

SANDIA REPORT

SAND2020-9099
Printed August 2020



Sandia
National
Laboratories

Automated Segmentation of Porous Thermal Spray Material CT Scans with Geometric Uncertainty Estimation

Carianne Martinez
Dan S. Bolintineanu
Aaron Olson
Theron Rodgers
Brendan Donohoe
Kevin M. Potter
Scott A. Roberts
Nathan W. Moore

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico
87185 and Livermore,
California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: <http://www.osti.gov/scitech>

Available to the public from

U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: <https://classic.ntis.gov/help/order-methods/>



ABSTRACT

Thermal sprayed metal coatings are used in many industrial applications, and characterizing the structure and performance of these materials is vital to understanding their behavior in the field. X-ray Computed Tomography (CT) machines enable volumetric, nondestructive imaging of these materials, but precise segmentation of this grayscale image data into discrete material phases is necessary to calculate quantities of interest related to material structure. In this work, we present a methodology to automate the CT segmentation process as well as quantify uncertainty in segmentations via deep learning. Neural networks (NNs) are shown to accurately segment full resolution CT scans of thermal sprayed materials and provide maps of uncertainty that conservatively bound the predicted geometry. These bounds are propagated through calculations of material properties such as porosity that may provide an understanding of anticipated behavior in the field.

ACKNOWLEDGEMENTS

Portions of this work were funded by the Laboratory Directed Research & Development program at Sandia National Laboratories (Credible Automated Meshing of Images LDRD Project #213016).

We are grateful for the discussion with the Uncertainty Quantification Team and for their insightful reviews of this work. Special thanks to David Stracuzzi and Gabriel Popoola for their suggestions and peer reviews.

CONTENTS

1. Introduction.....	11
1.1. Problem Motivation	11
1.2. Project Goals	12
1.3. Technical Approach	12
1.3.1. Synthetic Training Dataset Development.....	13
1.3.2. Algorithmic Segmentation of Real CT Scans.....	13
1.3.3. Neural Network Training.....	13
1.4. Contributions.....	14
2. Background/Related Work.....	15
2.1. Deep Learning Segmentation Models.....	15
2.2. Uncertainty Quantification for Deep Learning Predictions	15
2.3. Synthetic CT scans.....	16
3. Methods.....	17
3.1. Dataset generation and preparation	17
3.1.1. Synthetic material models	17
3.1.1.1. Ground truth data synthesis.....	17
3.1.1.2. Synthetic CT scans.....	18
3.1.2. Labeled CT data	19
3.1.2.1. CT examples	20
3.1.2.2. Naïve thresholding labels.....	21
3.1.2.3. Random walk algorithm labels.....	24
3.2. Deep Learning model architecture.....	26
3.3. Model Training.....	27
3.4. Analysis of pore characteristics.....	27
4. Results.....	31
4.1.1. One material thermal spray data	31
4.1.1.1. Trained with Random walk labels only.....	31
4.1.1.2. Trained with synthetic data only.....	32
4.1.1.3. Trained with synthetic data then refined with real CT examples	33
4.1.2. Two material thermal spray data.....	33
4.1.2.1. Trained with Naïve labels only	34
4.1.2.2. Trained with Random walk labels only.....	34
4.1.2.3. Trained with both Naïve and Random walk labels.....	35
4.1.2.4. Trained with synthetic data only.....	36
4.1.2.5. Trained with synthetic data then refined with real CT examples	37
4.1.3. Trained with binarized labels.....	37
5. Discussion.....	39
5.1. Comparison of Deep Learning predictions with labels.....	39
5.1.1. Qualitative comparison	39
5.1.2. Prediction characteristics.....	39
5.1.3. Accuracy to labels.....	40
5.1.4. Uncertainty characteristics	40
5.2. Interpretation of Uncertainty Maps	43
5.3. Downstream Analysis and Applications.....	43
5.4. Future Work	44

6. Conclusion	45
---------------------	----

LIST OF FIGURES

Figure 1: Distribution of total pore fraction from 25-statistically equivalent microstructure generation simulations.....	17
Figure 2: Slice of simulated one-material example with yellow representing the metal phase and dark purple representing the pore phase (left). Slice of synthetic CT scan of this example generated with ASTRA (right).	19
Figure 3: Slice of synthetic two-material example where distinct metal phases are represented in gray and black, and the pore phase is represented in white (left). Slice of synthetic CT scan generated for this (padded) example before cropping (right).....	19
Figure 4: Example image slices from CT scans of various samples. Left panel in each subfigure: significant portion of the field of view, excluding sample edges. Middle panel: magnified view of selected region of image in left panel (axis labels indicate region). Right panel: Histogram of grayscale intensity values.....	21
Figure 5: Examples of grayscale image slices and corresponding results of naive segmentation for 1-metal samples.	22
Figure 6: Examples of grayscale image slices and corresponding results of naive segmentation for 2-metal samples.	23
Figure 7: Examples of grayscale image slices and corresponding results of random walk segmentation for 1-metal samples	25
Figure 8: Examples of grayscale image slices and corresponding results of random walk segmentation for 2-metal samples	26
Figure 9: Example segmentations of a one material slice, along with low and high bounds and \bar{s} and σ_s	28
Figure 10. Summary of pore size distribution (see text for definition). Left panel shows linear scale, right panel shows same data on logarithmic scale.	30
Figure 11. Probability of percolation as a function of length scales in all three directions. The z direction corresponds to the sample spray direction.	30
Figure 12.: Sample cross section of held out CT scan (top left), associated random walk label (top right) and our model's predictions (lower right) and corresponding uncertainty map (lower left) for one material model trained on random walk labels.....	31
Figure 13.: Sample cross section of held out CT scan (top left), associated random walk label (top right) and our model's predictions (lower right) and corresponding uncertainty map (lower left) for one material model trained on synthetic data.....	32
Figure 14: Sample cross section of held out CT scan (top left), associated random walk label (top right) and our model's predictions (lower left) and corresponding uncertainty map (lower right) for the full pipeline one material model.....	33
Figure 15: Sample cross section of held out CT scan (top left), associated random walk label (top right) and our model's predictions (lower right) and corresponding uncertainty map (lower left) for two material model trained on naïve labels.	34
Figure 16: Sample cross section of held out CT scan (top left), associated random walk label (top right) and our model's predictions (lower right) and corresponding uncertainty map (lower left) for two material model trained on random walk labels.....	34
Figure 17: Sample cross section of held out CT scan (top left), associated random walk label (top right) and our model's predictions (lower right) and corresponding uncertainty map (lower left) for two material model trained on both naïve and random walk labels.	35

Figure 18: Sample cross section of real CT scan (top left), associated random walk label of that scan (top right) and our model’s predictions and uncertainties for two material model trained on synthetic data.....36

Figure 19: Sample cross section of held out CT scan (top left), associated random walk label (top right) and our model’s predictions (lower right) and corresponding uncertainty map (lower left) for the full pipeline two material model.37

Figure 20: Sample cross section of two material CT scan (top left), associated random walk label (top right), our model’s prediction (lower left) and the difference between ML and RW (lower right) for the full pipeline one material model.....38

Figure 21: Grayscale histograms with the number of voxels predicted to belong to each class. Pores are shown in blue, material one is shown in red, and material two is shown in yellow.39

Figure 22: Average uncertainties across correct and incorrect voxel predictions, using one-material synthetically trained model.40

Figure 23: Average uncertainties across correct and incorrect voxel predictions, using two-material synthetically trained model.41

Figure 24: Percentages of uncertain correct and incorrect voxel predictions, using one-material synthetic labels to train model. (Left) Full distribution. (Right) Zoomed to show small regions of distribution plot.42

Figure 25: Percentages of uncertain correct and incorrect voxel predictions, using two-material synthetic labels to train model.....43

LIST OF TABLES

Table 1: Mean statistics across all scans in the test set. Accuracy with respect to random walk labels, mean false positive and false negative rates across scans of both types, and average porosity for random walk segmentations and the predictions from our model.....31

Table 2: Prediction accuracies with respect to test set examples for one material model trained on synthetic data.32

Table 3: Mean statistics across all scans in the test set. Accuracy with respect to naïve and random walk labels, mean false positive and false negative rates across scans of both types, and average porosity for naïve labels, random walk segmentations, and the predictions the full pipeline one material model produces.33

Table 4: Prediction accuracies with respect to naïve and random walk labels, and uncertainty means, standard deviations, minimums and maximums, both overall and with respect to individual classes for two material model trained on naïve labels.34

Table 5: Prediction accuracies with respect to naïve and random walk labels, and uncertainty means, standard deviations, minimums and maximums, both overall and with respect to individual classes for two material model trained on random walk labels.....34

Table 6: Prediction accuracies with respect to naïve and random walk labels, and uncertainty means, standard deviations, minimums and maximums, both overall and with respect to individual classes for two material model trained on both naïve and random walk labels.35

Table 7: Prediction accuracies with respect to test set examples for two material model trained on synthetic data.36

Table 8: Prediction accuracies with respect to Naïve and random walk labels, and uncertainty means, standard deviations, minimums and maximums, both overall and with respect to individual classes for the full pipeline two material model.37

Table 9: Mean statistics across all scans in the test set. Accuracy with respect to naïve and random walk labels, mean false positive and false negative rates across scans of both types, and average porosity for naïve labels, random walk segmentations, and the NN model.38

This page left blank

ACRONYMS AND DEFINITIONS

Abbreviation	Definition
AS	Air-sprayed
BCNN	Bayesian convolutional neural network
CNN	Convolutional neural network
CS	Cold-sprayed
CT	Computed tomography
GPU	Graphics processing unit
MCDN	Monte Carlo dropout network
NN	Neural network
PSD	Pore size distribution

1. INTRODUCTION

In this section, we present the problem motivation, goals, and the technical approach we have developed under this project.

1.1. Problem Motivation

The work summarized in this report is motivated by a broader project characterizing the structure and performance of thermal-sprayed metal coatings. These coatings are currently used in applications such as airplane wings and turbines [1] and the linings of nuclear fusion reactors [2]. A precise understanding of structure informs understanding of structural properties and predictions of performance in environments including these high-mechanical-stress and thermal environments. Characterization of pore morphology is particularly important for several downstream quantities-of-interest and therefore receives more focus in this investigation than distinguishing between various metallic phases. In some samples, a second metal was sprayed with the first to provide imaging contrast and aid in the characterization process. Recent work has enabled collection of 3D images generated using X-ray computed tomography (CT) of these coatings [3] but a challenge remains in performing precise segmentation of the 3D grayscale image data to discrete material phases.

Perhaps the simplest segmentation technique to apply is a simple, “naïve thresholding,” segmentation in which an analyst chooses CT intensity value thresholds above and below which individual voxels are determined to be different materials. This technique relies on the analyst to choose threshold values based on either what looks reasonable or outside information such as how much of each material is present in the samples. This method requires analyst attention for segmenting each sample individually and its accuracy/reproducibility is contingent on analyst judgment or often hard-to-get external information. Additionally, this method has a lack of achievable accuracy since often no threshold can be chosen for which all intensity values below the threshold are one material and all values above the threshold are another material. One contributing factor to this inherent ambiguity is that CT reconstructions have flaws wherein the local intensity is dependent upon the structure of surrounding materials which can cast “shadows”.

In previous work [3], several segmentation techniques from the literature were explored for segmenting these datasets. The “random walker” algorithm was believed to be the most accurate of those surveyed. Nonetheless, questions remain about the accuracy of segmentations produced with the method on our datasets.

Additionally, neither the “naïve” thresholding method nor the random walker method provide uncertainty estimates of their segmentation predictions. It is therefore highly desirable to develop a segmentation method for these datasets that can make use not only of intensity values but also complex material autocorrelation behavior to yield precise segmentation and provide a meaningful estimate of segmentation uncertainty. This will yield more accurate downstream analysis of thermal-sprayed coatings as well as the ability to propagate uncertainty caused by segmentation to uncertainty in downstream quantities-of-interest.

Accurate segmentation of grayscale CT images and associated uncertainty quantification will be further used to characterize thermal-sprayed material structures for a host of different analyses. Analyses include:

- Assessing percolation properties of thermal-sprayed structures to help inform what gases fill the pores [4]
- Measuring spatial statistical properties from samples to enable tuning of models that produce synthetic realizations including the SPPARKS thermal-spray model [5] [6]

- Establishing the ability of cheaper and quicker 2D mesostructures [7] to work as a surrogate of 3D analysis for evaluation of various properties dependent on mesostructure
- Enabling effective comparisons of images of structures before and after non-destructive hydrodynamic tests such as those performed at the National Ignition Facility [8]
- Informing hydrodynamic simulations in thermal-sprayed structures [3]

1.2. Project Goals

The goal of the work represented in this report is to develop an accurate method for segmenting grayscale images of CT datasets and provide uncertainty estimates on the segmentation. Ultimately, accurate segmentation will be used to understand internal coating structures, inform tuning of the thermal-spray process, and assess performance of the coatings in various environments including mechanical and thermal environments.

1.3. Technical Approach

Our technical approach to the automatic segmentation with geometric uncertainty quantification is driven toward the development of a deep learning model – a neural network (NN) – capable of learning from examples to perform accurate, consistent segmentations. The approach to this supervised learning problem consists of several steps including:

1. Development of an entirely synthetic training dataset
2. Algorithmic segmentation of real CT scans to serve as another labeled training dataset
3. Training a NN with both synthetic and real data to automatically segment CT scans with geometric uncertainty characterization
4. Inference of segmentations with uncertainty maps of all real CT scans

Several recent advances in computer vision, such as those described in Section 2.1 herein, have led to deep learning methods being overwhelmingly adopted as the standard approach to image segmentation in the literature. Specifically, the most successful segmentation models have been trained in a supervised fashion, where labeled examples of accurate segmentations are presented to the NN and training is guided to match the examples.

In this work, we have no ground truth examples of perfect segmentations of the CT scans that are the target of the project. Our domain of interest is a set of real CT scans of two types of thermal spray materials. The first two steps of our approach are required to produce training sets such that the deep learning model can learn with synthetic examples that provide ground truth that is close to our domain of interest as well as two sets of viable but imperfect labels of the real target domain. The next steps in the process are training and inference typical of deep learning applications.

In this section, we provide a summary of each of these steps. We present detailed descriptions of the methods in Section 3.1.

1.3.1. Synthetic Training Dataset Development

In the absence of ground truth segmentations of the CT scans of the real materials with microstructures that are our target domain, we use simulated examples of these microstructures to generate an initial training set for our deep learning model. Synthetic data can provide examples of textural features that are important for NNs to capture to perform accurate segmentation. Using structures generated with the SPPARKS simulation library [5] further described in Section 3.1, we construct a training set with synthetic material examples of varied porosity intended to capture the viable range of porosities present in the real material examples that we intend to segment.

Starting from this numerical representation of the microstructure of the target domain, for each simulated material example, we used open source software to generate a synthetic CT scan of the numerical object. We compared the synthetic CT scans to real CT scans of the target material, and added noise to the synthetic scans to produce scans that are qualitatively close in appearance to the real scans that are the target of the project. A full description of the code developed to generate the synthetic CT scans is presented in Section 3.1.

1.3.2. Algorithmic Segmentation of Real CT Scans

In addition to the synthetic training dataset, we used two algorithmic techniques to generate segmentation examples of the real CT scans from our target domain that we used to train the deep learning model. Ideally, to generate ground truth labels for a segmentation training dataset, a human would be tasked with labeling each voxel in a CT scan by the type of material the voxel represents. For our target domain, the real CT examples comprise on the order of 1 billion voxels, presenting an impractical task for a human to label manually. As such, we have leveraged two algorithmic approaches to generate segmentations that are useful for the deep learning algorithm to learn how to identify features in the real data relevant to a successful segmentation. The use of these flawed segmentations of real CT examples gives the network exposure to the real domain of interest. Between the varied examples from different algorithms and the synthetic data, the network learns a useful segmentation as shown qualitatively in Section 4.

The details of the two algorithmic approaches are presented in Sections 3.1.3.2 and 3.1.3.3 herein.

1.3.3. Neural Network Training

We use a V-Net model [9] with spatial dropout layers to calculate segmentations and uncertainties [10] for synthetic and real three-dimensional thermal spray CT scans. We separate training of our model into two primary phases. For our first phase of training, we train the model using synthetic examples and segmentations generated using the SPPARKS microstructure library [5]. For our second phase of training, we train the model on real world CT data, for which ground truth segmentations are not available, instead being estimated using two different approaches: (a) naïve thresholding and (b) the random walk algorithm.

1.4. Contributions

The primary contributions of this work are as follows:

- We have augmented the existing V-Net architecture to generate segmentations and uncertainty maps for inputs of multiple sizes, employing a stitching approach to limit overall graphics processing unit (GPU) memory consumption.
- We have implemented a training pipeline combining three different labeling approaches (synthetic, thresholding, and random walk) to produce useful segmentations in the absence of expert labelled ground truth examples.
- Uncertainty maps generated by our model appear to align with common intuitions – with substantially higher uncertainty around edges and novel features and lower uncertainty elsewhere.

2. BACKGROUND/RELATED WORK

In this section, we describe recent advances in deep learning segmentation, uncertainty quantification for deep learning model predictions, and synthetic CT scan software.

2.1. Deep Learning Segmentation Models

Image segmentation is a key problem with applications to a number of computer vision related tasks, including medical imaging analysis [11], scene recognition and understanding [12], and robotics [13]. Classical techniques for performing image segmentation include thresholding [14], clustering [15], and region growing methods [16]. In recent years, however, supervised deep learning methods for image segmentation have grown increasingly popular due to their ability to leverage human knowledge via labelled examples and achieve high accuracy, and as such have set benchmarks across a variety of datasets.

In [17], Shelhamer et al. investigate the application of fully convolutional neural networks (CNNs) to segmentation tasks on two-dimensional images, employing convolutional and pooling layers, as well as additional skip connections at intermediate levels of the network for the purposes of combining finer, lower level information as well as coarser, higher level information. This approach motivated the U-Net architecture for two-dimensional images presented by Ronneberger et al. [18] which further refined this methodology by using an architecture comprised of symmetric downsampling and upsampling paths, with skip links allowing information to be copied from an intermediate layer of the downsampling path directly to the corresponding layer of the upsampling path. In this way, the network is able to combine information from multiple resolutions to produce more refined segmentations. Milletari et al. [9] later released V-Net, an extension of the U-Net architecture built to handle and segment three-dimensional data.

2.2. Uncertainty Quantification for Deep Learning Predictions

Deep learning models are powerful tools with the ability to leverage patterns found in data to achieve highly accurate results in computer vision tasks like object detection, image classification, and segmentation. However, predictions from these models do not come with error bars by default. Uncertainty quantification for deep learning predictions is an active and open research field, and several approaches have been set forth in the literature.

Ensemble methods [19] have been shown to provide uncertainty estimates for deep learning predictions by training several NNs independently and pooling their output to calculate a final prediction while retaining the variance over the model predictions to inform the uncertainty in the model's output. Deep learning models learn sets of filters that are relevant to the task at hand, but there is not a unique set of filters that will result in a high-performing model. Depending on differences in weight instantiation of a model, these filters can vary significantly while still producing accurate results, and different models may leverage different filters more heavily in the process of computing a final prediction. This variance in learned filters allows for a diverse set of predictions that can be interpreted as uncertainty. One disadvantage of the ensemble method is that several models must be trained and retained on the same task, likely with redundant sets of filters.

Another method proposed in the literature is the Bayesian neural network [20], where instead of point estimates of optimal NN weights that produce accurate output, a distribution over each NN weight is learned. Each weight in the network is represented by two parameters: a mean and a variance, each of which is learned via variational inference that approximates the posterior distribution of each weight given a prior distribution along with the training data observations.

These networks require double the parameters when compared with their non-Bayesian counterparts that learn point estimates of the optimal values of each network weight, but are capable of retaining variance both internally to the NN at each weight, and in the output space, where variance over predictions is retained. This team is currently developing Bayesian Convolutional Neural Networks (BCNNs) [21] under the CAMI LDRD Project #213016, and have found that they are sensitive to the data and can be difficult to train. We intend to process the subject thermal spray data with the BCNN in future work.

Finally, the approach we take in this work leverages work by Gal, et al. [10], where dropout layers remain active in the NN at inference time to introduce stochasticity in the model's predictions. The resulting Monte Carlo Dropout Network (MCDN) takes several samples of forward passes through the NN for each input example, producing a set of viable segmentations of the input example. Since the NN includes a final softmax layer, the model's output consists of a vector of length N comprised of floating point values between 0 and 1, where N is the number of possible materials or classes being predicted. Dropout layers are typically used during training to regularize NNs to prevent overfitting to the training data by dropping the output of a randomly sampled subset of NN weights during training. Gal, et al. [10] showed that active dropout layers at inference time approximate a Gaussian process leading to variance in the NN's predicted outputs. The final segmentation is taken as the mean over all inference runs for the same input, and the standard deviation over the outputs is interpreted as the uncertainty of the model's prediction. The main advantage of the MCDN is that dropout layers are easily integrated into already proven accurate NN models without the need to alter the architecture, increase the number of NN parameters, or train multiple models.

2.3. Synthetic CT scans

Several software packages are publicly available to generate synthetic X-rays and CT scans of numerical objects. For this work, we use the ASTRA Toolbox python library [22] to generate synthetic X-ray CT scans of the simulated microstructures generated from the SPPARKS thermal spray library [5]. The python API available in ASTRA toolbox allows for GPU-accelerated image processing and reconstruction with several tunable settings such as detector size, number of X-ray projections, and source and detector position relative to the numerical object. The software is able to produce cone beam CT scans similar to those that are within the target segmentation domain for this project.

3. METHODS

In this section, we present the details of our methods from dataset curation to NN model development and training, and finally through inference and interpretation of our geometric uncertainty characterization.

3.1. Dataset generation and preparation

In order to train a NN to perform segmentation, labeled examples are necessary to accommodate supervised learning. The CT scans that are the target of this investigation lack ground truth labels, and as such we have developed training sets that consist of synthetic examples to augment training as well as algorithmically labeled examples from the real CT domain.

3.1.1. Synthetic material models

Synthetic training used in the study was re-used from previous investigations of input parameter variation and stochastic model variation. The microstructures were not quantitatively tuned to recreate the experimental structures, but contained many relevant features such as multiscale porosity, unmelted particles, and layer-structures. The synthetic structures do not contain some features found in the experimental data that create additional complications such as oxides and diffraction ring artifacts from the CT process.

Two-material volumes were generated in a study of stochastic model variation. 25 simulations were performed with identical input parameters but unique random number generator seeds [3]. Model parameters used in the study resulted in a total pore fraction distribution with a mean of 0.059 and standard deviation of 0.0122. The distribution of total pore fractions is shown in Figure 1.

An additional dataset of 20 simulations (10 single material volumes and 10 two material volumes) was generated to create synthetic volumes with total pore fractions closer to the expected experimental mean of ~ 0.02 . The range of total pore fractions in this second set varied between 0.02 and 0.05.

3.1.1.1. Ground truth data synthesis

Ground truth data was generated using the Thermal Spray app in the SPPARKS microstructure simulation library [5] [3]. The thermal spray model generates synthetic microstructures at experimentally relevant length-scales using a rules-based model of particle incidence, spreading, and solidification during thermal spray processes. Synthetic microstructures consist of one pore/void phase and an arbitrary number of solid phases. Relative abundances of the solid phases are specified by user parameters, while the abundance of the pore phase emerges from the simulation behavior and cannot be directly specified with an input parameter.

As material phases are represented by integer values, the synthetic volumes are “pre-segmented” and do not require interpretation from continuum intensity values as with experimental CT data. The simulation also generates a unique label for each

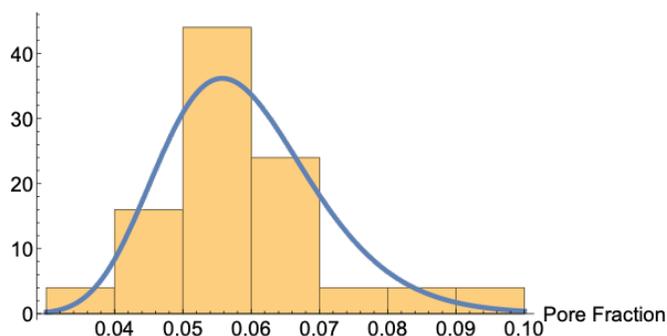


Figure 1: Distribution of total pore fraction from 25-statistically equivalent microstructure generation simulations.

individual particle/splat; this is not currently used in the segmentation training but would provide ground-truth data for the identification of individual splats, rather than just material phases. Two classes of microstructures were generated for training – one-material volumes representing Ta coatings and two-material volumes representing Ta & Nb coatings. The Ta/Nb coatings were created to match volume fractions of .7 Ta and .3 Nb. The two material volumes were easily transformed to one material volumes by assigning all solid phases to the same integer.

3.1.1.2. Synthetic CT scans

After establishing the ground truth synthetic material examples, we leveraged an open source python library to develop a synthetic CT generation framework with parameters set to result in images that qualitatively match the real CT domain. The ASTRA toolbox [22] is a library that provides GPU-enabled capabilities to generate simulated X-rays with flexibility to set options such as detector resolution, X-ray source type, reconstruction algorithms, and physical dimensions that define the geometric setup to match the fielded CT machine used to generate scans of the materials that are the subject of this work.

ASTRA takes parameters that define the physical setup of the CT machine to be simulated. We experimented with various parameters to qualitatively match features like resolution, feature scale and noise we observed in the real CT data. Geometric parameters were set to use 1800 cone beam projections with a 1536x1536 pixel detector with spacing of 4 units between adjacent pixel centers over a full 360-degree rotation of the synthetic object. The synthetic object was set to have a 1:3 ratio between the distance to the simulated X-ray source from the object and the distance from the detector to the object. The FDK_CUDA algorithm from the ASTRA library reconstructed the volumetric CT image of the object from the synthesized X-rays, and was executed on the GPU.

To ensure the ground truth material volume would be encapsulated in the imageable area of the simulated CT framework, the volume was padded with zero values along all axes, then the simulated CT scan was generated for each example. Several post-processing steps were taken to transform the synthetic scans to have voxel intensity values similar to the distribution found in the real CT domain. Voxel values were limited to the range [0,255] by first replacing all negative values within the synthetic CT scan volume with zeros then dividing by the maximum value in the volume, multiplying the result by 255 and rounding to the nearest integer value. Next, the volume was cropped to remove the zero padding needed for the CT simulation. Finally, noise was introduced to the volume via elementwise multiplication with a volume of randomly generated floating point values sampled from a normal distribution with mean 1 and 0.1 variance. All volumetric post-processing leveraged the Numpy python library [23] for efficient calculations.

The simulated CT framework was implemented on Sandia's Synapse GPU cluster, specifically, on a DGX-2 machine with eight 32 GB GPUs and approximately 1.5 TB of system memory.

Examples of the simulated CT scans of one-material and two-material simulated objects are shown in Figure 2 and Figure 3.

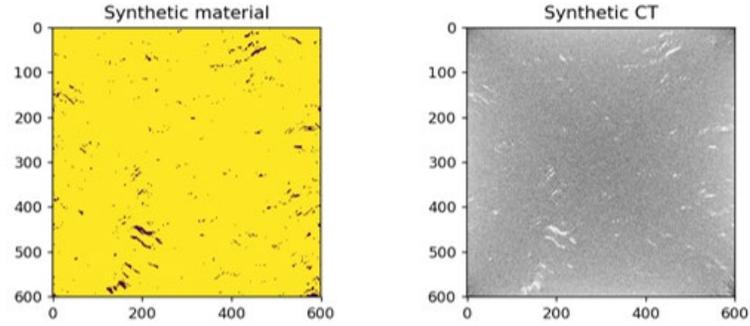


Figure 2: Slice of simulated one-material example with yellow representing the metal phase and dark purple representing the pore phase (left). Slice of synthetic CT scan of this example generated with ASTRA (right).

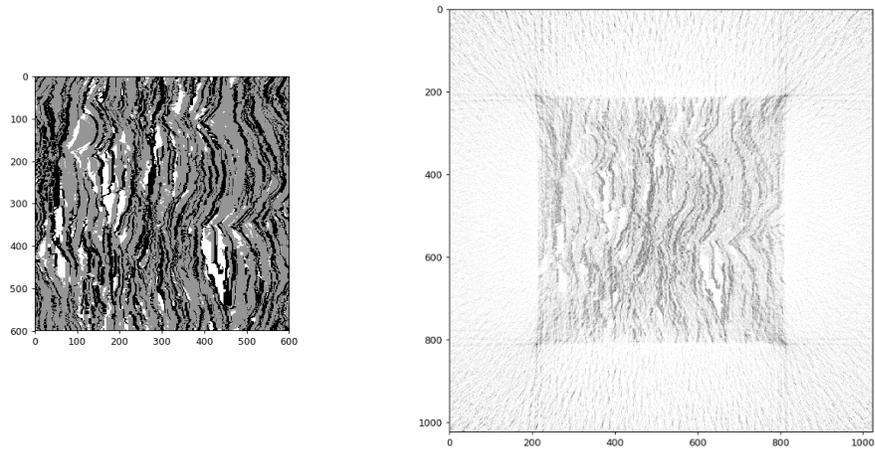


Figure 3: Slice of synthetic two-material example where distinct metal phases are represented in gray and black, and the pore phase is represented in white (left). Slice of synthetic CT scan generated for this (padded) example before cropping (right).

3.1.2. Labeled CT data

X-ray computed tomography experiments were carried out at the Advanced Photon Source on several dozen samples, spanning both 1-metal and 2-metal formulations, as well as atmospheric and cold spray processing conditions. For details of sample preparation and X-ray characterization, the interested reader is referred to [24] [3]. As discussed previously, the primary goal of X-ray image segmentation is to separate porous regions from solid metal-filled regions in the CT scans for purposes of quantifying pore morphology. Porous regions (whether gas-filled or vacuum) have very low X-ray attenuation and result in low grayscale intensity/dark regions in reconstructed images; conversely, high density/high attenuation metal regions result in much higher grayscale intensities, and appear as light regions in reconstructed images.

Despite this large difference in grayscale intensity, two key factors result in significant challenges for image segmentation: first, pore space features are occasionally comparable to the X-ray CT imaging resolution ($<1 \mu\text{m}$), which leads to voxels that contain both pore and metal phases, resulting in intermediate grayscale intensity values. Second, the presence of metal oxides in the samples results in

regions that likewise have intermediate intensity values, further complicating the identification of pore and solid phases. In particular, during atmospheric plasma spray processing, which takes place in a high-temperature, oxygen-rich environment, a variety of metal oxides are formed in both one- and two-material samples. These metal oxides have lower X-ray attenuation coefficients in comparison to the pure metals, but also higher attenuation in comparison to porous regions, resulting in grayscale values intermediate between pore and metal. In the following section, we provide additional details along with pertinent example images of these scans to illustrate these challenges in the context of image segmentation.

3.1.2.1. CT examples

Figure 4 shows example image slices from a selection of CT scans of one-metal air-sprayed (AS) and cold-sprayed (CS) samples, as well as two-metal AS samples. Images include a significant portion of the full field of view, as well as magnified regions to show detailed features of interest. The darkest regions (intensity $< \sim 20$) correspond to porous regions. Intermediate gray values correspond to metal oxides in 1-material regions, as well as a combination of metal oxides and one of two metals in the two-metal case. In all cases, the brighter gray regions correspond to pure metal. The precise intensity values of each of these phases are not known, and likely overlap significantly as discussed above, which makes image segmentation a challenging task.

We have initially focused efforts on separating pore space from any solid phase (i.e. oxide or metal), but separation of oxides from metal, as well as separation of different metals in two-metal samples is also of interest for future work. For all examples in Figure 4, a histogram of all intensity values is also shown. While in some cases there appears to be a separation of the pore space intensity (a trough in the histograms at low intensity values), thresholding images based on the intensity at the bottom of this trough does not produce satisfactory results. Likewise, when separating intermediate and dark gray regions (oxides and/or different metals), the separation is even more ambiguous, with clear overlaps between the two intensity ranges. These considerations ultimately motivated us to pursue more sophisticated segmentation techniques than typical global thresholding techniques that establish a global threshold based on features of the histogram.

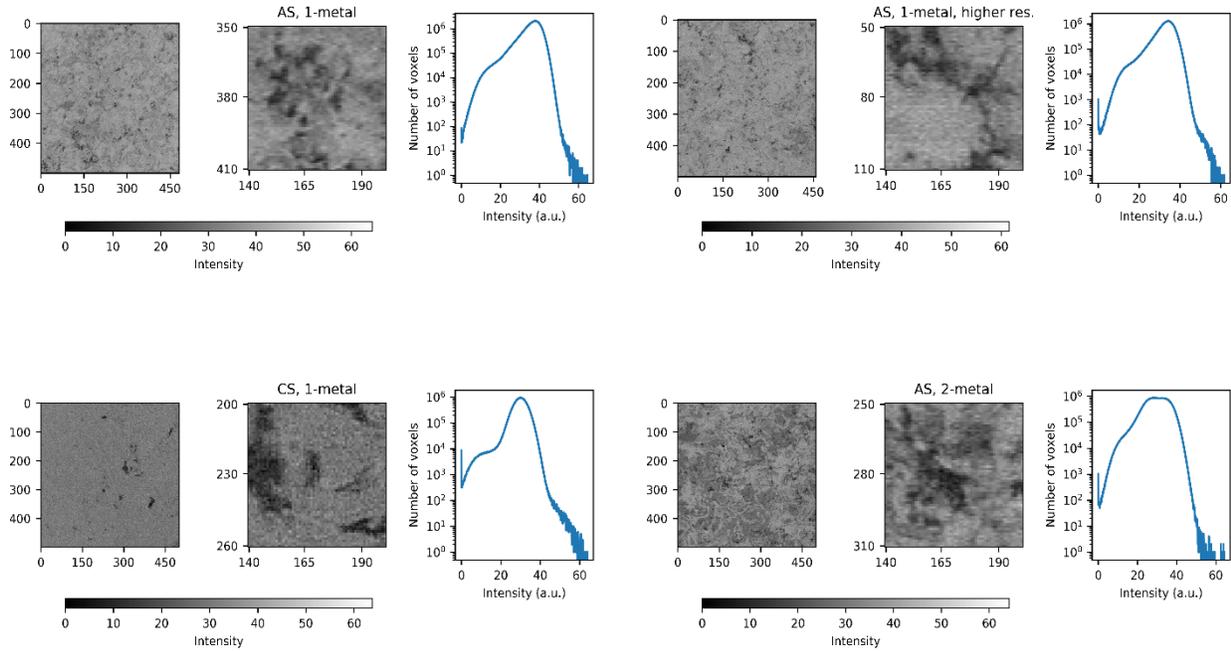


Figure 4: Example image slices from CT scans of various samples. Left panel in each subfigure: significant portion of the field of view, excluding sample edges. Middle panel: magnified view of selected region of image in left panel (axis labels indicate region). Right panel: Histogram of grayscale intensity values.

3.1.2.2. Naïve thresholding labels

We denote as “naïve thresholding” a labelling method that selects a global threshold on intensity based on analyst judgment and/or known pore and material volume fractions. The approach used for the “naïve thresholding” in this work was for an analyst:

1. Proposes a thresholding value
2. Segments the image based on that value
3. Compares the grayscale and proposed segmented images by eye and evaluate how reasonable the segmentation looked
4. Iterate on this process until thresholding values were settled on what appeared to produce a reasonable segmentation

A particular trait used by the analyst in this process was to select thresholding values that tended to minimize the amount of “speckling” in regions that were likely to be all pore or a particular metallic phase. If a threshold value is chosen larger than it probably should be, small clusters of the less-dense material show up like “speckles” in regions that are likely the higher-density material. Likewise, if a threshold value is chosen smaller than it probably should be, small clusters of the more-dense material show up like “speckles” in regions that are likely the lower-density material. Resulting volume fractions were then checked to ensure they were within plausible ranges based on outside information. In previous work, an estimate of 6% porosity was attained by mechanical grinding and polishing followed by SEM imaging [3]. It is possible this is a slight overestimate caused by material pull-out during grinding. Other estimates put upper bounds on the

porosity higher than 6%, though these assessments lack precision, likely due to the presence of metallic oxides in the samples.

Another possible application of this method would be to select thresholding values that yielded volume fractions ascertained or supposed based on external information or supposition. Intensity values could then be sorted in ascending order and the value found and selected that corresponds to the volume-fraction percentile of interest (e.g. 6%).

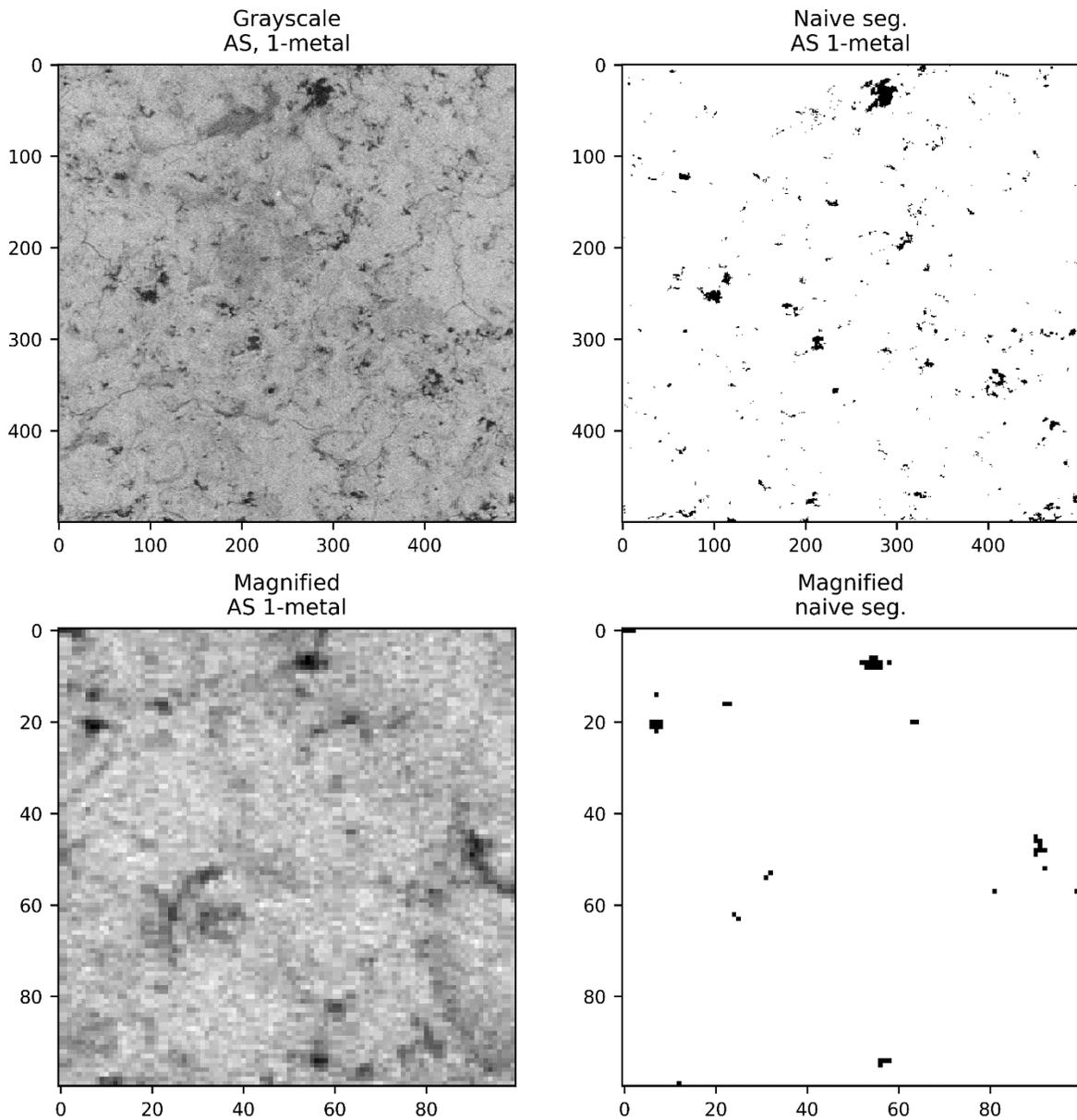


Figure 5: Examples of grayscale image slices and corresponding results of naive segmentation for 1-metal samples.

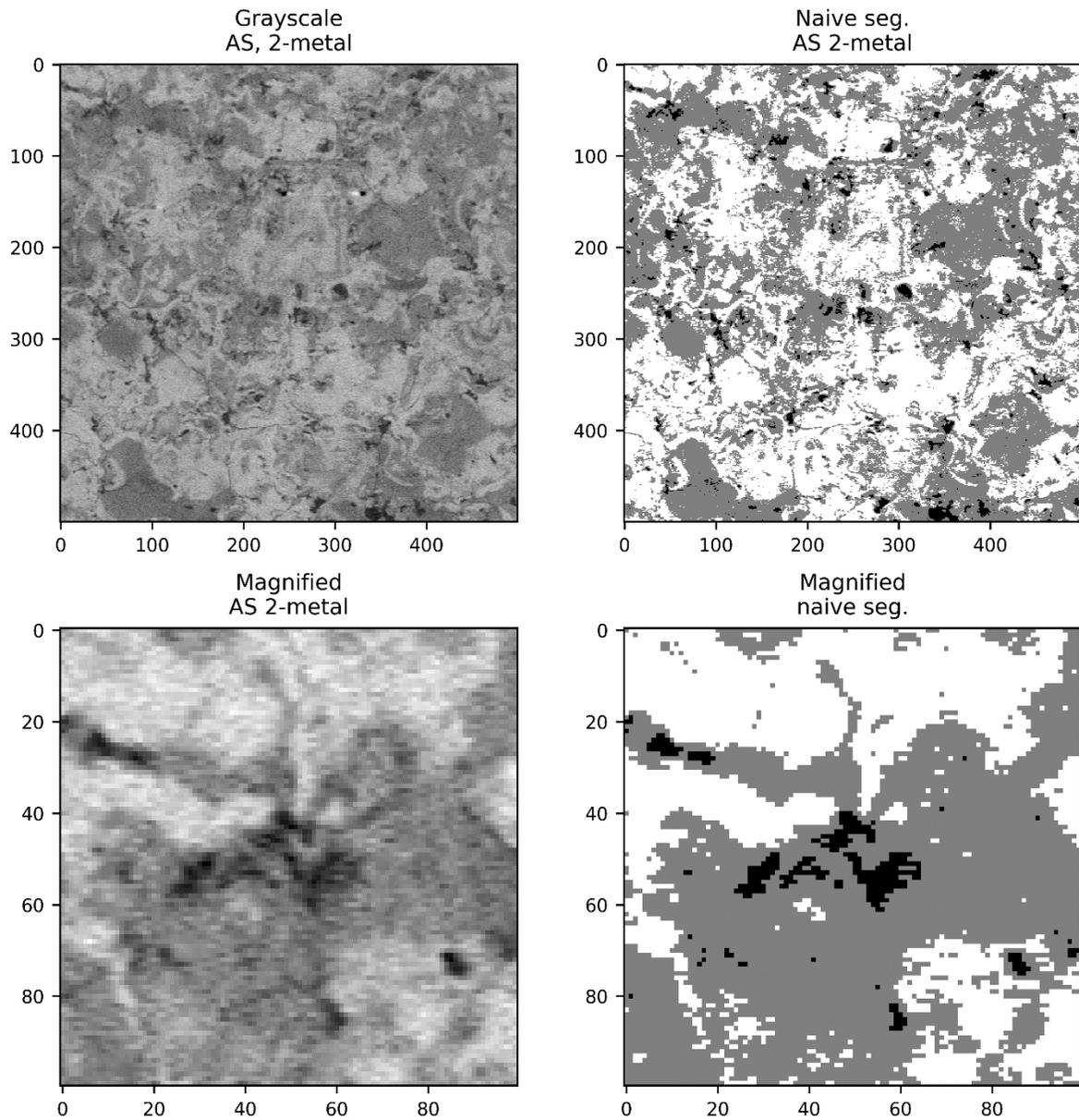


Figure 6: Examples of grayscale image slices and corresponding results of naive segmentation for 2-metal samples.

As this approach is based on a global threshold, it does not consider any local pixel information, which results in erroneously noisy features. This is apparent in Figures 5 and 6, particularly in the magnified views in the bottom rows of the figures. Furthermore, the target volume fractions are not generally known with high accuracy; instead, it would be highly valuable to obtain an estimate of these volume fractions (particularly porosity) based on analysis of the CT scans, rather than imposing them as part of the segmentation process.

3.1.2.3. Random walk algorithm labels

We have also attempted a more sophisticated segmentation approach that takes advantage of local pixel information, as developed by Grady [25]. This algorithm starts with assigning phase labels to a subset of voxels for which phase labels are known with high certainty, which we refer to as ‘seed’ labels. Conceptually, segmentation then proceeds by placing random walkers on all of the remaining, unlabeled voxels, and allowing them to carry out a random walk, where the likelihood of a move to a neighboring voxel is inversely proportional to the difference in grayscale intensities of those neighboring voxels. The probability of a given voxel belonging to each phase is then assigned according to the probability that a random walker beginning on the given voxel first reaches one of the seed label voxels corresponding to that particular phase. Computing this probability amounts to a linear algebra problem, so the random walk simulations do not need to be carried out directly. Additional details are provided by Grady [25]. In all random walk segmentations, we have used the open-source Python implementation in the *scikit-image* library [26].

Seed labels for random walk segmentation were assigned following an approach similar to the naïve segmentation, but with percentiles chosen very conservatively to ensure high certainty in the assigned voxels. The 1st percentile of intensity was used to assign seed labels for the pore phase, and the 60th percentile was used to assign labels to the solid phase. All voxels with intermediate intensity values were labelled via the random walk segmentation described above. Figure 5 and Figure 6 show resulting segmentations for the same slices as the naïve segmentation above, for both one- and two-material samples. In the case of two-material samples, similarly conservative percentiles were used in assigning seed labels to a small subset of the image. As seen from the images, the resulting segmentations contain much less noise/single-voxel features, which likely represents more meaningful segmentations. More importantly, the algorithm does not require a known porosity as an input, but rather provides an estimate of this important quantity based on the image segmentation. However, there are occasional pore regions that are completely missed by this segmentation, as seen in the magnified view of the one-metal sample (bottom right panel of Figure 7). We attribute this to an inadequate assignment of seed labels, wherein the random walk algorithm requires at least one voxel to be assigned as a seed label for each topologically connected pore region. With the conservative seed labeling strategy used here, this occasionally fails for very small pore regions, where partial voxel effects result in higher relative values of intensity. Nevertheless, the segmentations produced by the random walk algorithm are generally satisfactory and expected to provide good training data for the machine learning algorithms deployed herein.

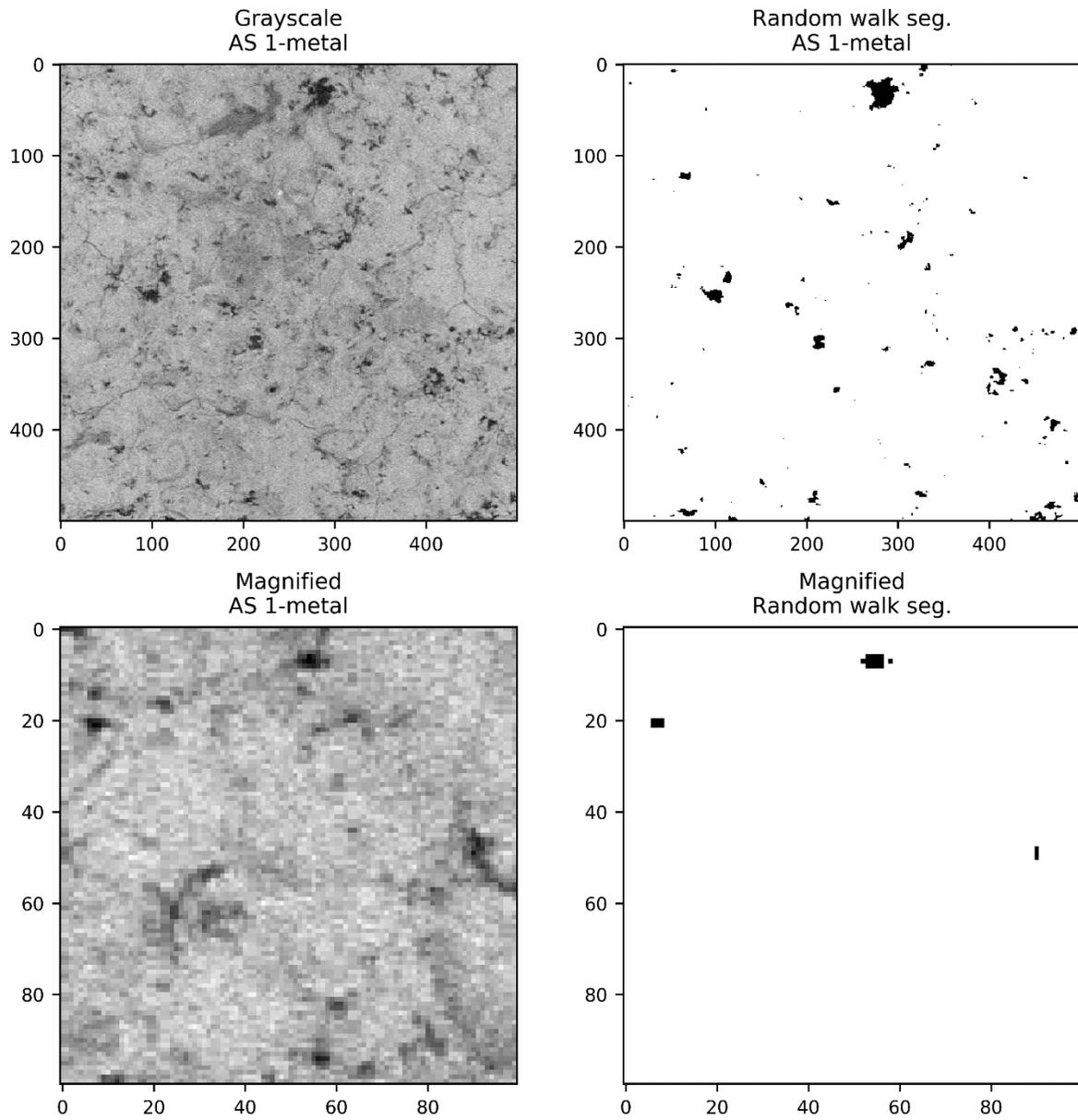


Figure 7: Examples of grayscale image slices and corresponding results of random walk segmentation for 1-metal samples

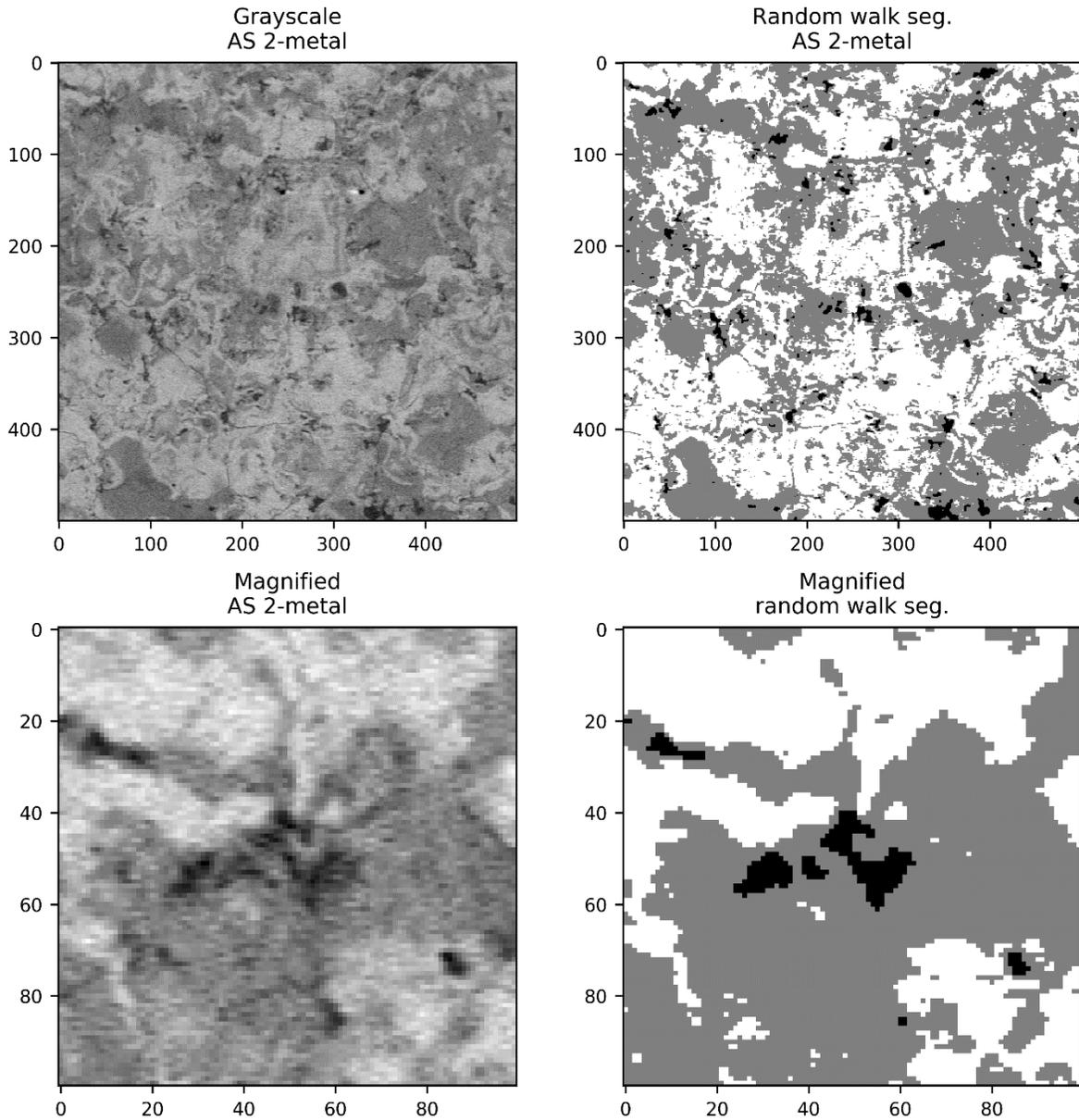


Figure 8: Examples of grayscale image slices and corresponding results of random walk segmentation for 2-metal samples

3.2. Deep Learning model architecture

The development of the Deep Learning methods used in this project was primarily funded by the CAMI LDRD, Project #213016. In our experiments, we use a modified V-Net, implemented in Keras [27] with a TensorFlow [28] backend taking a volumetric input chunks of size $(240, 240, 240)$ and producing a softmax output vector map of size $(240, 240, 240, c)$, where c is the number of output classes within each softmax output vector. The V-Net consists primarily of four downsampling blocks and four upsampling blocks, with skip connections between the corresponding downsampling and upsampling blocks. At each block, the input undergoes a series

of three-dimensional convolutional operations with unit stride length, same padding, and a fixed number of input channels for all layers in the block, followed by a final operation serving to either downsample or upsample the input. In the case of the downsampling block, this final operation is a downsampling convolutional operation using a kernel size of 2 and a stride of 2, reducing the input size in all dimensions by two and doubling the number of input channels. In the case of the upsampling block, this final layer is an upsampling transposed convolutional layer using a kernel size of 2 and a stride of 2, upsampling the input size in all dimensions by two and reducing the number of input channels by a factor of two. Each convolutional and transposed convolutional operation in the upsample blocks is followed by a three-dimensional spatial dropout operation, which deactivates a random fraction of feature maps in the input during each forward pass.

The dropout layers serve to add stochasticity to the model. We take advantage of this stochasticity, using the model to perform several rounds of inference on a single, fixed volumetric input chunk. We take the various resulting output maps and compute their means and standard deviations. The former gives us an improved, reliable output map estimate from which class predictions may be calculated, and the latter gives us an uncertainty map, yielding estimates of the model’s uncertainty at different voxel locations within the image.

3.3. Model Training

As is often the case with three-dimensional computed tomography data, the size of our datasets is such that it is impractical to keep an entire scan in GPU memory at once. Rather than passing full three-dimensional scans when training, we instead randomly take a uniform sample of size (240, 240, 240) from our input and target volumes, respectively. During inference, we partition each input volume into evenly spaced, overlapping chunks of the same size, with a stride length of 208 in each direction. We infer on these chunks individually and take the average mean and uncertainty for each voxel in the input volume. For mean and uncertainty calculation, we perform multiple iterations of dropout at each location.

We use the negative log likelihood loss, weighted by the logarithm of the inverse frequency of each class to account for class imbalance within each chunk. For optimization, we use Keras’ implementation of Adadelta with a learning rate α of 0.001 and a decay rate ρ of 0.95.

3.4. Analysis of pore characteristics

In order to analyze pore characteristics via different morphological metrics, segmented three-dimensional images are required. Translating per-voxel uncertainty outputs of the model to error bounds in various pore characteristics is an ongoing area of research, but we adopt a relatively simple approach here to obtain conservative (that is, very loose) bounds on these pore characteristics. For a given three-dimensional (grayscale) image, the CNN segmentation provides ten distinct inferences, corresponding to ten distinct randomly selected dropout configurations. In each inference, a softmax function value is assigned to every voxel for every phase label, where the softmax value is between 0 and 1, indicating a relative likelihood of a given voxel belonging to each phase. The mean and standard deviation of the values of the softmax output associated with the pore phase class label are then computed at each voxel across these ten inferences. These mean and

standard deviation values are denoted as \bar{s} and σ_s . The value of σ_s is associated with a relative level of uncertainty. The most likely segmentation, which we denote as the “base case”, is generated by simply setting a threshold of 0.5 on \bar{s} , i.e. all voxels with $\bar{s} > 0.5$ are assigned to the pore phase, regardless of the value of σ_s . To generate additional segmentations based on the associated uncertainty σ_s , a value of σ_s is first chosen to denote voxels with moderately high segmentation uncertainty. This choice of $\sigma_s > 0.02$ is somewhat arbitrary, but empirically was found to correspond to the value of σ_s that captures the vast majority of voxels that result in different class labels across any of the ten inferences of the CNN (i.e. the majority of voxels that change labels across any of the ten inferences satisfy $\sigma_s > 0.02$). To generate segmentations that represent bounds on the base case, all voxels that satisfy $\sigma_s > 0.02$ are set to the pore phase, resulting in a porosity “upper bound” segmentation; conversely, setting all voxels that satisfy $\sigma_s > 0.02$ to the solid phase produces a porosity “lower bound” segmentation. Figure 9 depicts the quantities \bar{s} and σ_s , as well as the base case and lower and upper bound segmentations for a sample two-dimensional slice of a three-dimensional dataset.

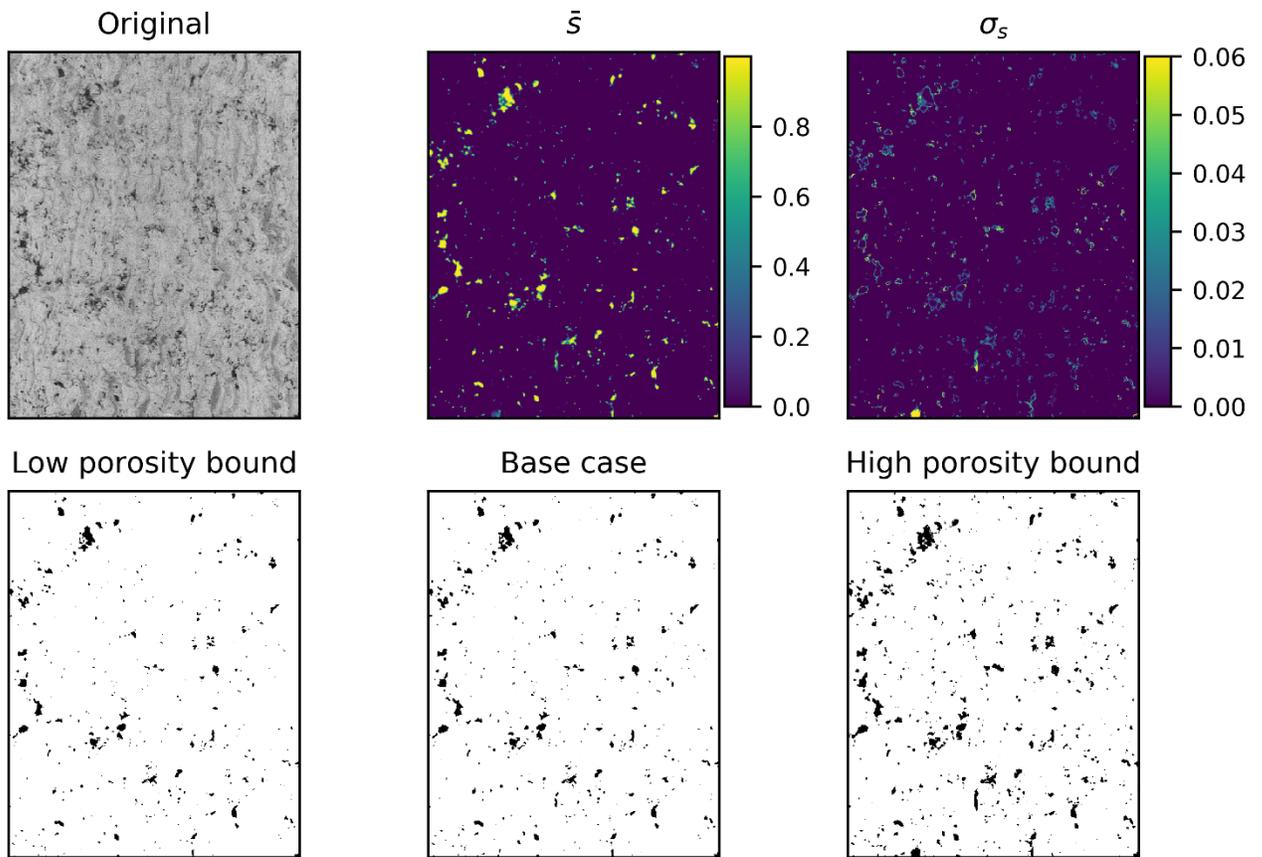


Figure 9: Example segmentations of a one material slice, along with low and high bounds and \bar{s} and σ_s .

With the low, base and high porosity segmentations as bounds, we compute several metrics that quantify the pore morphology. In the present context, the nature of these metrics and their physical implications are not discussed in detail; the interested reader is referred to other works for this

discussion [3] [4] [5]. Here, these metrics are only presented to illustrate quantitative differences arising from image segmentation.

First, we compute the two-point correlation function $S_2(\mathbf{r})$, which is a common measure used to describe the spatial distribution of a component within a heterogeneous material. It can be interpreted as the probability of two points that are separated by a line segment \mathbf{r} randomly placed in a volume both belonging to the pore phase [29] [30]. Given the symmetry of the samples, we plot this metric as a function of $\mathbf{r} = \mathbf{z}$ defined as the distance in the spray direction, as well as a radial distance r in the plane perpendicular to the spray direction (denoted as non-spray direction). The value of S_2 at both $\mathbf{r} \rightarrow \mathbf{0}$ and $\mathbf{z} \rightarrow \mathbf{0}$ is the porosity, i.e. the total fraction of the volume occupied by pore space. Figure 10 below summarizes the results for this metric in both the spray and non-spray directions. Lines for each segmentation bound correspond to the mean value across thirty scans, and error bars correspond to the standard deviation across the same scans. Additional details are provided by Rodgers, et al. [6].

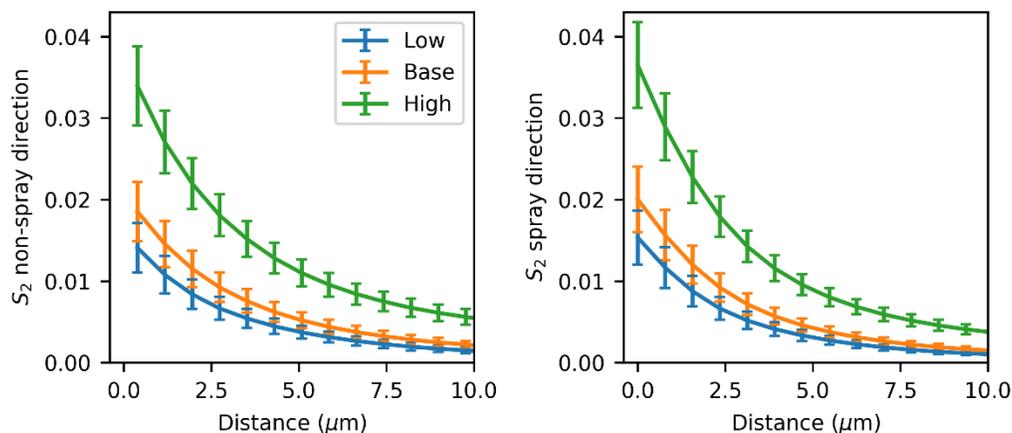


Figure 10. Summary of pore space two-point correlation function.

Another metric of interest is the pore size distribution (PSD). Since identifying individual pores in these microstructures is difficult and ambiguous, we adopt a more general definition of pore size distribution that relies on identifying the fraction of the volume that can be ‘swept out’ with spheres of a given size, such that the spheres are wholly contained in the pore phase. This approach was originally suggested by Munch and Holzer [31] and provides a general measure of the distribution of the local pore space size that does not rely on arbitrary definitions of individual pores. Since spatial correlation is closely related to pore size, the information in this PSD metric is similar to the two-point correlation function S_2 ; however, the pore size distribution has a slightly different focus on local size, as compared to local correlation. The results for the PSD are summarized in ## below, with lines corresponding to mean values across thirty scans, and error bars corresponding to the standard deviation across these scans.

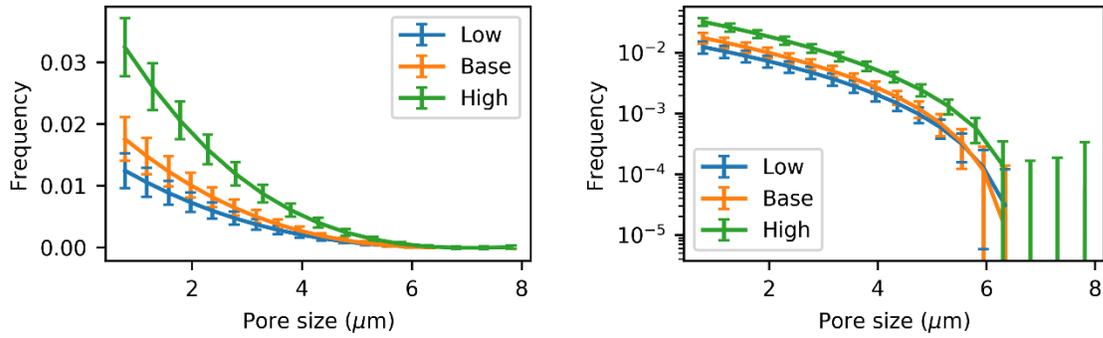


Figure 10. Summary of pore size distribution (see text for definition). Left panel shows linear scale, right panel shows same data on logarithmic scale.

Finally, we quantify pore space connectivity and topology based on a metric suggested by Hilfer, et al. [32] [33]. Additional details are given by Olson et al. [4] and Moore, et al. but we briefly describe the analysis methods here. For a given scan, many cubic subvolumes of size L are randomly selected, and the fraction of subvolumes for a given choice of L that contain a percolating pore phase in each direction is computed. In each subvolume, percolation in a given direction is simply determined based on the presence of at least one connected cluster that spans the subvolume in that direction. For purposes of defining connected clusters, a simple connected component labeling algorithm [34] is applied to each subvolume. Results are summarized in Figure 11 as probability of percolation in each direction as a function of subsample size.

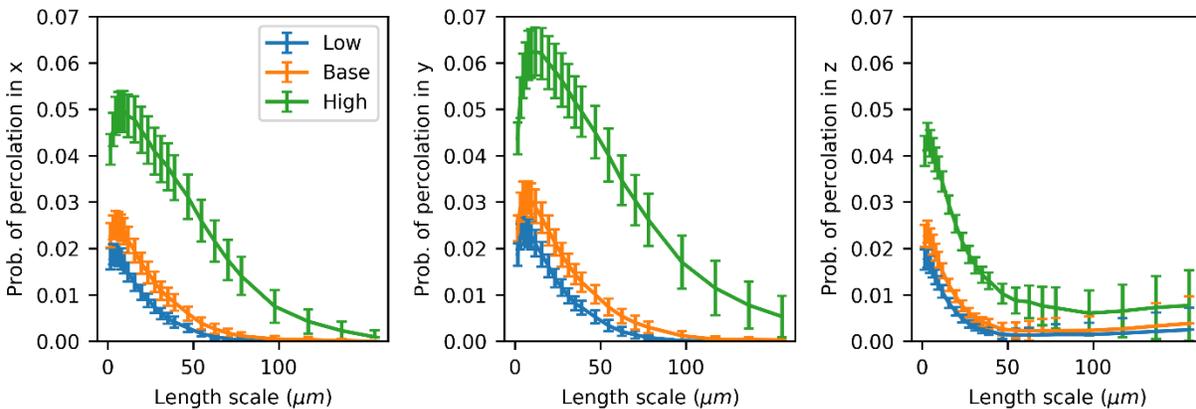


Figure 11. Probability of percolation as a function of length scales in all three directions. The z direction corresponds to the sample spray direction.

4. RESULTS

In this section, we present CNN segmentation results from the various training methods using distinct combinations of real and synthetic one- and two-material examples as training data.

4.1.1. One material thermal spray data

We performed experiments and gathered metrics for models trained on thermal spray scans consisting of two classes to segment – either material or pore. We train our models on both synthetic and real examples, and in the case of real one material thermal spray scans, we train on random walk segmentations. Once finished, we assess the overall quality of our model’s segmentations in the different cases.

4.1.1.1. Trained with Random walk labels only

We trained a model solely on real examples and labels generated via random walker segmentation. We used twenty-three scans for training, two scans for testing, and held five scans out for post-training inference and analysis.

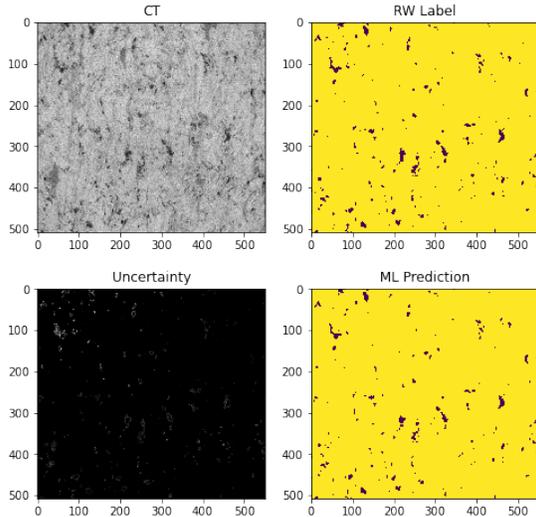


Figure 12.: Sample cross section of held out CT scan (top left), associated random walk label (top right) and our model’s predictions (lower right) and corresponding uncertainty map (lower left) for one material model trained on random walk labels.

Acc to random walker	98.70
False positive rate	0.38%
False negative rate	28.02%
Porosity (random walker)	0.0281
Porosity (model)	0.0226

Table 1: Mean statistics across all scans in the test set. Accuracy with respect to random walk labels, mean false positive and false negative rates across scans of both types, and average porosity for random walk segmentations and the predictions from our model.

4.1.1.2. Trained with synthetic data only

We trained a model on synthetic one material CT scans and their respective generated ground truth labels. In this case, our training set consisted of eleven such scans, and our validation set consisted of two such scans. Once trained, we used our model to perform inference on nine real examples.

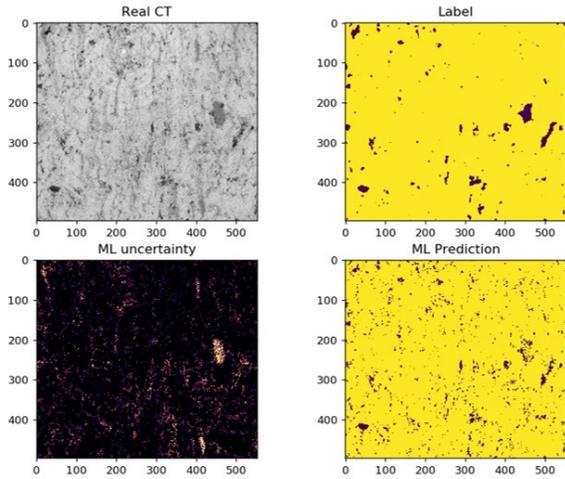


Figure 13.: Sample cross section of held out CT scan (top left), associated random walk label (top right) and our model's predictions (lower right) and corresponding uncertainty map (lower left) for one material model trained on synthetic data.

Sample	Accuracy
1	94.88%
2	95.05%
3	92.21%
4	95.19%
5	95.98%
6	95.38%
7	95.06%
8	95.03%
9	95.05%
Mean	94.87%

Table 2: Prediction accuracies with respect to test set examples for one material model trained on synthetic data.

4.1.1.3. Trained with synthetic data then refined with real CT examples

We trained a model using the full pipeline. Initially, we train the model on two material synthetic examples, using an identical experimental setup to that in 4.1.1.4. Afterward, we continue training on real data, using an identical experimental setup to that in 4.1.1.2 with both naïve and random walker labels but initializing the weights to those of the model trained on synthetic data.

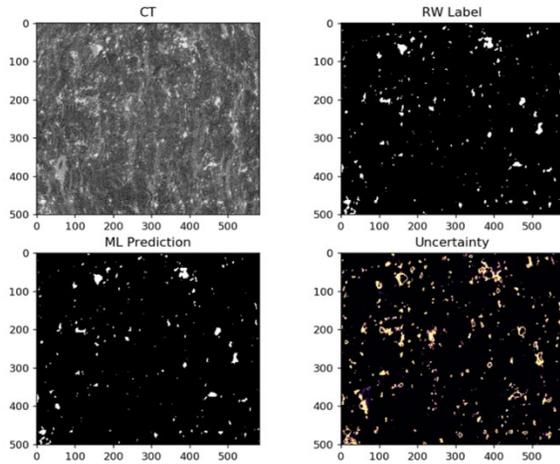


Figure 14: Sample cross section of held out CT scan (top left), associated random walk label (top right) and our model's predictions (lower left) and corresponding uncertainty map (lower right) for the full pipeline one material model.

Acc to random walker	98.80%
False positive rate	0.55%
False negative rate	20.27%
Porosity (random walker)	2.31%
Porosity (model)	2.22%

Table 3: Mean statistics across all scans in the test set. Accuracy with respect to naïve and random walk labels, mean false positive and false negative rates across scans of both types, and average porosity for naïve labels, random walk segmentations, and the predictions the full pipeline one material model produces.

4.1.2. Two material thermal spray data

We performed experiments and gathered metrics for models trained on thermal spray scans comprised of two metallic phases to segment in addition to the pores – three classes in all. We train models on different combinations of the three labeling methods (synthetic, naïve grayscale intensity thresholding, random walk) and assess the overall quality of their segmentations.

4.1.2.1. Trained with Naïve labels only

We trained a model on real CT scan examples and grayscale intensity (naïve) labels. We used two scans for training, one scan for validation, and held out one scan for post-training inference and result calculation.

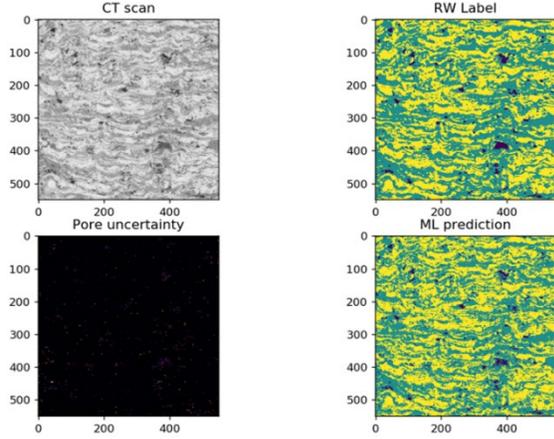


Figure 15: Sample cross section of held out CT scan (top left), associated random walk label (top right) and our model's predictions (lower right) and corresponding uncertainty map (lower left) for two material model trained on naïve labels.

Acc to naïve	98.15%		
Acc to rw	93.08%		
	Uncertainty mean	Min	Max
Overall	0.0005±0.0053	0.0	0.4897
Pore	0.0006±0.0063	0.0	0.4897
Phase 1	0.0008±0.0064	0.0	0.4867
Phase 2	0.0002±0.0014	0.0	0.4897

Table 4: Prediction accuracies with respect to naïve and random walk labels, and uncertainty means, standard deviations, minimums and maximums, both overall and with respect to individual classes for two material model trained on naïve labels.

4.1.2.2. Trained with Random walk labels only

We trained a model on real CT scan examples and labels generated via random walker segmentation. We used two scans for training, one scan for validation, and held out one scan for post-training inference and result calculation.

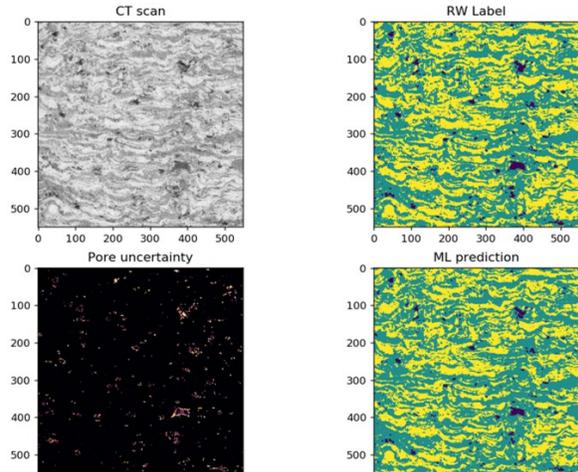


Figure 16: Sample cross section of held out CT scan (top left), associated random walk label (top right) and our model's predictions (lower right) and corresponding uncertainty map (lower left) for two material model trained on random walk labels.

Acc to naïve	91.90%		
Acc to rw	95.32%		
	Uncertainty mean	Min	Max
Overall	0.0034±0.0120	0.0	0.4308
Pore	0.0017±0.0121	0.0	0.4308
Phase 1	0.0051±0.0144	0.0	0.4051
Phase 2	0.0033±0.0086	0.0	0.4215

Table 5: Prediction accuracies with respect to naïve and random walk labels, and uncertainty means, standard deviations, minimums and maximums, both overall and with respect to individual classes for two material model trained on random walk labels.

4.1.2.3. Trained with both Naïve and Random walk labels

We trained a model on real CT scan examples and labels generated via both grayscale intensity thresholding and random walker segmentation. Our training and validation set consisted of two copies of each scan – one with the naïve label and one with the random walker label. We used two scans for training, one scan for validation, and held out one scan for post-training inference and result calculation.

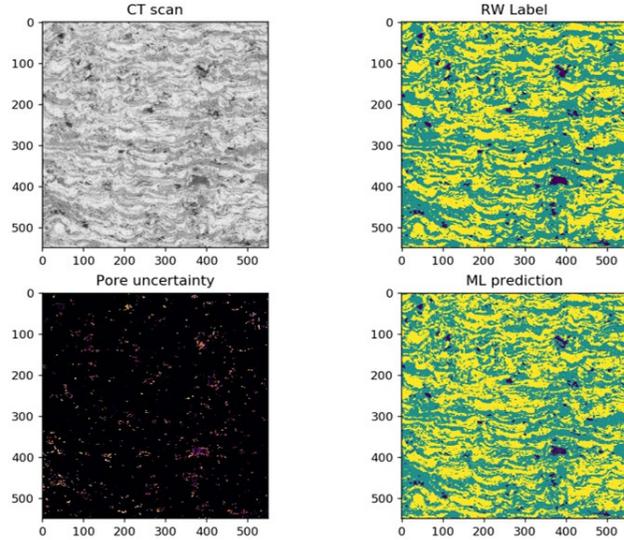


Figure 17: Sample cross section of held out CT scan (top left), associated random walk label (top right) and our model's predictions (lower right) and corresponding uncertainty map (lower left) for two material model trained on both naïve and random walk labels.

Acc to naïve	97.50%		
Acc to rw	94.50%		
	Uncertainty mean	Min	Max
Overall	0.0017 ± 0.0088	0.0	0.3122
Pore	0.0017 ± 0.0106	0.0	0.3122
Phase 1	0.0030 ± 0.0107	0.0	0.3122
Phase 2	0.0009 ± 0.0021	0.0	0.0897

Table 6: Prediction accuracies with respect to naïve and random walk labels, and uncertainty means, standard deviations, minimums and maximums, both overall and with respect to individual classes for two material model trained on both naïve and random walk labels.

4.1.2.4. Trained with synthetic data only

We trained a model on synthetic two material CT scans and their respective generated ground truth labels. In this case, our training set consisted of eleven such scans, and our validation set consisted of two such scans. Once trained, we used our model to perform inference on five real examples.

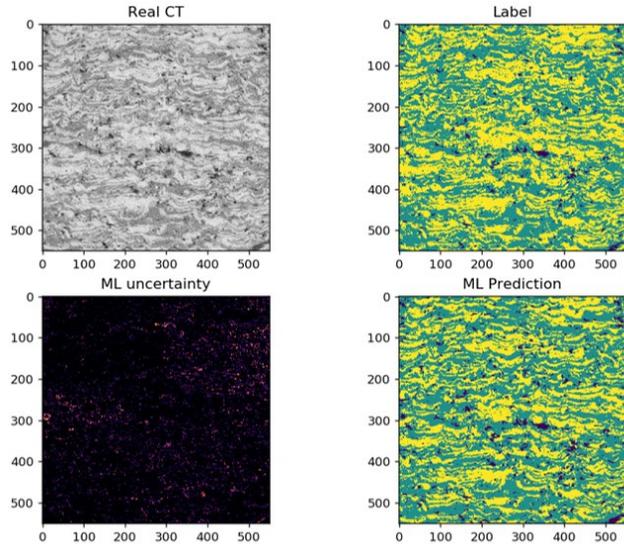


Figure 18: Sample cross section of real CT scan (top left), associated random walk label of that scan (top right) and our model's predictions and uncertainties for two material model trained on synthetic data.

Sample	Accuracy
1	88.94%
2	86.17%
3	86.68%
4	88.75%
5	89.02%
Mean	87.91%

Table 7: Prediction accuracies with respect to test set examples for two material model trained on synthetic data.

4.1.2.5. Trained with synthetic data then refined with real CT examples

We trained a model using the full pipeline. Initially, we train the model on two material synthetic examples, using an identical experimental setup to that in 4.1.2.4. Afterward, we continue training on real data, using an identical experimental setup to that in 4.1.2.3 with both naïve and random walker labels but initializing the weights to those of the model trained on synthetic data.

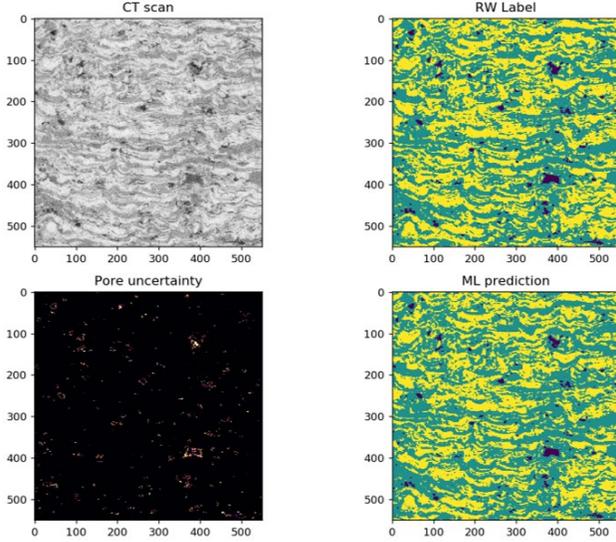


Figure 19: Sample cross section of held out CT scan (top left), associated random walk label (top right) and our model's predictions (lower right) and corresponding uncertainty map (lower left) for the full pipeline two material model.

Acc to naïve	92.5%)		
Acc to rw	96.00%		
	Uncertainty mean	Min	Max
Overall	0.0028 ± 0.0101	0.0	0.4609
Pore	0.0011 ± 0.0094	0.0	0.4609
Phase 1	0.0042 ± 0.0121	0.0	0.4571
Phase 2	0.0030 ± 0.0081	0.0	0.4188

Table 8: Prediction accuracies with respect to Naïve and random walk labels, and uncertainty means, standard deviations, minimums and maximums, both overall and with respect to individual classes for the full pipeline two material model.

4.1.3. Trained with binarized labels

We took our one material full pipeline model from 4.1.1.3 and ran inference on several two material examples for which naïve or random walker segmentation labels were available. We binarized these labels to two classes – material and nonmaterial – and evaluated the model's performance with

respect to both naïve labels (seven examples in total) and random walker labels (four examples in total).

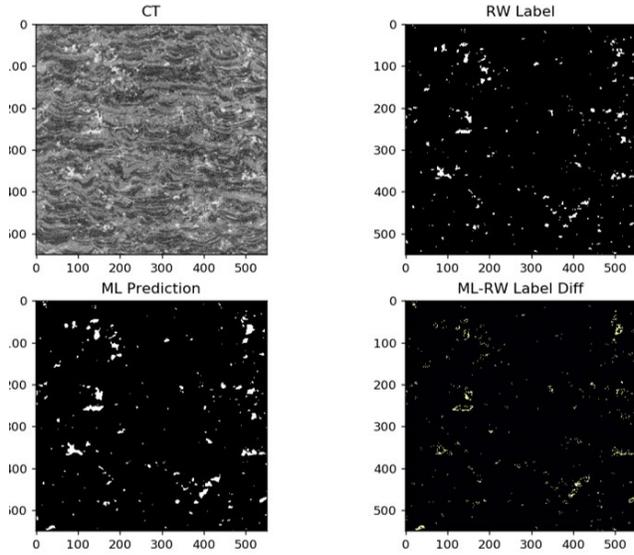


Figure 20: Sample cross section of two material CT scan (top left), associated random walk label (top right), our model's prediction (lower left) and the difference between ML and RW (lower right) for the full pipeline one material model.

Acc to naïve	98.93%
Acc to rw	98.45%
False positive rate	0.73%
False negative rate	15.62%
Porosity (naïve)	0.0238
Porosity (random walker)	0.0248
Porosity (model)	0.0258

Table 9: Mean statistics across all scans in the test set. Accuracy with respect to naïve and random walk labels, mean false positive and false negative rates across scans of both types, and average porosity for naïve labels, random walk segmentations, and the NN model.

5. DISCUSSION

In this section, we present a comparison of the results from our CNN model to labels, and discuss the impact of geometric uncertainty estimation on analysis of relevant quantities of interest.

5.1. Comparison of Deep Learning predictions with labels

Herein, we discuss the accuracy of our segmentations and compare the prediction maps of our model to those of the three methods on which the model was originally trained.

5.1.1. Qualitative comparison

Across all experiments, we observe that our models appear to segment the original image relatively faithfully – though there are still some obvious differences between models. Some of the most prominent differences originate from the models trained on synthetic data alone (4.1.1.4 and 4.2.2.4). In these cases, the model tends to predict a much higher proportion of pore voxels in comparison to other approaches. This discrepancy highlights the differences between real and synthetic examples, and points to the importance of training on real segmentations in order to ensure that the resulting outputs are accurate. Notably, although the two-material synthetically trained model tends to overpredict pore voxels, the metallic phase segmentations appear to be relatively consistent with the segmentation produced by the random walker algorithm.

The two-material model trained on naïve labels (4.2.2.1) produces pore segmentations that much more closely match the corresponding random walker labels, but for larger pores, these predictions tend to be noticeably spottier in comparison to the random walk labels used for comparison. This illustrates some of the same shortcomings the intensity thresholding algorithm possesses – namely, the lack of locality in the resulting segmentations.

In comparison, models which include the random walker segmentations (4.1.1.1, 4.1.1.3, 4.1.2.2, 4.1.2.3, 4.1.2.5, 4.1.3) tend to avoid these issues – producing segmentations of visually high quality.

5.1.2. Prediction characteristics

We analyze the complexity of the predictions our model is making, and whether our model is performing grayscale thresholding (as in the naïve case) or something more sophisticated (as in the random walker segmentation case). We took a two-material example and constructed a histogram of the grayscale values and the corresponding classes for: (a) the naïve label, (b) the random walker label, and (c) the label produced by our pipelined two-material model.

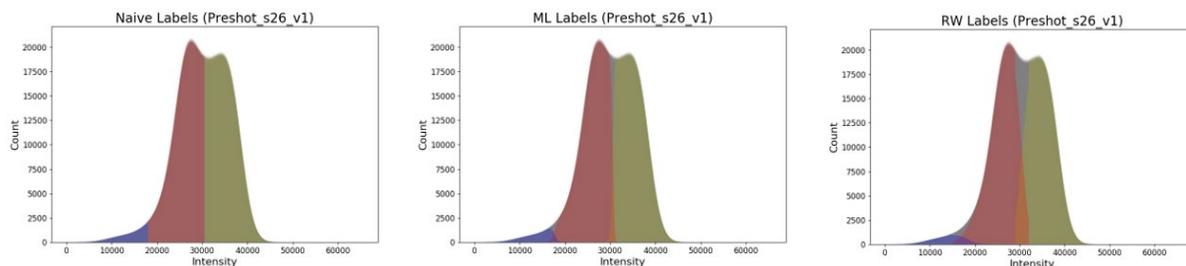


Figure 21: Grayscale histograms with the number of voxels predicted to belong to each class. Pores are shown in blue, material one is shown in red, and material two is shown in yellow.

We observe that the class distributions for the pipelined model overlap – providing evidence that our model’s decision-making process is more complex than intensity thresholding.

5.1.3. Accuracy to labels

We now estimate each model’s performance based on the average accuracies with respect to the random walk labels across each experiment’s testing set. Among the one material examples, the full pipeline model (4.1.1.3) outperforms the real data model (4.1.1.1) slightly, with an accuracy of 0.9880 as opposed to the previous accuracy of 0.9870. Another point of note is the porosity, for which there is far less of a disparity between the random walker labels and model predictions in the pipeline case (0.231 and 0.222, respectively) than in the real data case (0.0281 and 0.226, respectively).

Among the two material examples, we again observe that the full pipeline model (4.1.2.5) performs the best out of all of the experiments, with an accuracy of 0.96 across the testing set, followed shortly by the model trained on random walker labels (4.1.2.2), with an overall accuracy of 0.9532.

5.1.4. Uncertainty characteristics

In order to examine the role that uncertainty plays in our models, we calculate the average uncertainties for voxels in which the predicted classes were either correct or incorrect with respect to the labels using the one material (4.1.1.4) and two-material (4.1.2.4) synthetically trained models.

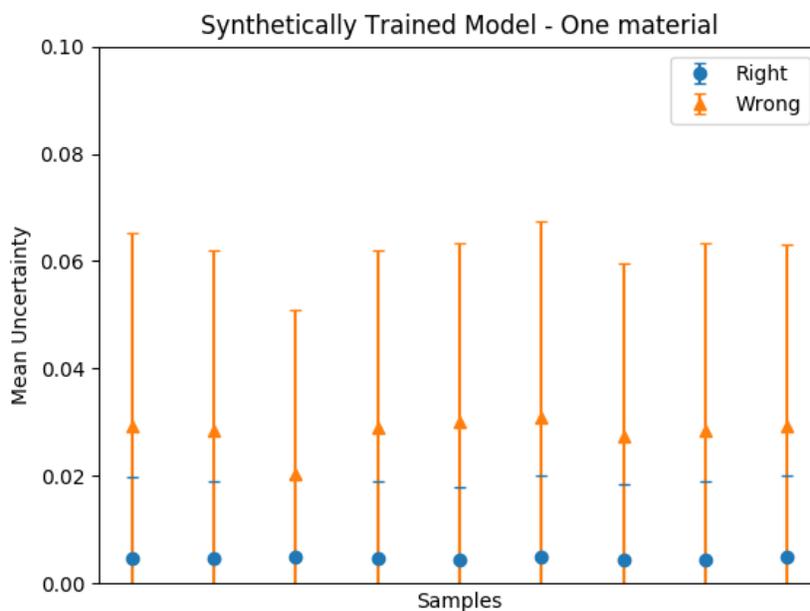


Figure 22: Average uncertainties across correct and incorrect voxel predictions, using one-material synthetically trained model.

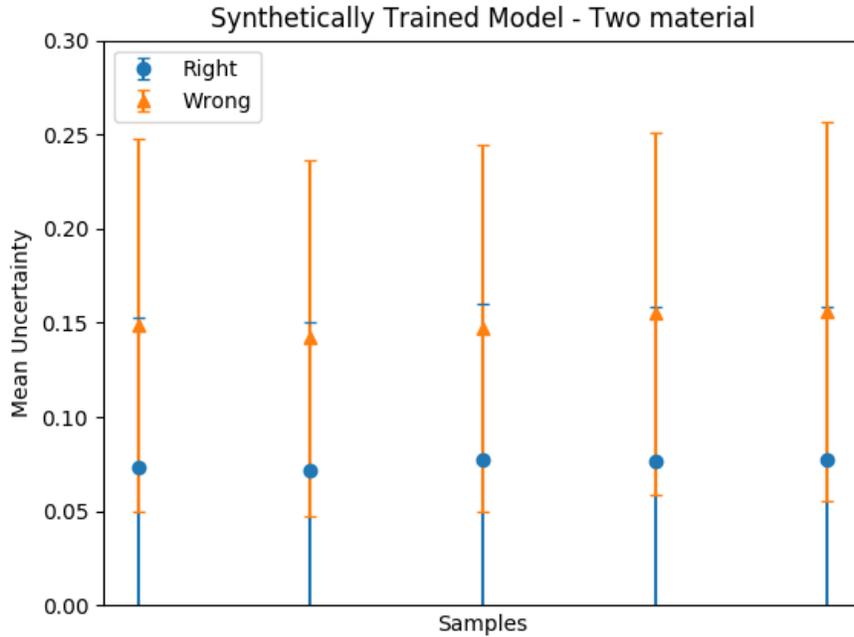


Figure 23: Average uncertainties across correct and incorrect voxel predictions, using two-material synthetically trained model.

In both cases, we observe in Figure 22 and Figure 23 that on average, uncertainties tend to be higher in incorrectly predicted regions than incorrectly predicted regions, which we believe shows promise that our findings are in line with intuitions about uncertainty. Next, we denote a voxel within a volume uncertain if $v_u > \mu_u + \sigma_u$, where v_u is the voxel uncertainty, and μ_u and σ_u are the per-volume uncertainty mean and standard deviation, respectively. We then calculate the proportions of uncertain correct and incorrect voxel predictions for these models.

For both models, we observe that in cases in which the model made a correct prediction, most predictions tend to be certain (though the margin is substantially wider in the one-material case than the two-material case). The confidence distributions are shown in Figure 24 and Figure 25. In cases where the model makes an incorrect prediction, the one-material model tends to be confident in some of these predictions, whereas the two-material model behaves closer to what we would expect – wrong predictions tend to be uncertain. We believe that this may be attributed to the presence of the two phases in the two-material case. While both models tend to predict an overabundance of pore voxels within the real examples, the two-material model manages to predict between the two material phases with relatively high accuracy, and we believe that the majority of “wrong” cases above might involve discrepancy between these two phases rather than that between phase and pore – for which the results tend to be more accurate and hence, uncertainty more reliable. Regardless, we believe that this result warrants closer investigation.

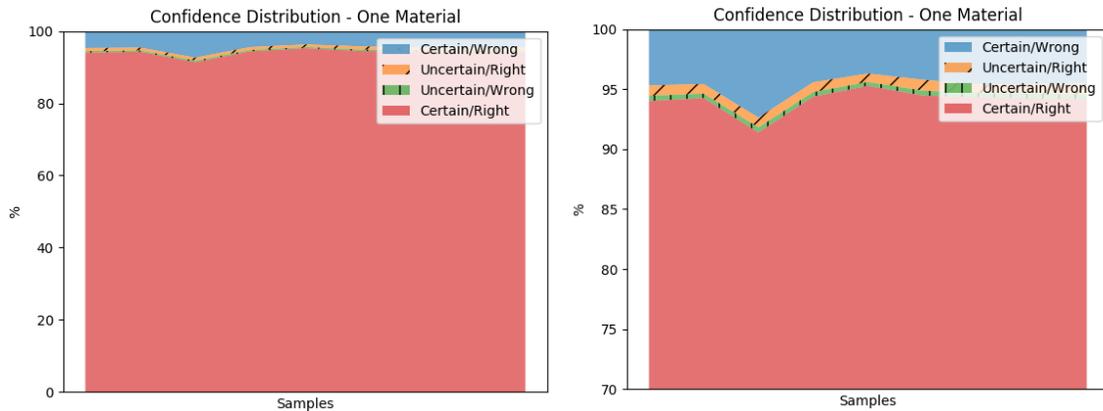


Figure 24: Percentages of uncertain correct and incorrect voxel predictions, using one-material synthetic labels to train model. (Left) Full distribution. (Right) Zoomed to show small regions of distribution plot.

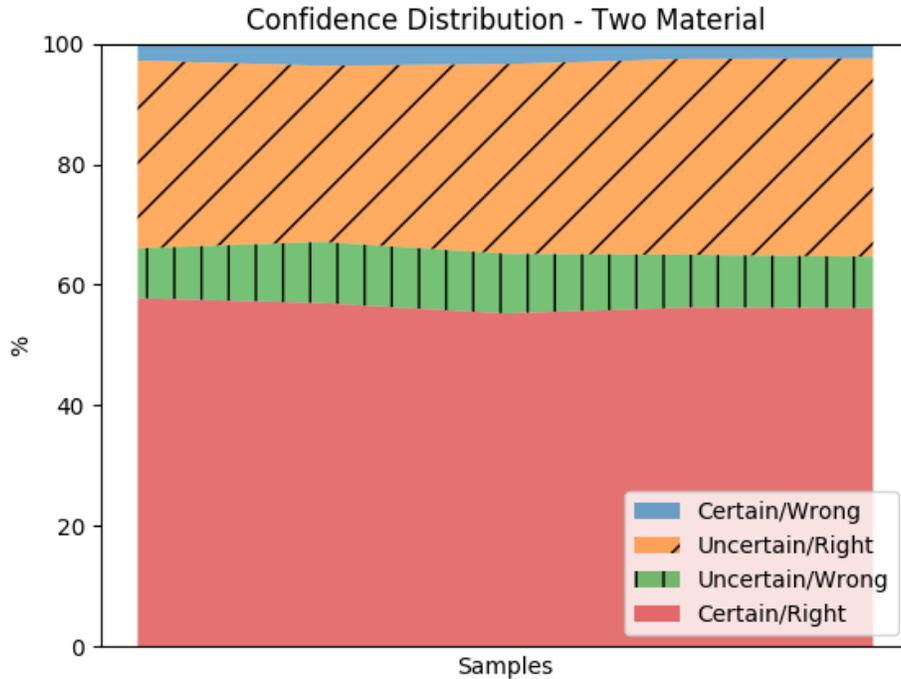


Figure 25: Percentages of uncertain correct and incorrect voxel predictions, using two-material synthetic labels to train model.

5.2. Interpretation of Uncertainty Maps

The uncertainty maps generated by the NN are calculated as the standard deviation over the model's prediction for each voxel from several inference runs each with a different, random subset of neuron activations dropped out of the calculation. Using the method described in Section 3.4, we identify a threshold value for which voxels with predictions whose standard deviation is greater than the threshold are deemed to have uncertain predictions. We interpret this to mean those voxels might represent a pore or a material. By generating 3 versions of the segmentation where (1) all uncertain voxels are labeled as pores (2) all uncertain voxels are labeled as material and (3) all voxels are labeled according to the mean prediction over all inference runs, we produce bounds as well as a nominal prediction for the porosity of the scanned material. This is a conservative bound of the geometry. In none of the NN predictions are all of the 'uncertain' voxels all predicted to belong exclusively to one phase - there is no sampled output that realizes either the upper or lower extreme for the pore geometry. Another potential interpretation of the uncertainty is to take the sampled segmentations as a whole and calculate the variance in porosity measured over the realized, sampled segmentations. We currently choose the conservative approach to ensure the true material porosity is captured within the bounds of our model, but we hypothesize that a tighter bound exists, and research toward that end is left for future work.

5.3. Downstream Analysis and Applications

The metrics discussed in Section 3.4 describe the spatial distribution, local size, and topology of the pore space. Segmentation clearly plays a significant role and is reflected in differences between these metrics. The differences arising due to segmentation are particularly apparent in the “high porosity” bound segmentation, suggesting that uncertainty is skewed such that some voxels that are initially assigned as solid ($\bar{s} < 0.5$) have higher uncertainty as compared to voxels that are initially assigned as pore space. This effect is noticeable in all metrics, but particularly pronounced in the percolation-related metrics; this is not surprising, considering that only a few voxels added or removed to the pore phase can facilitate or break connectivity, resulting in highly nonlinear effects on topology-related metrics such as these. As mentioned previously, the segmentation bounds used here likely represent very conservative (i.e. loose) bounds based on the uncertainty. As such, the differences in the metrics across these different segmentations are likely exaggerated, and the present case represents an overall upper bound on uncertainty.

5.4. Future Work

There are several areas related to this work that are ripe for exploration. We found that synthetic training examples improved performance of the deep learning model on the segmentation of real CT scans. While we experimentally determined a set of parameters that generated simulated CT scans that were qualitatively similar to the real target CT domain, there are several potential methods that would serve to bring the synthetic examples closer to the real data. Deep learning techniques such as domain transfer [35] could be employed to more closely mimic the noise and artifacts associated with scans produced by a particular CT machine, for example. Neural style transfer approaches [36] could also be applied to improve the similarity of the image textures. If we can reliably train the model on only synthetic data, we can appropriately calibrate the model’s uncertainty since we have ground truth by definition with synthetic training data.

While we have conservatively estimated the geometric bounds of the subject materials, a tighter bound could be pursued by gaining a better understanding of the relative importance between where the various predictions for a particular voxel fall on the number line (i.e. close to the decision boundary between classes) and the variance over the predictions for the voxel. Our current approach considers only the variance, and while based on results from the literature, a study on the distributions of predicted values could inform a tighter bound.

Finally, additional uncertainty metrics have been proposed in the literature [37] that can improve the validation of our uncertainty estimation, and potentially the choice of threshold that determines which voxel predictions are designated as ‘uncertain’ in the volumes of interest.

6. CONCLUSION

We have developed a method to perform automated, replicable segmentation of CT scans of complex materials while characterizing the uncertainty in geometric predictions. Our deep learning model produces viable segmentations along with uncertainty maps that can be used downstream to calculate and understand material properties and to bound quantities of interest. Leveraging simulations of target materials as well as synthetic CT software enables pretraining of the NN to recognize features typical of the materials before refining NN weights with examples of real data. Uncertainty maps are generated by performing inference multiple times on the same input example, with stochasticity introduced into the NN with dropout layers that are active during inference. The variance among the inference runs on the same example characterizes the per-voxel uncertainty in the model's prediction for the classification at each voxel. Bounds on the relative prevalence of each material present in a scanned sample are estimated by shifting all uncertain predictions together to identify the bounds on the amount of a particular material present, enabling estimations of the range of geometry-sensitive quantities in the materials.

REFERENCES

- [1] R. W. Smith and R. D. Fast, "The future of thermal spray technology," *Welding Journal*, vol. 73, no. 7, 1994.
- [2] A. M. Hassanein, "Thermal Effects of Erosion Rates from X-Ray Energy Deposition in ICF Reactor First Walls," *Journal of Nuclear Materials*, vol. 122 & 123, pp. 1459-1465, 1984.
- [3] N. W. Moore, C. C. Battaile, J. B. Carleton, A. J. Olson, T. M. Rodgers, A. Vackel, W. M. Scherzinger, R. Pokharel, D. S. Bolinteanu, H. H. Collis, S. J. Owen, C. D. Ernst, W. L. Davis, G. R. Chantler and W, "Stochastic Shock in Porous Refractory Metals," Sandia Report, SAND2019-14283, Albuquerque, 2019.
- [4] A. J. Olson and D. S. Bolinteanu, "Insights into pore gas content based on pore connectivity analysis in X-Ray CT scans," in *Uncertainty Quantification and Parameter Sensitivity Studies for Shock in Porous, Spray-Formed Materials*, Moore, N.W., et al., Sandia Report, SAND2020-XXXX, 2020.
- [5] T. M. Rodgers, J. Mitchell, A. Olson, D. Bolinteanu, A. Vackel and N. Moore, "Fast Three Dimensional Rules-Based Simulation of Thermal Sprayed Microstructures," *In Preparation*, 2020.
- [6] T. M. Rodgers, D. S. Bolinteanu and N. Moore, "Calibrating Thermal Spray Microstructure Simulations to Experimental Data," Sandia Report. SAND2020-XXXX, 2020.
- [7] M. R. Kracum and et al, "Controlled Atmosphere Plasma Spray Development Efforts for FY20," *In preparation*, 2020.
- [8] N.W. Moore, et al., "Morphological Changes in Porous, Refractory Metals After Pulsed X-Ray Exposure," Sandia Report, SAND2019-11816, 2019.
- [9] F. Milletari, N. Navab and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *3D Vision (3DV), 2016 Fourth International Conference on*, IEEE, 2016.
- [10] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning," in *Proceedings of the 33rd International Conference on Machine Learning*, 2016.
- [11] M. Hesamian, W. Jia and X. He, "Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges," *J Digit Imaging*, no. 32, pp. 582-596, 2019.
- [12] A. Bassiouny and M. El-Saban, "Semantic segmentation as image representation for scene recognition," in *2014 IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 981-985.
- [13] A. Mishra, Y. Aloimonos and C. Fermuller, "Active segmentation for robotics," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2009, pp. 3133-3139.
- [14] S. S. Al-amri, N. V. Kalyankar and K. S. D., "Image Segmentation by Using Threshold Techniques," *ArXiv*, vol. abs/1005.4020, 2010.
- [15] G. B. Coleman and H. C. Andrews, "Image segmentation by clustering," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 773-785, 1979.
- [16] M. M. S. J. Preetha, L. P. Suresh and M. J. Bosco, "Image segmentation using seeded region growing," in *2012 International Conference on Computing, Electronics and Electrical Technologies (ICCEET)*, 2012, pp. 576-583.
- [17] E. Shelhamer, J. Long and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640-651, 2017.

- [18] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *ArXiv*, vol. abs/1505.04597, 2015.
- [19] B. Lakshminarayanan, A. Pritzel and C. Blundell, "Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Red Hook, Curran Associates Inc., 2017, p. 6405–6416.
- [20] C. Blundell, J. Conrebise, K. Kavukcuoglu and D. Wierstra, "Weight uncertainty in neural networks.," in *Proceedings of the 32nd International Conference on Machine*, 2015.
- [21] T. LaBonte, C. Martinez and S. A. Roberts, *We Know Where We Don't Know: 3D Bayesian CNNs for Uncertainty Quantification of Binary Segmentations for Material Simulations.*, arXiv preprint arXiv:1910.10793, 2019.
- [22] W. van Aarle, W. J. Palenstijn, J. De Beenhouwer, T. Altantzis, S. Bals, K. J. Batenburg and J. Sijbers, "The ASTRA Toolbox: A platform for advanced algorithm development in electron tomography.," *Ultramicroscopy*, vol. 157, pp. 35-47, 2015.
- [23] S. van der Walt, S. C. Colbert and G. Varoquaux, "The NumPy array: a structure for efficient numerical computation.," *Computing in science & engineering*, vol. 13, no. 2, pp. 22-30, 2011.
- [24] D. S. Bolintineanu, A. J. Olson, H. Collis, A. Pokharel, A. Vackel, J. D. Ivanoff, J. D. Madison and N. W. Moore, "Quantifying pore space morphology in spray-formed materials using X-Ray microcomputed tomography," In preparation, 2020.
- [25] L. Grady, "Random walks for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1768-1783, 2006.
- [26] S. van der Walt, J. L. Schonberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu and scikit-image contributors, "scikit-image: image processing in Python," *PeerJ*, vol. 2, no. 2167-8359, p. e453, 2014.
- [27] F. Chollet, "Keras documentation," 2015. [Online]. Available: <https://keras.io>. [Accessed 17 March 2020].
- [28] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard and M. Kudlar, "Tensorflow: A system for large-scale machine learning.," in *In 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, 2016.
- [29] S. Torquato and H. W. Haslach Jr., "Random heterogeneous materials: microstructure and macroscopic properties," *Appl. Mech. Rev.*, vol. 55, no. 4, pp. B62-B63, 2002.
- [30] Y. Jiao, F. H. Stillinger and S. Torquato, "Modeling heterogeneous materials via two-point correlation functions: Basic principles," *Physical Review E*, vol. 76, no. 3, p. 031110, 2007.
- [31] B. Munch and L. Holzer, "Contradicting geometrical concepts in pore size analysis attained with electron microscopy and mercury intrusion," *Journal of the American Ceramic Society*, vol. 91, no. 12, pp. 4059-4067, 2008.
- [32] R. Hilfer, "Local-porosity theory for flow in porous media," *Physical Review B*, vol. 45, no. 13, p. 7115, 1992.
- [33] B. Biswal, C. Manwart and R. Hilfer, "Three-dimensional local porosity analysis of porous media," *Physica A: Statistical Mechanics and its Applications*, vol. 255, no. 3-4, pp. 221-241, 1998.
- [34] H. Samet and T. Markku, "Efficient component labeling of images of arbitrary dimension represented by linear bintrees," *IEEE transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 4, pp. 579-586, 1988.

- [35] J. Y. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks.," in *Proceedings of the IEEE international conference on computer vision*, 2017.
- [36] L. A. Gatys, A. S. Ecker, M. Bethge, A. Hertzmann and E. Shechtman, "Controlling perceptual factors in neural style transfer.," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- [37] D. J. Stracuzzi, M. C. Darling, M. G. Chen and M. G. Peterson, "Data-Driven Uncertainty Quantification for Multi-Sensor Analytics," *Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR IX*, vol. 10635, p. 106350N, 2018.

DISTRIBUTION

Email—Internal

Name	Org.	Sandia Email Address
Technical Library	01977	sanddocs@sandia.gov

Email—External (encrypt for OUO)

Name	Company Email Address	Company Name

Hardcopy—Internal

Number of Copies	Name	Org.	Mailstop

Hardcopy—External

Number of Copies	Name	Company Name and Company Mailing Address

This page left blank

This page left blank



**Sandia
National
Laboratories**

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.