Neural Networks for the Benchmarking of Detection Algorithms

Silvia Miramontes
mirasilva@lbl.gov
Lawrence Berkeley National Laboratory
University of California, Berkeley
Berkeley, California

Daniela Ushizima (advisor)
dushizima@lbl.gov
Lawrence Berkeley National Laboratory
University of California, Berkeley
Berkeley, California

Figure 1: Evaluation of a fiber detection model: (a) raw data and (b) detected fibers in white from cross-section in the middle of stack; (c) rendering of detected fibers to emphasize discontinuities that are artifact from detection model.

ABSTRACT
There are several automated methods to detect objects from grayscale images. However, materials scientists still lack basic tools to compare different detection results, particularly when working with microtomography. This paper introduces FibCAM, a convolutional neural network (CNN)-based method using TensorFlow that allows benchmarking fiber detection algorithms. Our contribution is three-fold: (a) the design of a computational framework to compare automated fiber detection models with curated datasets through classification; (b) lossless data reduction by embedding prior knowledge into data-driven models; (c) a scheme to decompose computation into embarrassingly parallel processes for future analysis at scale. Our results show how FibCAM classifies different structures, and how it illustrates the material’s composition and frequency distribution of microstructures for improved interpretability of machine learning models. The proposed algorithms support probing the specimen content from gigabyte-sized volumes and enable pinpointing inconsistencies between real structures known a priori and results derived from automated detections.

CCS CONCEPTS
• Computing methodologies → Artificial intelligence; Computer vision; 3D imaging; Object recognition;

KEYWORDS
Computer vision, Materials science, CNN

1 INTRODUCTION
Object detection algorithms often target specific image structures or corresponding polygons, aim to speed up the screening of large image datasets, and identify regions of interest before running more specialized algorithms. For example, a volume cross-section may contain fiber profiles, which are frequently correlated with the presence of elliptical shapes surrounded by high contrast coating. Counting and characterizing fibers are tasks particularly important on the research of ceramic matrix composites (CMCs) [2, 3]. CMCs have special properties for aviation manufacturing: it is a relatively light material that can withstand high temperatures and loads. Shape and structural properties of such compounds are imaged through X-rays, generating microtomography (microCT) images, used for materials quality control [1].

Scientific experiments to stress test CMCs can generate terabyte-sized datasets [7–9], which demand automated algorithms. Such
demand has been met with numerous models for finding "fiber profiles" within microCT cross-sections, also known as fiber detection algorithms \[4, 6, 11\]. One major challenge is to compare these different detection approaches quantitatively, including 3D information. Fig. 1 shows an example in which fiber detection results look appropriate when observed in 2D, however false fiber discontinuities permeate the whole result as shown in the 3D rendering. Another challenge is to process microCT with billions of voxels for discovery of new material configurations, and to achieve autonomous experimentation in many steps of the scientific process.

2 MATERIALS
The X-ray microtomography images acquired in the investigation of the microstructure evolution during the matrix impregnation and curing in unidirectional fiber beds are publicly available at the Materials Facility website \[9\]. These are also fully described in Larson et al \[6\].

FibCAM experiments used thousands of cross-sections from a single fiber-reinforced CMC specimen, specifically two modes: the "raw image stack" and the "ground-truth stack". The first set contains 2,160 image cross-sections from raw microCT data, while the second stack contains 1,000 image cross-sections of the segmented results by \[6\], as \[6\] only includes the segmentation results for image slices (159-1158). We ensure that in our experiments the image slices from the raw stack correspond to the same slices as those in the segmented stack.

3 METHODS
3.1 Specimen Analysis with CNN
In order to properly access inaccurate detections, FibCAM uses manual labeling of a human expert as part of the training and classification of images. Expertise includes the ability to identify the composition of CMC phases and placement of fibers within a specimen. We create two datasets containing image crops of fibers and void spaces via scikit-image \[12\] (also known as fiber bed Fig. 2). The baseline dataset is derived from the work of a human expert, containing 8,814 images.

The segmented crops dataset is composed of 9,195,819 image crops generated following the results of the fiber model in \[9\], which may or may not contain inaccurate detections. To successfully compare the expert’s manual labeling with the segmented crops dataset, we train our LeNet-5 \[10\] based CNN with the baseline dataset, targeting a binary classification problem ("fiber" vs "non-fiber"). With a 70% - 30% train-test split we obtain an accuracy
of 95.99% and a test loss of 0.1110, which is illustrated in Fig. 3. Henceforth, we proceed to classify a subset of the segmented crops dataset, and its augmentation achieved by the rotation of each image at 90°.

Moreover, FibCAM also contains python codes that automatically account for the number of fibers and fiber area per slice for the entire stack. FibCAM’s quantitative analysis emphasizes the large variation of fiber cross sectional area that goes beyond the expected fiber diameters specified by the manufacturer (13-20 pixels). Also notice the fiber count oscillations across the stack, which indicate inaccuracies in the segmentation model [9] since no fiber breaks were expected in that CMC sample.

4 RESULTS AND CONCLUSION

The classification performance of the CNN model in Table 1 in terms of the sensitivity and specificity rates indicate that the fiber detection model presented in [6] hardly mislabeled non-fibers as fibers, but still missed a few true fibers. FibCAM experiments presented 2 main issues: a relatively small training set and limited generalization to the variety of fiber profiles.

Future work will address training size and variability to improve the accuracy of our CNN-based benchmarking system during evaluation of fiber detection methods, including support to eliminate false negatives from the results of such models. Because of the parallel nature of the CNN fiber-profile prediction algorithm, chunks of the original microCT volume can be processed separately in different computing cores with minimum communication. In doing so, we expect to take advantage of multi-core/many-core architectures available at DOE scientific computing facilities. Further efforts will explore parallelization using libraries such as joblib and dask.

ACKNOWLEDGMENTS

This research is supported by the Office of Science of the US Department of Energy under Contract No. DE-AC02-05CH11231 with Center of Advanced Mathematics for Energy Research Applications (CAMERA), the Gordon and Betty Moore Foundation through Grant GBMF3834, and the Alfred P. Sloan Foundation through Grant 2013-10-27 to the University of California, Berkeley.

REFERENCES


