Reproducible 3D Segmentation and Reconstruction of Porous Electrodes.

Anonymous Authors[1]

Abstract

This work presents a reproducible pipeline for porosity analysis and 3D reconstruction of porous fuel cell cathode layers, from pFIB-SEM data. A major limitation in current practice is that existing reconstruction methods depend heavily on manual parameter tuning, which leads to inconsistent results across datasets and undermines reproducibility. This challenge is compounded by the difficulty to generate unbiased labeled 3D ground truth, making it difficult to develop and evaluate scalable, data-driven reconstruction approaches. To address this gap, we generate synthetic 3D porous structures using the Porous Microstructure Analysis (PuMA) and Blender softwares, enabling the creation of diverse training datasets with known porosity and geometric properties. Building on these synthetic volumes, our work establishes a reproducible framework that supports accurate porosity estimation, robust 3D reconstruction, and lays the groundwork for integrating future machine learning models into the analysis of both real and synthetic porous electrode materials.

1. Introduction

Image segmentation is used in multiple fields such as medicine, agriculture, and transport (Abramova et al., 2021; Sahin et al., 2023; Pal et al., 2022). More recently, there has been increased interest in this technique in the field of materials science (García et al., 2025; Nagda et al., 2023; Rezaei et al., 2025). One particularly interesting application of image segmentation is the efficient and reproducible prediction of electrode porosity from tomography data, such as for fuel cell electrodes (Ferner et al., 2024). Fuel cells are used to convert the chemical energy from a fuel (such as hydrogen) and an oxidant (such as oxygen) into electricity

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

(Appleby, 1990). They are made up of multiple layers: the cathode and anode layers where the electrochemical reactions happen, a Gas Diffusion Layer (GDL) to facilitate the transport of gases, water, and electrons to the catalyst sites, and a Polymer Electrolyte Membrane (PEM) to prevent the crossing of electrons and gases, while enabling protons to cross. Throughout this work, we will focus on the cathode layer where the oxygen, coming from the cathode GDL, and the protons, coming from the anode, come together to form water and electrons. This reaction is facilitated by an electrode with a high surface area, in other words, an electrode with a large number of pores. To improve the efficiency of fuel cells, researchers are interested in studying the microstructure of those layers (Park et al., 2012). This can be done via imaging methods such as pFIB-SEM as described in the paper by Ferner et al. (2024).

This work aims to improve on the shortcomings of previous image processing methods by building a machine learning pipeline that reconstructs the 3D porous volume from 2D image slices, followed by porosity analysis. Previous methods have been shown to be very sensitive to how parameter values are set (Ferner et al., 2024), our goal is to create a scalable and robust model that accurately depicts the microstructure of the material. Since it is hard to generate unbiased ground-truth labels for real-world data, we will generate our own synthetic data set for training. Therefore, our project contains three main steps: generating a synthetic dataset, training an image processing and reconstruction model, and validating the model pipeline.

2. Prior Work

This work builds directly on the imaging and segmentation techniques introduced by (Ferner et al., 2024). They developed a method that used high-resolution pFIB-SEM to create 3D images of fuel cell catalyst layers and performed porosity analysis through grayscale thresholding. Specifically, after the streak artifacts were removed and the image data were pre-processed, they used intensity thresholding to separate the pore space from the solid material. After processing and cleaning all slices, the binary results were combined into a full 3D volume. Although their method is powerful, the results are highly sensitive

to input parameters that were selected manually, which can lead to inconsistencies when applied across different datasets. Figure 1 shows how sensitive the obtained porosity value of this method is to different input parameters, showing as much as 10% variation in the output porosity value.
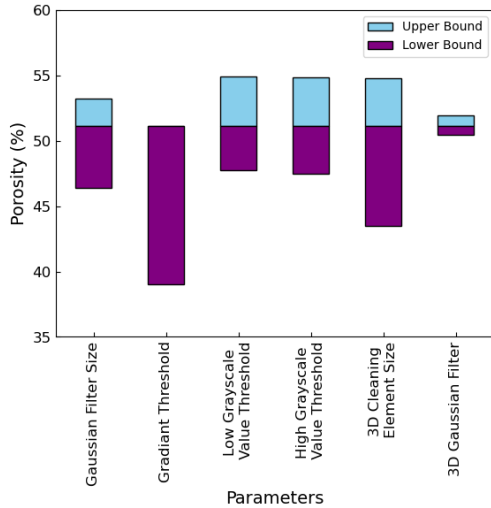

Figure 1. Small changes in processing parameters can significantly affect porosity results. Data digitized from Ferner et al. (2024).

Therefore, to develop a more reproducible method for analyzing the porosity and pore size distribution in the material, we propose a new hybrid approach that combines synthetic data generation using customizable porosity and pore geometry, with machine learning models trained for accurate 3D reconstruction and porosity estimation. By introducing a fully modular and extensible pipeline, our system advances reproducible image-based material characterization and addresses key limitations of previous methods, including parameter sensitivity, lack of scalability, and limited reproducibility.

3. Method

3.1. Synthetic Data Generation

Synthetic 3D porous structures were created using the Porous Microstructure Analysis software (PuMA) (Ferguson et al., 2018). Each structure was generated using the sphere generator, a lognormal particle size distribution, and 150 x 150 x 150 voxels. Blurring and noise were also added to simulate real experimental data. To add further realism, such as lightning and shadows, image rendering of the structure was performed within the 3D software Blender (Blender Foundation, 2023). The angle between the surface normal and the camera was also adjusted to 52° to obtain the same angle as the real data. Additionally, a Python script was

implemented using the Blender API (Blender Foundation, 2023) to systematically slice the 3D model multiple times into 2D stacks, while rendering an image of each slice. This allowed us to emulate how real pFIB-SEM images are taken.

3.2. Image Correction

The pipeline starts when users upload raw 2D/3D images or a stack of slices. The system detects tilt and uneven background illumination, then applies cropping and alignment to isolate the Region of Interest (ROI). The corrected image is saved and passed to the preprocessing module.

3.3. Data Preprocessing

Corrected images undergo noise filtering (e.g., Gaussian or median filters) followed by intensity normalization to ensure consistent pixel value ranges. These steps prepare the data for classification.

3.4. Segmentation Model

Image segmentation was performed using a 2D U-Net (Ronneberger et al., 2015), a widely used encoder–decoder architecture designed for dense prediction. The model utilizes bilinear interpolation in the upsampling block of the decoder, and takes a single raw grayscale slice as input and outputs a pixel-wise binary mask identifying pore versus material. We train the network entirely on our synthetic dataset, which provides clean ground-truth masks and allows systematic control over porosity, noise levels, and imaging conditions.

3.5. Reconstruction and Porosity Computation

We stack the 2D segmentation outputs into a 3D voxel grid and let the model refine this volume into a clean, consistent binary structure. The reconstruction network is a 3D U-Net (Özgün Çiçek et al., 2016) that learns how pores and solids should connect across slices, smoothing out noise and enforcing realistic continuity in the material. The reconstruction of three-dimensional structures from sequential 2D image slices is accomplished through multiple algorithmic approaches, each with distinct characteristics and trade-offs. The system implements seven reconstruction methods to enable comparative analysis and robust mesh generation (see the Annex for a detailed description): Basic, Smoothed, and Hole-Filled Marching Cubes, Direct Voxel Conversion, Convex Hull Construction, Dual Contouring Approach, and Multiscale Reconstruction. Reconstruction quality is assessed through multiple geometric metrics comparing generated meshes against ground truth models. Chamfer distance quantifies bidirectional point-to-surface distances (Fan et al., 2016), Hausdorff distance captures maximum deviations (Karimi and Salcudean, 2020), and volume/area ratios evaluate overall geometric fidelity. Porosity is computed as the

ratio of pore voxels to the total number of voxels in the reconstructed volume. Additionally, porosity calculations from the binary volume data provide domain-specific validation for porous material characterization, enabling quantitative assessment of both surface reconstruction accuracy and internal structure preservation.
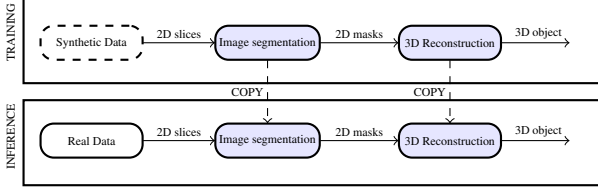

Figure 2. Dataflow diagram of the system

3.6. Experimental data

Real experimental data were obtained from Ferner et al. (2024). This data was obtained by imaging an iridium oxide catalyst layer deposited on Nafion D2020 using pFIB-SEM. The full experimental method can be found in the original publication.
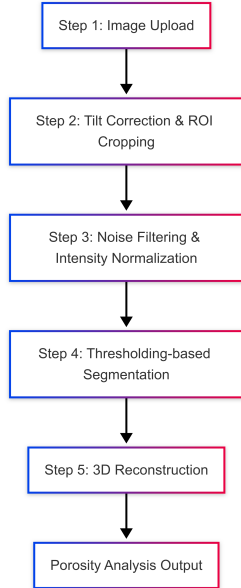

Figure 3. This flowchart shows the full pipeline we used to create our synthetic dataset, designed to match the target porosity and pore size distribution.

3.7. Evaluation Metrics

We evaluate our system using metrics that capture both segmentation quality and 3D reconstruction accuracy. Because real pFIB-SEM data lack exact ground truth, quantitative benchmarking relies on synthetic datasets, while real data are used for qualitative checks.

Pixel-Level Metrics To quantify the segmentation quality of our 2D U-Net, we utilize the Cross-Entropy (CE) Loss, Intersection over Union (IoU), and the Dice score. Cross-Entropy Loss measures pixel-wise classification error, while IoU is computed for both pore and solid classes to evaluate segmentation quality under class imbalance. Finally, the Dice score captures the structural overlap between prediction and ground truth and also acts as an optional regularizer during training.

Volumetric Metrics To measure the 3D reconstruction accuracy of our model, 3D IoU and Smoothness were used as metrics, where 3D IoU quantifies the overlap between predicted and ground-truth voxel volumes, while smoothness is a Laplacian-based measure of slice-to-slice consistency.

4. Results

We evaluated our system across multiple experimental conditions, such as realism and statistical correctness of the synthetic data and the performance of the segmentation model (CE loss, Dice score, IoU, and pixel accuracy). We also looked at the reconstruction model's ability to recover consistent 3D structure and how stable the entire pipeline is when synthetic or real data passes through both stages of our model end-to-end. The key experimental results we obtained are found in the subsections below.

4.1. Synthetic Pore-Size Distribution

To validate the realism of our PuMA-generated (Ferguson et al., 2018) structures, we compare their pore-size histogram against a log-normal fit. Figure 4 shows the resulting distribution.
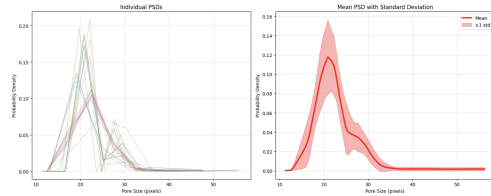

Figure 4. Synthetic pore-size distribution. X-axis: pore diameter (nm). Y-axis: normalized frequency. The distribution closely follows a log-normal curve, consistent with reported fuel-cell cathode measurements (Ferner et al., 2024).

The strong alignment with a log-normal trend indicates that our synthetic samples capture realistic pore statistics and are appropriate for training and evaluation.

4.2. Synthetic and Lightened Slices

Figures 5 and 6 illustrate how we transform clean 2D mask slices into visually realistic inputs using Blender (Blender Foundation, 2023). The Blender (Blender Foundation, 2023) rendering step intentionally introduces shadows, gradients, and textural noise that mimic pFIB-SEM artifacts. Because these images come from a 3D render rather than a flat mask, they appear volumetric and textured, improving model robustness by forcing the network to handle non-uniform illumination and realistic surface variation.
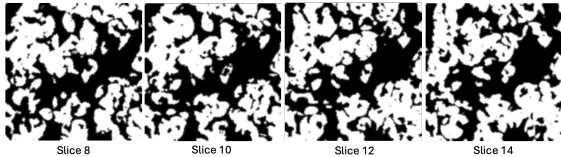

Figure 5. Raw synthetic mask slices extracted along the Z-axis of the PuMA volume(Ferguson et al., 2018). Each slice is a clean binary representation of pores (black) and solid material (white). These masks act as the ground-truth labels during training and are not used as input images.
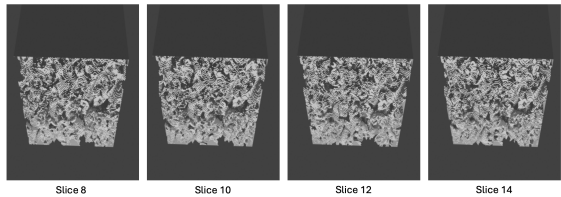

Figure 6. Blender-rendered slices (Blender Foundation, 2023). Instead of 2D masks, Blender renders the full 3D porous structure using a camera tilted at approximately 52°, matching the acquisition angle reported in the original pFIB-SEM workflow. This produces realistic SEM-style shading, depth cues, and surface noise. Although these images correspond to the same Z-locations as the masks above, the 3D lighting and viewing geometry make them appear volumetric rather than flat. These lightened renders are used as model inputs.

4.3. Segmentation Results

Our 2D U-Net (Ronneberger et al., 2015) achieves strong segmentation quality, reaching a validation pixel accuracy of **93.14 %** and a foreground IoU of **90.49 %** by Epoch 180. The model converges smoothly, with the training loss decreasing to **0.0368**. The corresponding training and validation curves are shown in Figure 7. Figure 8 illustrates the qualitative improvement in segmentation quality across different training epochs, while Figure 9 shows a representative comparison between the model prediction and the ground-truth mask.

Although quantitative evaluation is only possible on the synthetic dataset, our trained model also produces high-quality masks on the real pFIB-SEM stacks. Because these experimental samples lack ground-truth annotations, we cannot
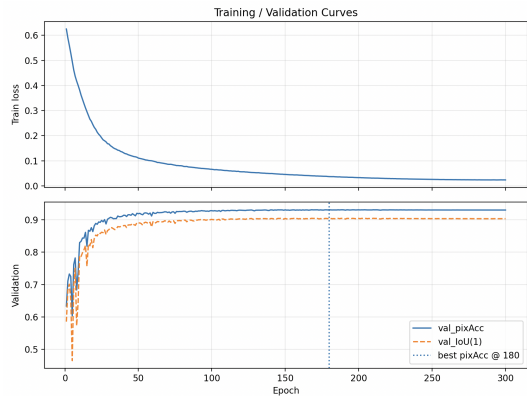

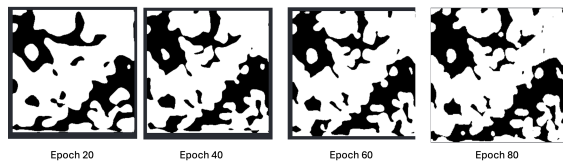*Figure 7.* Training and validation loss curves.

Figure 8. Segmentation result across different training epochs. Early predictions are noisy and overly smoothed; later epochs capture sharper pore shapes and more accurate boundaries.

compute metrics such as pixel accuracy or IoU. Instead, we perform qualitative assessment by visually comparing model predictions against raw slices. Figure 10 is one example of slice validation on the real dataset, the model successfully identifies major pore regions and preserves overall morphology. While this confirms that the model generalizes well to real microscopy data, full validation remains limited by the absence of annotated ground truth.
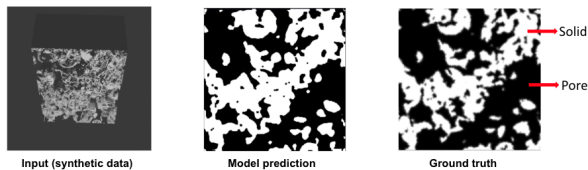

Figure 9. Segmentation quality comparison on synthetic dataset.

These results suggest that the model can reliably separate pore and solid regions, even when the input slices include challenging lighting variations. Additionally, models trained with combined CE + Dice loss consistently outperform CE-only loss, particularly on thin, high-curvature pore boundaries where pixel imbalance makes CE unstable. Dice regularization improves region cohesion, reduces boundary fragmentation, and yields higher IoU scores.

4.4. 3D Reconstruction Results

The comparative evaluation of seven reconstruction methods revealed significant performance differences across accu-
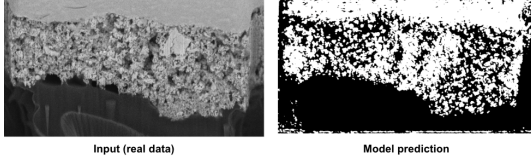
Figure 10. Segmentation quality comparison on real dataset.

racy and porosity estimation metrics. Figure 11 presents the Chamfer distance for each method, where lower values indicate better surface reconstruction accuracy. The Multiscale method achieved the lowest Chamfer distance of **74.04**, followed closely by Voxel Direct **(74.35)** and Marching Cubes variants **(74.72-75.44)**. The Convex Hull method exhibited the second highest error **(79.59)**, which is expected given its inability to capture concave features and internal structures, with Dual Contouring having the highest error **(82.02)**.
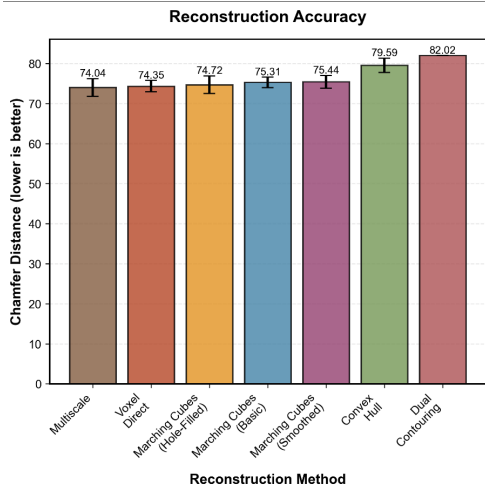

Figure 11. Reconstruction accuracy comparison across seven methods measured by Chamfer distance. Lower values indicate better surface reconstruction quality. Error bars represent standard deviation across multiple datasets.

Figure 12 shows the mean absolute error in porosity estimation, a critical metric for characterizing porous materials. Voxel Direct demonstrated exceptional performance with a MAE of only **0.064**, substantially outperforming all other methods. The Multiscale and Marching Cubes (Basic) approaches achieved moderate errors of **0.270** and **0.240** respectively, while Marching Cubes (Hole-Filled) performed poorly with a MAE of **0.933**, likely due to aggressive hole-filling operations that eliminated true pore structures.

Based on these results, we determined that **Voxel Direct** is the optimal reconstruction method for this application. While its surface reconstruction accuracy **(Chamfer distance of 74.35)** is comparable to other top-performing methods, its superior porosity estimation accuracy makes it uniquely suited for applications requiring precise charac-
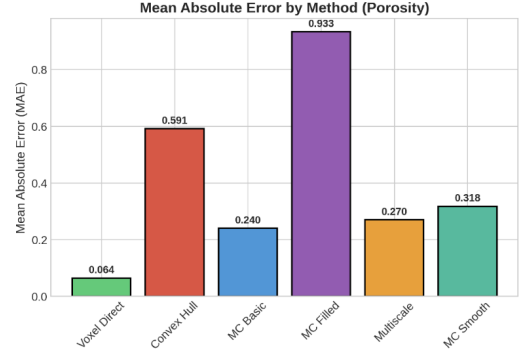

Figure 12. Mean absolute error in porosity estimation for each reconstruction method. Voxel Direct achieves significantly lower error (MAE=0.064) compared to all other approaches, demonstrating superior preservation of internal pore structures.

terization of internal pore structures. The method's direct correspondence between voxels and mesh geometry ensures faithful preservation of volumetric properties, a critical requirement for materials science and biomedical imaging applications.

4.5. Error Analysis

Across experiments, we observed three consistent patterns of errors, which are enumerated below:

Boundary Misalignment. Predicted pore boundaries were sometimes slightly offset or smoothed compared to the ground truth. This occurred because strong illumination gradients reduce edge contrast, and convolutional layers naturally smooth fine details. Future work should incorporate boundary-aware losses or sharper contrast augmentations.

Thin-Pore Under-Segmentation. Very narrow pore channels were occasionally missed. These structures contribute a small fraction of total pixels, so CE and Dice loss assign them low weight. Training with focal-Dice combinations or pore-size–aware sampling could alleviate this issue.

Synthetic–Real Gap. While synthetic data produced strong segmentation performance, real pFIB-SEM slices contain harsher noise and irregular artifacts not fully captured by Blender (Blender Foundation, 2023) lighting. This gap explains minor generalization drops on real samples. Adding domain-randomized noise models and artifact simulations (e.g., drift, charging) is a promising next step.

5. Discussion

During this work, we have successfully implemented a machine learning pipeline to reconstruct the 3D structure from sliced 2D images. The model was trained using synthetic

data, since the data we obtained did not have ground truth. Our innovative approach was able to train a model to **93%** validation accuracy. However, it is difficult to assess the validity of the model output for real images. Indeed, while we have taken considerable steps to replicate the slicing and lighting conditions, there are still key differences between our synthetic images and the real images. Among the imperfections of real data not currently captured by our model is the presence of non-planar slicing, non-cubic shapes, sample charging by the electron beam, unwanted movement during measurement (drift), and different resolutions in the input dataset.

All in all, there are several key aspects from the synthetic dataset that might explain why generalizing to real data is more difficult than anticipated. With the lack of ground truth, strong arguments are necessary to ensure the validity of the model on other datasets. It can be done by analyzing real data output, evaluating key attributes such as porosity, pore-size distribution and material characteristics. The robustness of the model can be further increased by adding variability and noise to the training data to ensure that the model does not overfit to specific parameters.

It is not yet fully clear that the results obtained using this method will generalize reliably to all datasets or accurately reflect the underlying material structure. However, we believe that investing future effort, particularly in diversifying the synthetic data, would be worth trying. These steps would strengthen the model's credibility and significantly enhance its applicability to real-world fuel-cell microstructure analysis.

6. Conclusion

Our work focuses on using pFIB-SEM images of cathode layers to generate a 3D reconstruction of the layer in fuel cells. Due to the absence of ground truth for our data, we developed an innovative approach that relies on synthetic data to train our machine learning pipeline.

The work we have done is two-fold. First, we have crafted and generated a synthetic dataset using PuMA software (Ferguson et al., 2018) to generate porous material and Blender (Blender Foundation, 2023) to add light and perspective, to slice the volume and to render images. Secondly, we used our synthetic dataset to train a machine learning pipeline. We implemented a 2D U-Net to extract information from the raw slices and identify pores, achieving 93% accuracy. Then, we relied on a 3D U-Net to reconstruct the final 3D volume.

The results show great promise in using synthetic datasets to replicate and train machine learning pipelines when data is sparse and incomplete. Future work is needed to ensure that the model does not overfit to the synthetic data and can generalize to produce a reliable results on a variety of real

data.

7. Appendix

Compute Requirements

- **AWS g-family GPU Instance:** A `g5.4xlarge` (or equivalent g-family instance) is sufficient for training the 2D and 3D U-Net models.

- **GPU Memory:** Approximately 24 GB VRAM is adequate for 3D U-Net training at moderate patch sizes.

- **CPU Resources:** Multi-core CPU for Blender-based rendering and PuMA simulations, which are CPU intensive.

- **Storage Requirements:** Approximately 200 GB for storing synthetic volumes, rendered slices, training logs, and model checkpoints.

7.1. 3D Reconstruction methods

Marching Cubes Variants. The primary reconstruction approach employs the marching cubes algorithm, a classical isosurface extraction technique that processes volumetric data by evaluating voxel configurations against a threshold value. Three variants are implemented to address different reconstruction requirements. Basic Marching Cubes operates directly on the raw volume data with a fixed threshold of 127 (mid-grayscale), extracting the isosurface by analyzing each voxel's eight corners and determining triangle configurations based on which corners exceed the threshold. While computationally efficient, this approach is sensitive to noise and produces meshes with visible artifacts from the underlying voxel structure. Smoothed Marching Cubes applies Gaussian filtering with configurable sigma values (typically $\sigma = 2.0$) to the volume data prior to isosurface extraction, reducing high-frequency noise and creating more gradual intensity transitions that result in meshes with smoother surfaces and fewer geometric discontinuities. Hole-Filled Marching Cubes addresses incomplete data through morphological operations by first binarizing the volume, then applying slice-by-slice hole filling using binary flood-fill algorithms, with morphological closing operations further refining the volume by connecting nearby structures and eliminating small gaps. This approach proves particularly valuable for porous materials where imaging artifacts may create false voids.

Direct Voxel Conversion. The voxel-direct method bypasses traditional isosurface extraction by converting the binarized volume directly into a mesh representation. Each occupied voxel in the binary grid is transformed into geometric primitives, producing a blocky but topologically consistent mesh. This approach guarantees volume preservation and

maintains strict correspondence with the original voxel data, though at the cost of surface smoothness.

Convex Hull Reconstruction. For objects with predominantly convex geometries, the convex hull method extracts surface points via marching cubes and computes the minimal convex polyhedron enclosing these points. When dealing with large point sets, the algorithm employs random subsampling (typically 5,000 points) to maintain computational tractability. This method provides the smoothest possible surface for convex structures but cannot represent concavities or internal features.

Dual Contouring Approach. The dual contouring variant implements adaptive thresholding by computing the median intensity of non-zero voxels rather than using a fixed threshold. Combined with moderate Gaussian smoothing ($\sigma = 1.5$), this approach adapts to varying intensity distributions across different datasets. The median-based threshold proves more robust to lighting variations and imaging inconsistencies than fixed thresholding.

Multiscale Reconstruction. The multiscale method employs pyramid-based processing to capture features at different spatial frequencies. Three Gaussian filters with increasing sigma values (0.5, 2.0, and 4.0) generate volume representations at fine, medium, and coarse scales. These volumes are combined through weighted averaging (0.5, 0.3, and 0.2 respectively), emphasizing fine details while incorporating broader structural information. This multi-resolution approach enhances robustness to noise while preserving important geometric features.

References

Valeriia Abramova, Albert Clèrigues, Ana Quiles, Deysi Garcia Figueredo, Yolanda Silva, Salvador Pedraza, Arnau Oliver, and Xavier Lladó. 2021. Hemorrhagic stroke lesion segmentation using a 3D U-Net with squeeze-and-excitation blocks. *Computerized Medical Imaging and Graphics*, 90:101908.

A.J. Appleby. 1990. From sir william grove to today: fuel cells and the future. *Journal of Power Sources*, 29(1):3–11. Proceedings of the Grove Anniversary Fuel Cell Symposium.

Blender Foundation. 2023. Blender - a 3d modelling and rendering package.

Haoqiang Fan, Hao Su, and Leonidas Guibas. 2016. A Point Set Generation Network for 3D Object Reconstruction from a Single Image. *arXiv preprint*. ArXiv:1612.00603 [cs].

Joseph C. Ferguson, Francesco Panerai, Arnaud Borner, and Nagi N. Mansour. 2018. Puma: the porous microstructure analysis software. *SoftwareX*, 7:81–87.

Kara J. Ferner, Janghoon Park, Zhenye Kang, Scott A. Mauger, Michael Ulsh, Guido Bender, and Shawn Litster. 2024. Morphological analysis of iridium oxide anode catalyst layers for proton exchange membrane water electrolysis using high-resolution imaging. *International Journal of Hydrogen Energy*, 59:176–186.

Antonio García, Javier Monsalve-Serrano, Javier Marco-Gimeno, and Carlos Guaraco-Figueira. 2025. Comparative CT scan and U-net segmentation analysis of NMC811 batteries after thermal runaway triggered by various abuse methods and states of charge. *Journal of Power Sources*, 629:236045.

Davood Karimi and Septimiu E. Salcudean. 2020. Reducing the Hausdorff Distance in Medical Image Segmentation With Convolutional Neural Networks. *IEEE transactions on medical imaging*, 39(2):499–513.

Vinit Nagda, Artem Kulachenko, and Stefan B. Lindström. 2023. Image-based 3D characterization and reconstruction of heterogeneous battery electrode microstructure. *Computational Materials Science*, 223:112139.

Kushagra Pal, Piyush Yadav, and Nitish Katal. 2022. RoadSegNet: a deep learning framework for autonomous urban road detection. *Journal of Engineering and Applied Science*, 69(1):110.

Jae-Hyun Park, Sung-Dae Yim, Taeyoung Kim, Seok-Hee Park, Young-Gi Yoon, Gu-Gon Park, Tae-Hyun Yang, and Eun-Duck Park. 2012. Understanding the mechanism of membrane electrode assembly degradation by carbon corrosion by analyzing the microstructural changes in the cathode catalyst layers and polarization losses in proton exchange membrane fuel cell. *Electrochimica Acta*, 83:294–304.

Shahed Rezaei, Kianoosh Taghikhani, Alexandre Viardin, Reza Najian Asl, Ali Harandi, Nikhil Vijay Jagtap, David Bailly, Hannah Naber, Alexander Gramlich, Tim Brepols, Mustapha Abouridouane, Ulrich Krupp, Thomas Bergs, and Markus Apel. 2025. Digitalizing metallic materials from image segmentation to multiscale solutions via physics informed operator learning. *npj Computational Materials*, 11(1):262. Publisher: Nature Publishing Group.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. *Preprint*, arXiv:1505.04597.

Halil Mertkan Sahin, Tajul Miftahushudur, Bruce Grieve, and Hujun Yin. 2023. Segmentation of weeds and crops

using multispectral imaging and CRF-enhanced U-Net. *Computers and Electronics in Agriculture*, 211:107956.

Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. 2016. 3d u-net: Learning dense volumetric segmentation from sparse annotation. *Preprint*, arXiv:1606.06650.