rate was also reduced by a factor of 4 (default value of $2E-03$ to $5E-04$). The dataset was split 80/20 and the validation accuracy was monitored until it plateaued after which the model training was ceased to prevent over-fitting. With this new approach, the validation accuracy increased to 96%.

Figure 1 shows a random subset of the images used in the training of the model as well as the confusion matrix produced using the results of the K-fold cross validation predictions from the fine-tuned model. A confusion matrix is a visual and quantitative tool to compare classification model accuracy and performance. The values in the cells along the downward diagonal (dark-blue) indicate the accuracy of the model on that particular set or class of images. For example, on row 2, it is observed that the model had 99% validation accuracy for class 2. The model was wrong 1% of the time where it had predicted a class label of 1 instead of 2.

The dense layer outputs were collected for both the pre-trained VGG16 and fine-tuned VGG16 model, run through PCA, and plotted in two dimensions to visualize clustering of images by pedigree. A significant improvement in cluster definition was observed in the fine-tuned model over the pre-trained model, which is in agreement with the improved classification accuracy.

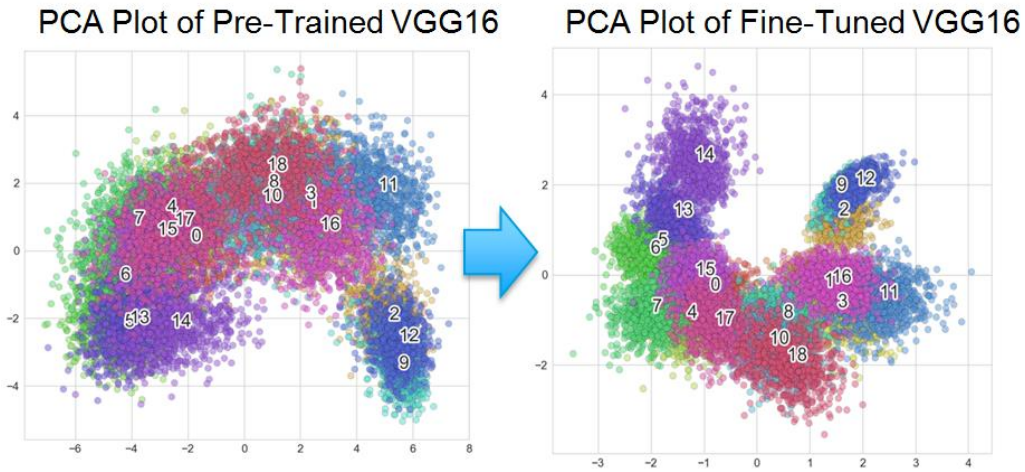


Figure 2. Two-dimensional PCA visualization of VGG16 dense layer output for stock model (left) and a model fine-tuned on microstructures (right).

ASSESSING THE MICROSTRUCTURE RVE

In many applications, it is important to be able to identify the statistically equivalent representative volume element (SERVE), which is the minimum area/volume needed to statistically describe a microstructure. Using the logistic regression model performance as the baseline, we compare logistic regression model cross-validation accuracies as a function of image size using the pipeline described in section 2.2. It was found that small image sizes ($52\mu\text{m} \times 52\mu\text{m}$), -3 sigma cross-validation accuracies were less than 40%. For $104\mu\text{m} \times 104\mu\text{m}$ images, the CV accuracy increases to approximately 70%. At $209\mu\text{m}$ and $419\mu\text{m}$ images, the CV accuracy plateaus around 91-93%, respectively. This exercise suggests that the SERVE for this titanium dataset is approximately $209\mu\text{m} \times 209\mu\text{m}$, below which the classifier is unable to reliably separate the various microstructures.

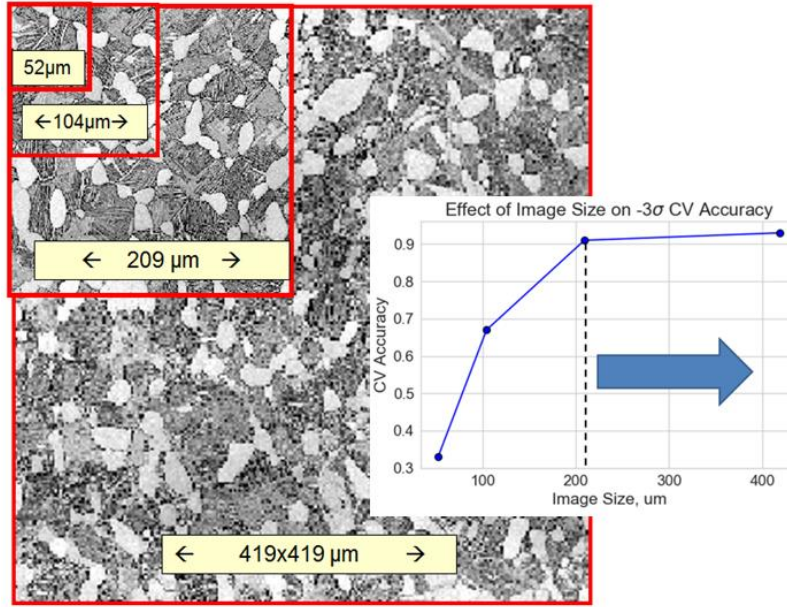


Figure 3. Image classification model performance being used to approximate the microstructure RVE

CONTINUOUS PROPERTY PREDICTION

All the prior work has focused on classification of microstructure images. A natural extension of this work is to infer continuous value properties from image directly e.g. material properties or processing routes –without explicitly quantifying features and building response models.

To accomplish this, the output dimension of the stock VGG16 model was changed to 1 with a linear activation. All the model weights were frozen except for the dense layers and output node. A synthetic, binary two-phase dataset (5000 images) was created for which the exact area fractions of white pixels are able to be calculated. The data was split 80/20 and the VGG16 regression model trained to predict area fraction of white pixels. After 100 epochs, the validation accuracy converged with a mean absolute error of approximately 2.5%. The test RMSE measured on an additional 500 images was less than 0.5% --indicating that the model is able to generalize to new images.

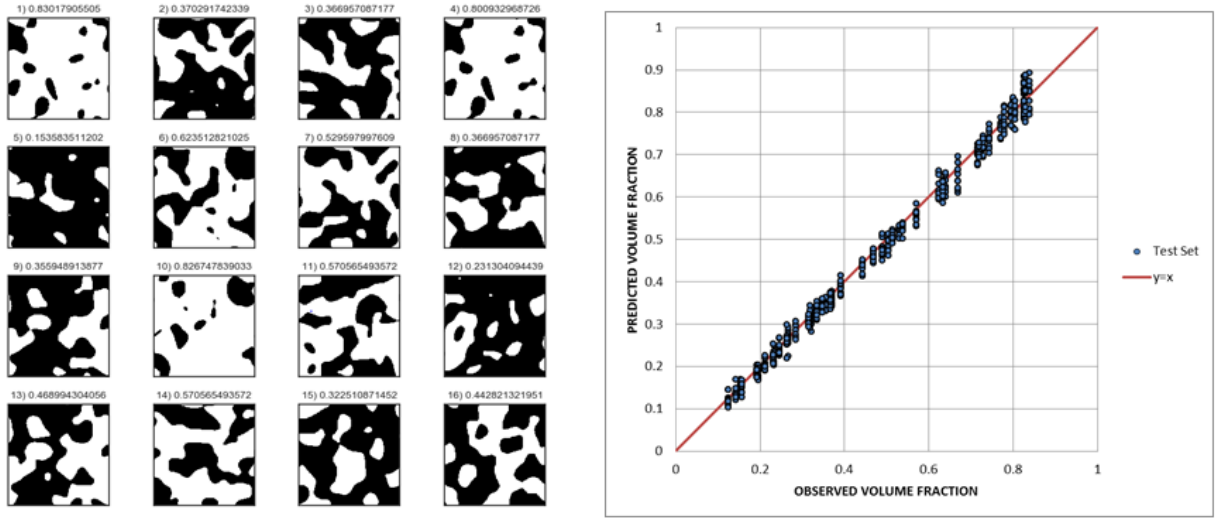


Figure 4. Sample images from the synthetic two phase dataset (left) and the actual versus predicted white pixel area fraction on test set using a convolutional network (right)

This exercise proves that the neural network is able to quantify relative volume fraction of well-controlled, two-phase images i.e. with no noise. The same network was then trained to predict alpha volume fraction in real dual-phase titanium micrographs. The ~1100 image dataset was split 90/10 into training and validation. Training was halted once the model achieved validation mean absolute error below 2% (in terms of volume fraction). The actual vs. predicted results are shown in Figure 5. The mean test error on a test set of 50 images was $1.72 \pm 2.54\%$. With the exception of a few images, the network was able to learn how to identify and predict primary alpha volume fraction without performing any image segmentation.

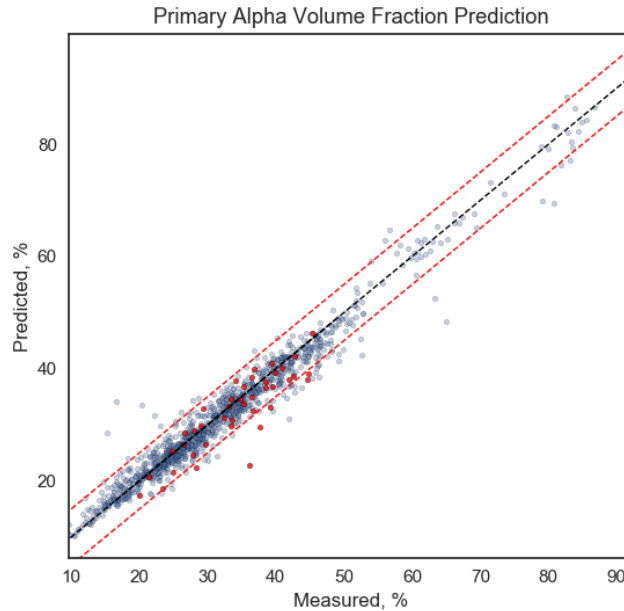


Figure 5. Actual vs. predicted primary alpha volume fractions for alpha-beta titanium microstructures. Blue points are part of the training/validation set. Red points belong to the test set (~50 new images). Dotted lines represent $\pm 5\%$ error bounds.

CONCLUSION

We have introduced and successfully evaluated deep learning convolutional networks for automated microstructure classification, quantification, and SERVE assessment. The outputs of stock convolutional neural networks appear suitable for input into other machine learning algorithms. Additional improvements can be realized via fine-tuning of stock CNN's, although this comes with increased computational expense and risk of over-fitting. The ability of a CNN to predict continuous response values was also demonstrated and validated on synthetic and real alpha-beta microstructure datasets. Automated, high fidelity, scalable methods enable analysis of more data to support engineering decision making. The reduction in variability due to human or institutional bias lends itself to the idea of standardized methods for microstructure analysis in the future. Development of process-structure-property databases and models could enable repeatable results regardless of end user.

ACKNOWLEDGMENTS

The authors thank the support of Pratt & Whitney and United Technologies Research Center for a cohesive and collaborative environment with sustained investment towards this work.

REFERENCES

- [1] Fullwood, W., Niezgoda, S., Kalidindi, S., "Microstructure reconstructions from 2-point statistics using phase-recovery algorithms", *Acta Materialia* 56(5):942:948, 2008.
- [2] Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image Recognition", arXiv preprint arXiv:1409.1556. 2014.
- [3] Gatys, L., Ecker, A.S. and Bethge, M., "Texture synthesis using convolutional neural networks. In *Advances in Neural Information Processing Systems*", (pp. 262-270). 2015
- [4] Lubbers, N., Lookman, T., Barros, K., "Inferring low-dimensional microstructure representations using convolutional neural networks", arXiv:1611.02764v1. 2016
- [5] Larsen, A.B.L, Sonderby, S, Larochelle, H. and Winther, O., "Autoencoding beyond pixels using a learned similarity metric", arXiv:1512.09300v2 [cs.LG] 10 Feb 2016
- [6] LeCun Y, Bengio Y., "Convolutional networks for images, speech, and time series", *The Handbook of Brain Theory and Neural Networks*. Apr; 3361(10):1995.
- [7] Erhan, D., Courville, A., and Bengio, Y., "Understanding representations learned in deep architectures. Department d'Informatique et Recherche Operationnelle, University of Montreal, QC, Canada, Tech. Rep, 1355, 2010.