Application of Data Analytics and Deep Learning for Improving and Accelerating Large-Scale Ptychographic Imaging Problems

Anthony DiBenedetto[1]
Data Science and Learning, Argonne National Laboratory, Lemont IL 60439 (USA)[1]

This study presents an advanced approach to 3D image segmentation using deep learning techniques, specifically applied to tomographic images of borosilicate glass spheres encased in a polypropylene matrix. We implemented a modified 3D U-Net architecture to process volumetric data. We had a main focus on maintaining spatial context across the z-axis while performing convolutions and pooling operations primarily in the x and y dimensions. Also we employed K-means clustering and depth maps for segmentation, this provided a comprehensive analysis of different methodologies. The model was trained on a dataset of sphere samples with varied volume fractions, scanned at different sample-to-detector distances. Results demonstrate the great use of our approach in accurately segmenting complex 3D structures, with applications in materials science.. The study achieves a Dice coefficient of 0.934, indicating excellent overlap between predicted and ground truth segmentation's. Further, we validated our deep learning model on two additional datasets, achieving consistent performance.

I. INTRODUCTION

A. Background on 3D Image Segmentation

Three-dimensional image segmentation is an important technique used in various fields. Some of these fields includes medical imaging, materials science, and industrial quality control. This involves splitting a 3D image into multiple sections where each section is corresponding to a specific part of the image. Traditional methods of image segmentation, such as thresholding and region growing, sometimes don't perform as well with complex 3D structures and noisy data[1].

In recent years, deep learning approaches, particularly convolutional neural networks (CNNs), have shown to be useful in image segmentation tasks[2]. These methods can learn hierarchical features directly from the data. This ends up often outperforming those traditional algorithms in terms of accuracy[3].

B. Importance in Materials Science

In materials science, accurate 3D image segmentation is crucial for understanding the microstructure and properties of composite materials. For instance, in the case of sphere-filled composites, the spatial distribution, size, and volume fraction of the spheres can significantly influence the material's properties. These include mechanical, thermal, and optical properties[4].

Precise segmentation enables researchers to perform various analyses on data. With accurate segmentation, researchers can simulate and predict the composite's behavior under various conditions, leading to more informed design decisions. Additionally, this information can be used to optimize manufacturing processes.

C. Challenges in 3D Segmentation

Despite the advancements in deep learning, 3D image segmentation presents unique challenges. One significant hurdle is the high computational requirements due to the large

size of 3D datasets. Processing such voluminous data demands substantial computational resources and efficient algorithms.

Another challenge lies in the limited availability of annotated 3D training data. Unlike 2D image datasets, which are relatively abundant, high-quality labeled 3D datasets are scarce.

Maintaining spatial context across all three dimensions while processing the data efficiently is another complex task. Unlike 2D images, 3D volumes require algorithms that can effectively capture and utilize spatial relationships in all directions, which can be computationally intensive.

D. Study Objectives

This study aims to address these challenges and advance 3D image segmentation. Our primary objective is to implement and evaluate a 3D U-Net model specifically designed for segmenting borosilicate glass spheres in a polypropylene matrix. This model is great with handling the unique characteristics of our dataset and the specific requirements of materials science applications.

Additionally, we want to test the effectiveness of K-means clustering and depth maps in 3D image segmentation. This approach combines traditional and advanced image processing techniques, potentially offering a more efficient segmentation pipeline.

Finally, we aim to analyze the model's performance across different volume fractions and scanning conditions. This evaluation will provide insights into the model's robustness and generalizability, which are crucial for its practical application in diverse materials science scenarios.

II. DATASET AND PREPROCESSING

A. Sphere Dataset

Our study utilizes a dataset comprising samples with varied volume fractions of borosilicate glass spheres encased in a polypropylene matrix[4]. The sphere diameters in these samples follow a Gaussian distribution, ranging from 38 to 45 µm. This variation in sphere size adds some complexity to the segmentation task, which is good as it makes it more of a realistic test for our proposed method. To capture different imaging conditions, the samples were scanned using X-ray tomography at two distinct sample-to-detector distances: 25 mm and 60 mm. A total of 2000 projections were acquired for each scan, providing comprehensive angular coverage for high-quality 3D reconstruction. The pixel size was set to 0.65 µm, offering sub-micron resolution that allows for detailed visualization of the spheres and matrix interface.

B. Data Preprocessing

Our preprocessing pipeline is designed to enhance the quality of the input data and prepare it for efficient processing by the neural network. It consists of several key steps, each addressing specific aspects of the data preparation challenge.

We begin by loading the tomographic images using specialized image processing libraries which are optimized for handling TIFF stacks commonly used in tomographic data storage.

We first start with standard normalization. This is done by setting pixel values to the range [0, 1], which ensures consistent input scales for the neural network. This normalization is important for stable training and improved convergence of the deep learning model.

Our initial mask generation process com-

bines two powerful image processing techniques. First, we apply adaptive histogram equalization, which enhances the contrast of the images. Following this, we use Otsu thresholding to automatically determine the optimal threshold value for separating the foreground (spheres) from the background (matrix). This two-step approach provides a initial segmentation that serves as a starting point for our deep learning model.

To handle the large 3D volumes efficiently, we implement a data chunking strategy. The 3D volume is divided into chunks of 10 slices along the z-axis. This approach allows us to process the data in manageable parts while still maintaining spatial context. It enables efficient batch processing and helps in managing memory constraints during both training and inference stages.

III. SEGMENTATION METHODS

A. 3D U-Net

At the core of our approach is a modified 3D U-Net architecture, which we have adapted to address the specific challenges of our segmentation task. The U-Net, originally proposed for 2D biomedical image segmentation[5], has been extended to three dimensions to handle our volumetric data.

Our model is designed to process 3D image chunks, maintaining spatial context across the z-axis while performing convolutions and pooling operations primarily in the x and y dimensions.

The architecture consists of four main components: an encoder path, a bridge, a decoder path, and an output layer. The encoder path is responsible for capturing hierarchical features from the input volume. It consists of four blocks, each containing two 3D convolutional layers followed by batch normalization and ReLU activation. After each block, a max pooling layer reduces the spatial dimensions of the feature maps, retaining the most significant features. Importantly, the pooling operation is performed only in the x and y dimensions (2x2x1) to maintain the spatial context along the z-axis. The number of filters doubles after each pooling operation, starting from 32 and increasing to 256, allowing the network to capture increasingly complex features.

The bridge connects the encoder and decoder paths, consisting of two 3D convolutional layers with 512 filters each, followed by batch normalization and ReLU activation.

The decoder path mirrors the structure of the encoder, with four blocks that progressively increase the spatial dimensions of the feature maps. Each block starts with an up-sampling layer, followed by two 3D convolutional layers with batch normalization and ReLU activation. Skip connections from the encoder to the decoder ensure that fine-grained spatial information is preserved. The number of filters halves after each up-sampling operation, decreasing from 256 to 32.

The final output layer is a 3D convolutional layer with a single filter and sigmoid activation, producing a probability map for segmentation.

B. K-means Clustering

We also employed K-means clustering for 3D image segmentation. This method clusters the voxel intensities along with their spatial coordinates. We standardized the features using StandardScaler and then applied K-means clustering with three clusters. Thresholding was applied before clustering to enhance the distinction between spheres and the matrix.

The process begins by loading the 3D image and analyzing its intensities. A threshold is applied to create an initial mask, which is then used for clustering. K-means clustering segments the image into clusters, which are then reshaped back into a 3D volume for visualization and analysis.

C. Depth Maps

Depth maps were created by calculating the gradient along the z-axis and summing the absolute values. The depth map is smoothed using a Gaussian filter, and a threshold is applied to segment the image. Connected component analysis is then performed to identify individual objects.

This method provides an alternative approach to segmentation by focusing on the depth information, which can be particularly useful for distinguishing objects based on their relative depths within the volume.

IV. TRAINING PROCESS

A. Data Preparation and Model Training

Our dataset was carefully partitioned into training which was (10%) of the total data. This division ensures that we have a substantial portion of data for training and then plenty of data for final testing. To manage memory usage and maintain spatial context, we processed the data in chunks of 10 slices along the z-axis.

The model was compiled using the Adam optimizer, known for its efficiency in handling sparse gradients and its ability to adapt the learning rate for each parameter. We set an initial learning rate of 1e-4, which provides a good balance between convergence speed and stability. We set a maximum of 100 epochs for training and the batch size was set to 4, which was determined to balance between computational efficiency and model performance.

B. GPU Utilization

Our system was configured with four of these GPUs, allowing for parallel processing and significantly reducing training time. To fully utilize this multi-GPU setup, we implemented a distributed training strategy using TensorFlow's capabilities. This approach allows us to distribute the workload across all four GPUs simultaneously, leading to a near-linear speedup in training time. The distributed strategy not only accelerates the training process but also enables us to work with larger batch sizes and more complex models than would be possible on a single GPU.

V. RESULTS AND DISCUSSION

A. Segmentation Performance

A metric we used to gauge the models accuracy was Dice loss. The Dice coefficient is a measure of overlap between two samples and is commonly used to gauge the accuracy of image segmentation. It ranges from 0 to 1, where a Dice coefficient of 1 indicates perfect agreement between the predicted and ground truth segmentations[6].

Our 3D U-Net model demonstrated exceptional performance in segmenting borosilicate glass spheres from the polypropylene matrix. The final Dice coefficient achieved on the testing set was 0.934, indicating an excellent overlap between the predicted and ground truth segmentations. This high score suggests that our model can accurately delineate the boundaries of the glass spheres, even in complex 3D structures.

B. K-means Clustering and Depth Map Segmentation

In addition to the 3D U-Net, we applied K-means clustering and depth map techniques to the segmentation task. Figures 2 and 3 show the results of these methods on the sphere dataset.

The K-means clustering approach effectively segmented the spheres from the matrix, providing a clear distinction between different materials based on intensity and spatial

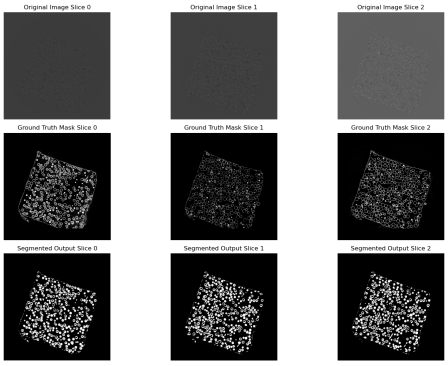

FIG. 1. 3D U-Net segmentation results. (a) Original tomographic image. (b) Ground truth Image. (c) 3D U-Net segmented image.

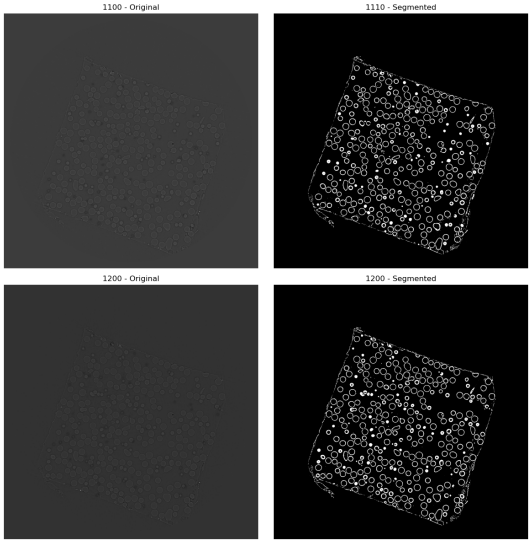

FIG. 2. K-means clustering segmentation results. (a) Original tomographic image. (b) K-means segmented image.

information. The depth map method utilized gradient information along the z-axis, resulting in a segmentation that highlights depth variations within the volume.

C. Validation on Additional Datasets

We validated our 3D U-Net model on two additional datasets: the Tomography Round-Robin datasets and the Lorentz dataset.

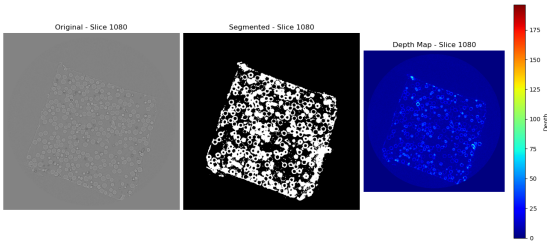

FIG. 3. Depth map segmentation results. (a) Original tomographic image. (b) Depth map segmented image. (c) Depth map.

1. Tomography Round-Robin Datasets

The Tomography Round-Robin datasets consist of two shale samples obtained from the North Sea (sample N1) and the Upper Barnett Formation in Texas (sample B1)[7]. These samples were scanned at three different synchrotron facilities to compare data quality and determine phase proportions and microstructure differences.

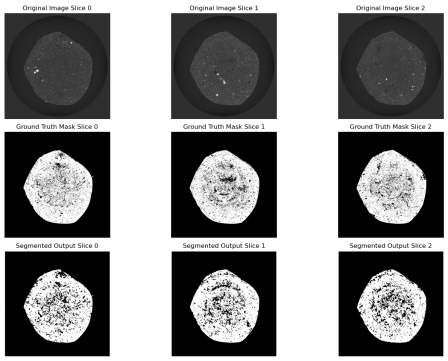

FIG. 4. 3D U-Net segmentation results. (a) Original tomographic image. (b) Ground truth Image. (c) 3D U-Net segmented image.

2. Lorentz Dataset

The Lorentz dataset is part of a demo data collection performed at the Lorentz Workshop. It consists of a wood stick sample scanned at the APS 2-BM-BM beamline.

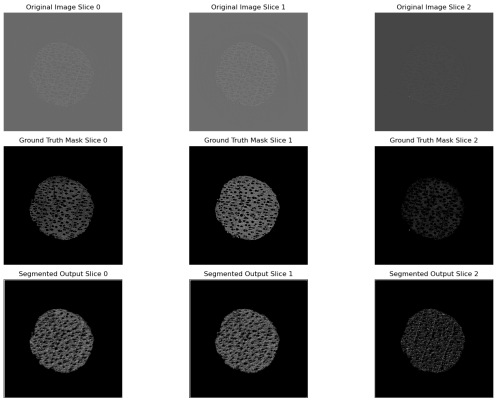

FIG. 5. 3D U-Net segmentation results. (a) Original tomographic image. (b) Ground truth Image. (c) 3D U-Net segmented image.

The consistent performance of our model across these datasets shows its robustness and generalizability. The Dice coefficients achieved on these datasets were comparable to those on the original sphere dataset.

VI. CONCLUSION

This study shows the effectiveness of a 3D U-Net for segmenting borosilicate glass spheres in a polypropylene matrix. Our method, combining adaptive histogram equalization and Otsu thresholding with deep learning, achieves high accuracy across different volume fractions and scanning conditions.

We also explored K-means clustering and depth maps for segmentation. These methods offered valuable insights and demonstrated the flexibility of different approaches in 3D image segmentation.

Three main take aways from our work start with our 3D U-Net model. This model effectively captures the spatial context of the spheres, accurately outlining their boundaries in three dimensions. This is confirmed by a high Dice coefficient of 0.934 and excellent visual segmentations. Second, the model shows strong performance across different volume fractions and sample-to-detector distances. It worked well on multiple datasets. Third, K-means clustering and depth maps provided some newer and interesting ways to approach 3D image analysis. Future work could build on depth maps, combining them with advanced techniques to improve accuracy and applicability.

In addition to the successes of this study, there are several potential directions for future research. One area of interest is the integration of more advanced deep learning architectures. These include transformer models and self-supervised learning techniques to address the lack of annotated 3D datasets. Creating a reliable self-supervised learning framework could significantly enhance the ability to utilize large volumes of unlabeled data. Also exploring the application of our segmentation methods to other materials and varying imaging techniques could help in generalizing and fine-tuning the models. Lastly, working on utilizing the GPUs in a more efficient manner to remove the need to chunk the data.

In conclusion, this study advances the application of deep learning in 3D image segmentation for materials science. By merging traditional image processing with cutting-edge models, we developed a robust, efficient, and accurate method for analyzing complex 3D structures in composite materials. This approach could accelerate research, development, and quality control, aiding the advancement of next-generation composite materials.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to the VFP program for their support of this research and for providing the necessary resources and guidance. I would also like to extend my heartfelt thanks to Dr. Abuomar and Dr. Bicer for their invaluable guidance and unwavering support throughout this study. Their expertise and insights have significantly contributed to the depth and quality of this research.

REFERENCES

[1]D. Shen, G. Wu, and H.-I. Suk, Annual review of biomedical engineering **19**, 221 (2017).

[2]G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, Medical image analysis **42**, 60 (2017).

[3]S.-Y. Fu, X.-Q. Feng, B. Lauke, and Y.-W. Mai, Composites Part B: Engineering **39**, 933 (2008).

[4]TomoBank, "Glass spheres in matrix," `https://tomobank.readthedocs.io/en/latest/source/data/docs.data.spheres.html` (2021), accessed: 2023-07-28.

[5]O. Ronneberger, P. Fischer, and T. Brox, in *International Conference on Medical image computing and computer-assisted intervention* (Springer, 2015) pp. 234–241.

[6]K. H. Zou, S. K. Warfield, A. Bharatha, C. M. Tempany, M. R. Kaus, S. J. Haker, W. M. Wells, F. A. Jolesz, and R. Kikinis, Academic radiology **11**, 178 (2004).

[7]Tomography Round-Robin, "Tomography round-robin data sets," `https://roundrobin.readthedocs.io/en/latest/source/data/docs.data.roundrobin.html` (2021), accessed: 2023-07-28.

Appendix A: Appendix A: Detailed Model Architecture

Table I provides a detailed breakdown of our modified 3D U-Net architecture.

TABLE I. Modified 3D U-Net Architecture

Layer	Output Shape
Input	(10, 256, 256, 1)
Conv3D	(10, 256, 256, 32)
Conv3D	(10, 256, 256, 32)
MaxPooling3D	(5, 128, 128, 32)
Conv3D	(5, 128, 128, 64)
Conv3D	(5, 128, 128, 64)
MaxPooling3D	(2, 64, 64, 64)
Conv3D	(2, 64, 64, 128)
Conv3D	(2, 64, 64, 128)
MaxPooling3D	(1, 32, 32, 128)
Conv3D	(1, 32, 32, 256)
Conv3D	(1, 32, 32, 256)
UpSampling3D	(2, 64, 64, 256)
Conv3D	(2, 64, 64, 128)
Conv3D	(2, 64, 64, 128)
UpSampling3D	(5, 128, 128, 128)
Conv3D	(5, 128, 128, 64)
Conv3D	(5, 128, 128, 64)
UpSampling3D	(10, 256, 256, 64)
Conv3D	(10, 256, 256, 32)
Conv3D	(10, 256, 256, 32)
Conv3D	(10, 256, 256, 1)

Appendix B: Appendix B: Hyperparameter Tuning

Table II shows the hyperparameters we explored during model optimization.

TABLE II. Hyperparameter Tuning Results

Hyperparameter	Range Tested	Optimal Value
Learning Rate	1e-2 to 1e-5	1e-4
Batch Size	2 to 16	4
Number of Epochs	50 to 200	100

Appendix C: Appendix C: Additional Validation Results

Table III presents the Dice coefficients achieved on the additional validation datasets.

TABLE III. Dice Coefficients on Additional Datasets

Dataset	Dice Coefficient
Round-Robin (Sample N1)	0.921
Round-Robin (Sample B1)	0.918
Lorentz Dataset	0.929