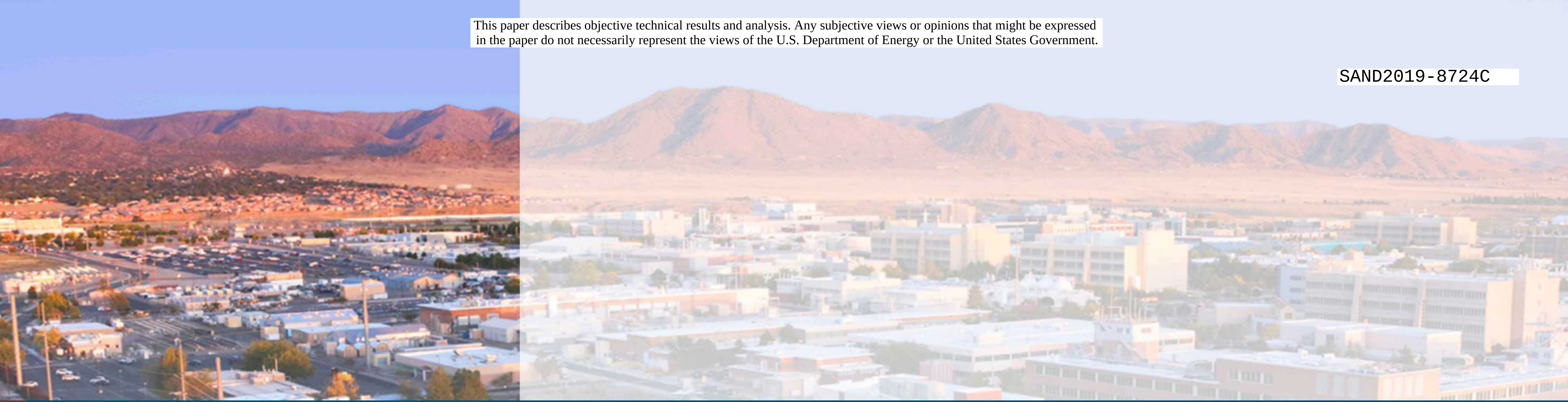
SAND2019-8724C



Deep Learning with Per-Voxel Uncertainty Quantification for Volumetric Segmentation of Battery Electrode Images

Cari Martinez, Kevin Potter, Matthew D. Smith, Scott A. Roberts

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.

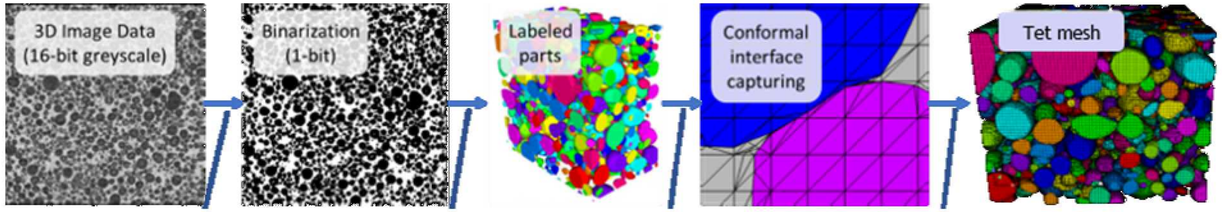
Can we predict the behavior of as-built parts with error bars?

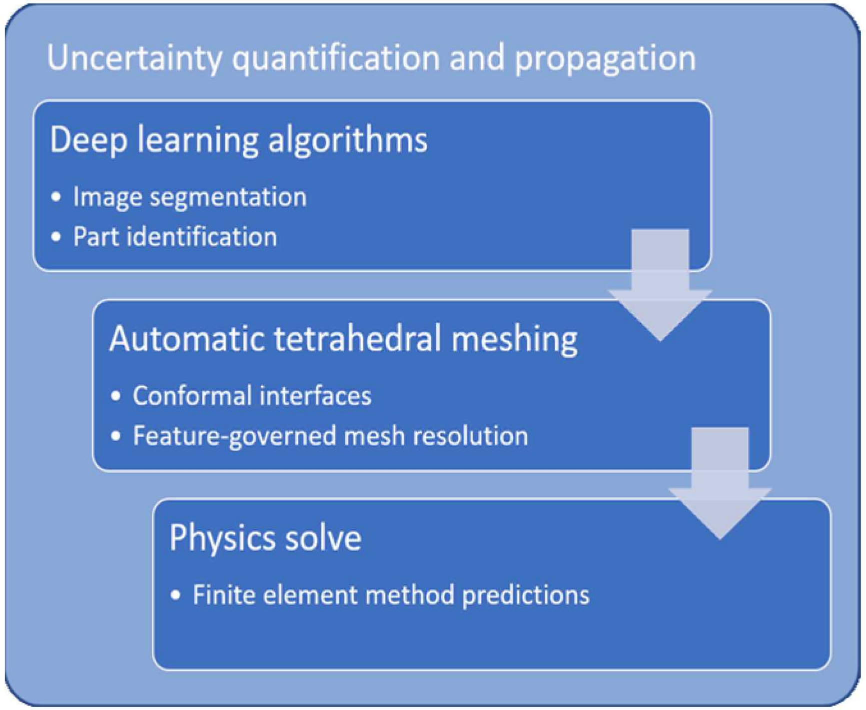
Hypothesis: We can develop an automated and credible image-to-mesh technology that can demonstrate the physics impact of per-unit variability on material, component, or system performance

Objective: We seek to develop a methodology for **automatically**, **efficiently**, and **reproducibly** creating **conformal** finite element meshes from **3D tomography** with **quantified uncertainty**.

Research thrusts – primary science questions:

1. Automatic CT segmentation via Machine Learning (ML)
2. Automatic conformal tetrahedral mesh creation (ATM)
3. Uncertainty quantification and propagation (UQ)





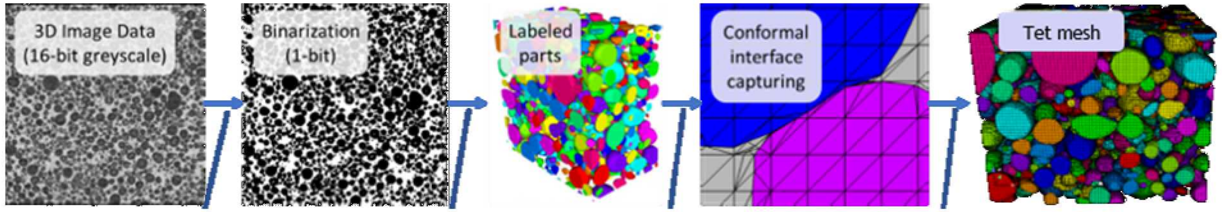
Can we predict the behavior of as-built parts with error bars?

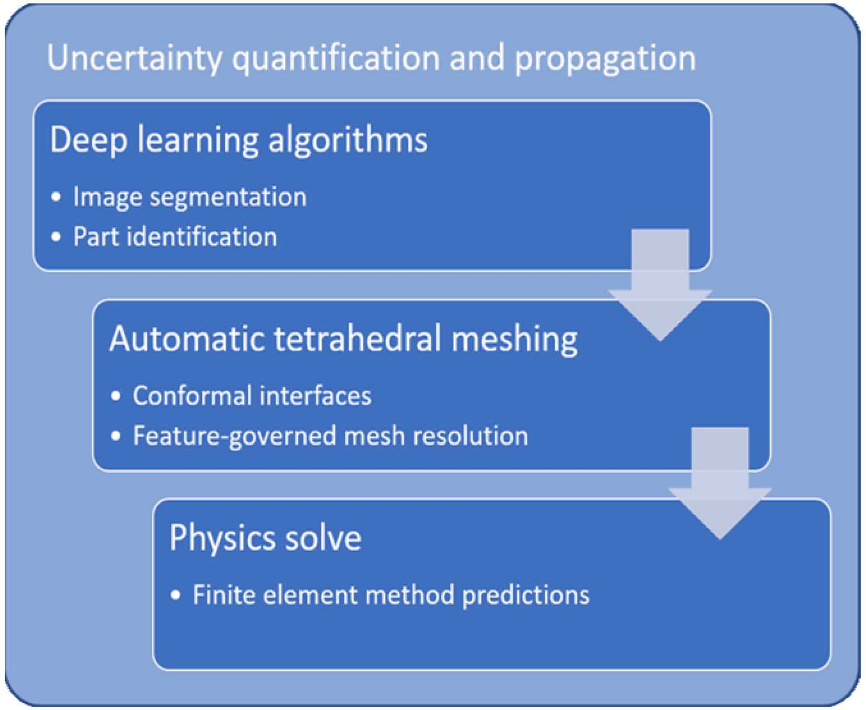
Hypothesis: We can develop an automated and credible image-to-mesh technology that can demonstrate the physics impact of per-unit variability on material, component, or system performance

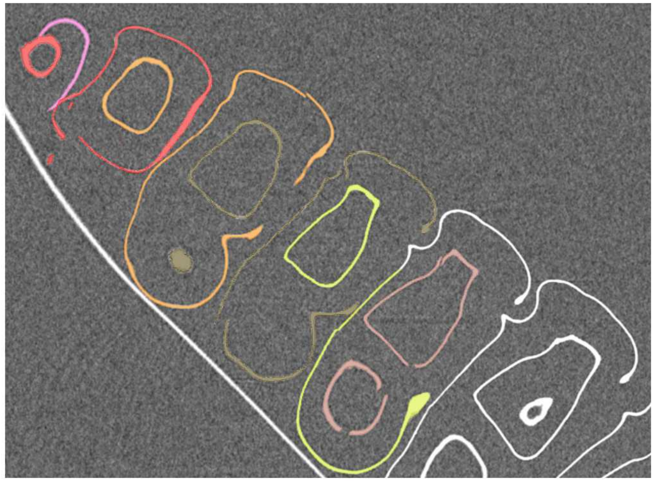
Objective: We seek to develop a methodology for **automatically**, **efficiently**, and **reproducibly** creating **conformal** finite element meshes from **3D tomography** with **quantified uncertainty**.

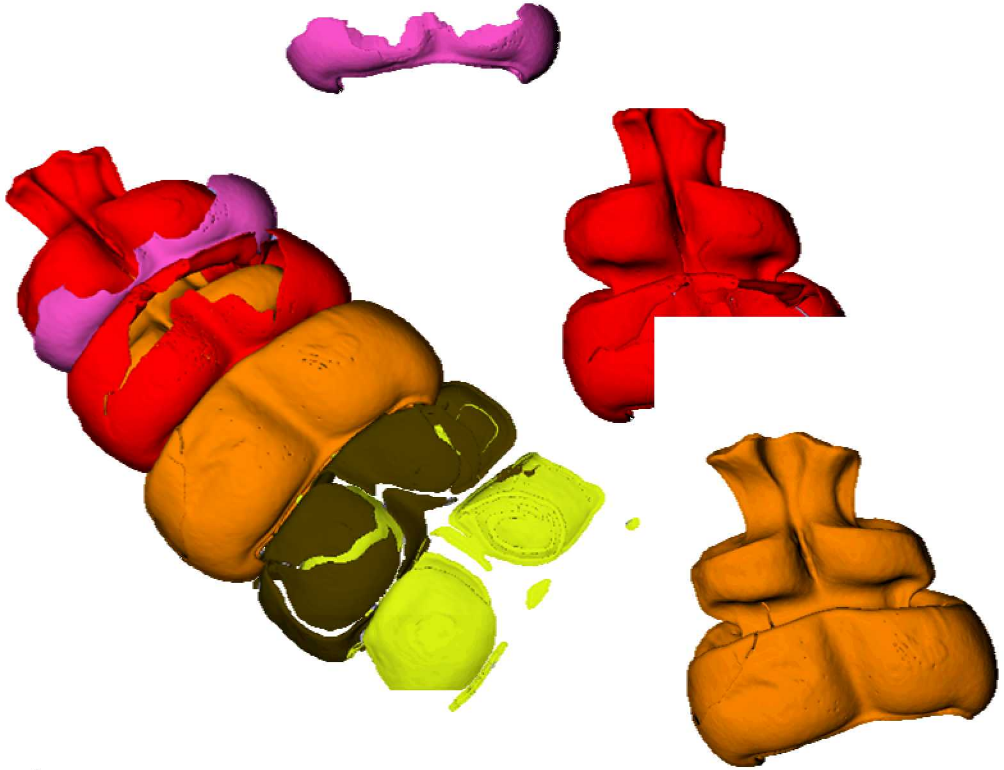
Research thrusts – primary science questions:

1. Automatic CT segmentation via Machine Learning (ML)
2. Automatic conformal tetrahedral mesh creation (ATM)
3. Uncertainty quantification and propagation (UQ)







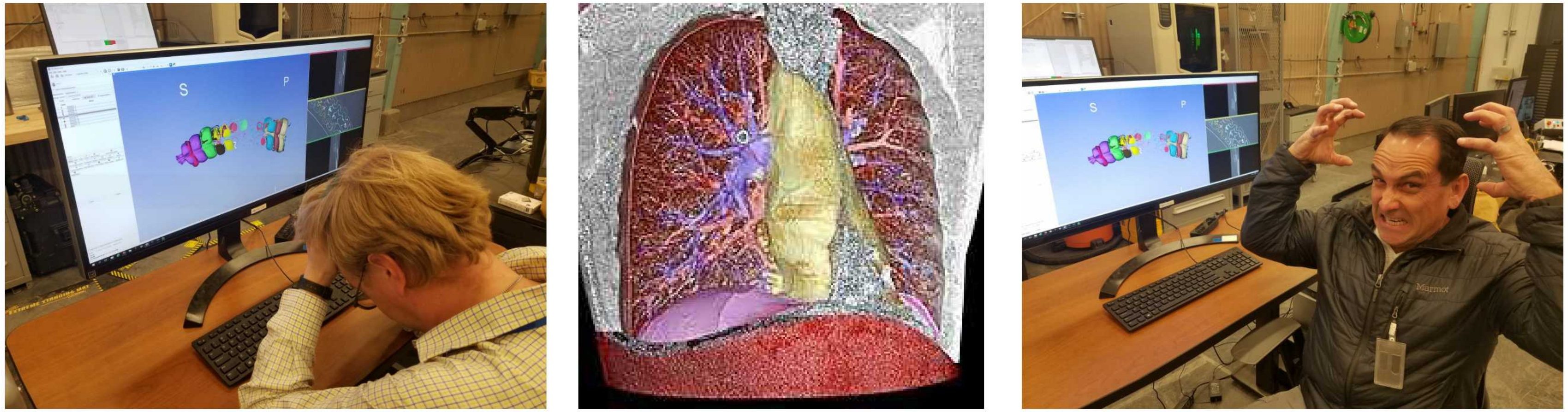


Using ML to save time and effort while improving accuracy

CT SEGMENTATION

CT Segmentation is hard for humans

CT scans must be labeled by component for simulations



https://en.wikipedia.org/wiki/Image_segmentation

Labeling by hand does not scale

Deep Learning algorithms

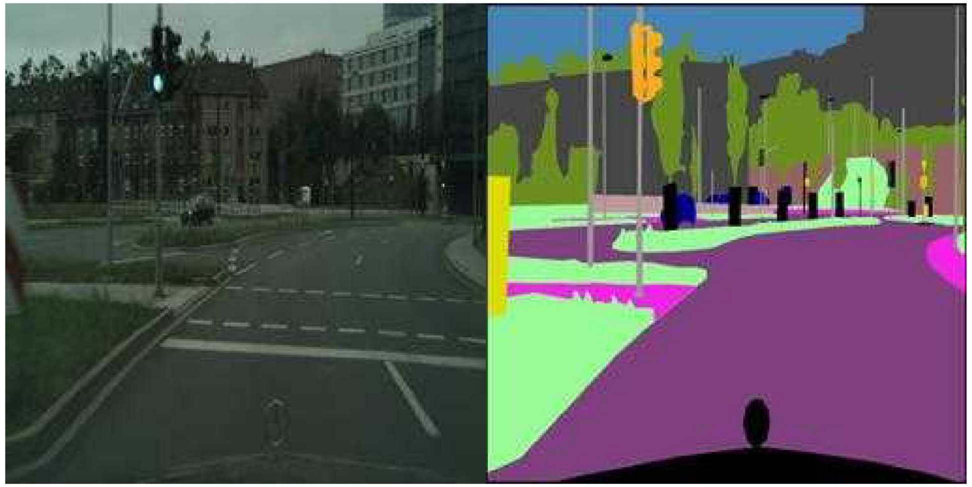
- Each voxel must be labeled by material
- Find any defects/anomalies
- Pass this to a usable form for numerical simulations

Segmentation is a classic computer vision problem

- Image segmentation is well studied
  - Small files
  - Large training sets

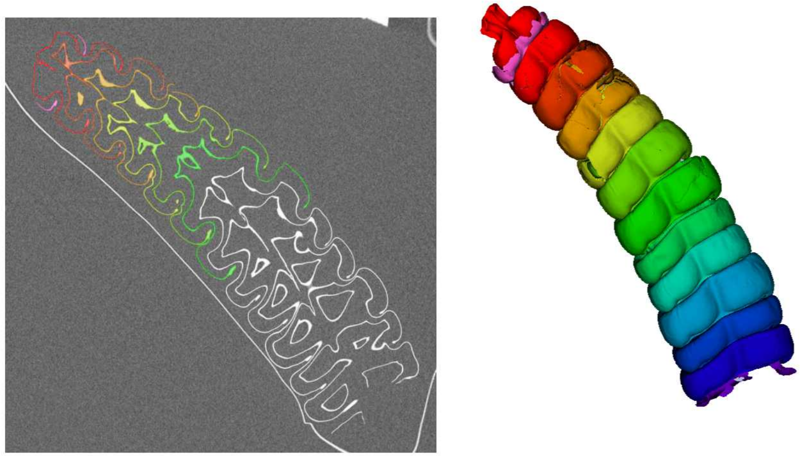
- Volumetric segmentation is different
  - Big data
  - Class imbalance (lots of background)
  - Small training sets with "bad" human labels
  - Humans can't label billions of voxels

- Medical researchers are leading this work toward Deep Learning solutions



https://www.cityscapes-dataset.com/

Cityscape (~1e5 pixels)



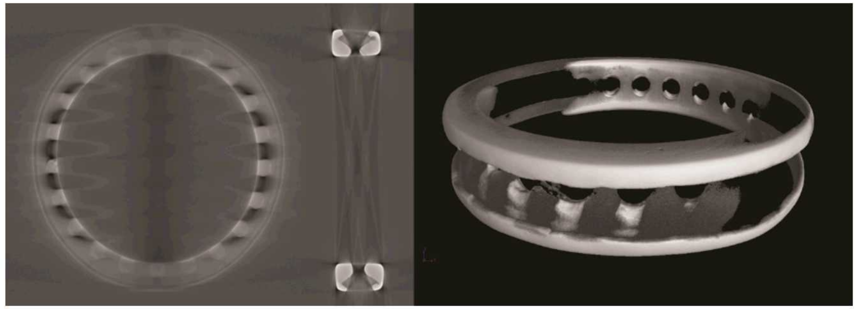
Rattlesnake Tail (~1e9 voxels)

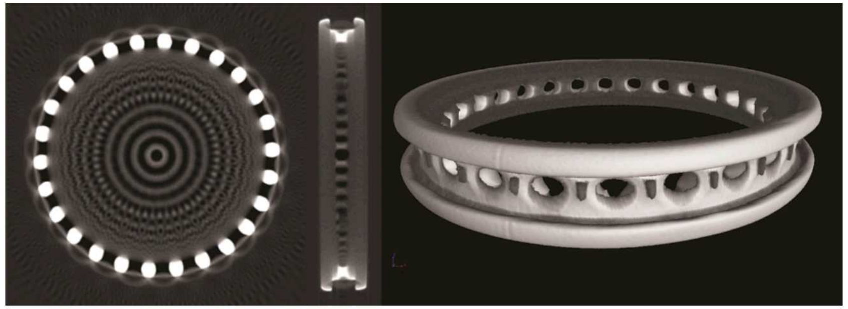
Automated Segmentation presents challenges

- CT scans are large
  - Medical literature: 128x128x64
  - Ours: ~1000x larger
  - Soon: ~10000x larger
- Class Imbalance
  - Empty space often dominates the scan
- Artifacts and noise
  - Difficult to separate materials of similar density
  - Shadow effects

Gold rings

Series of X-rays of rings

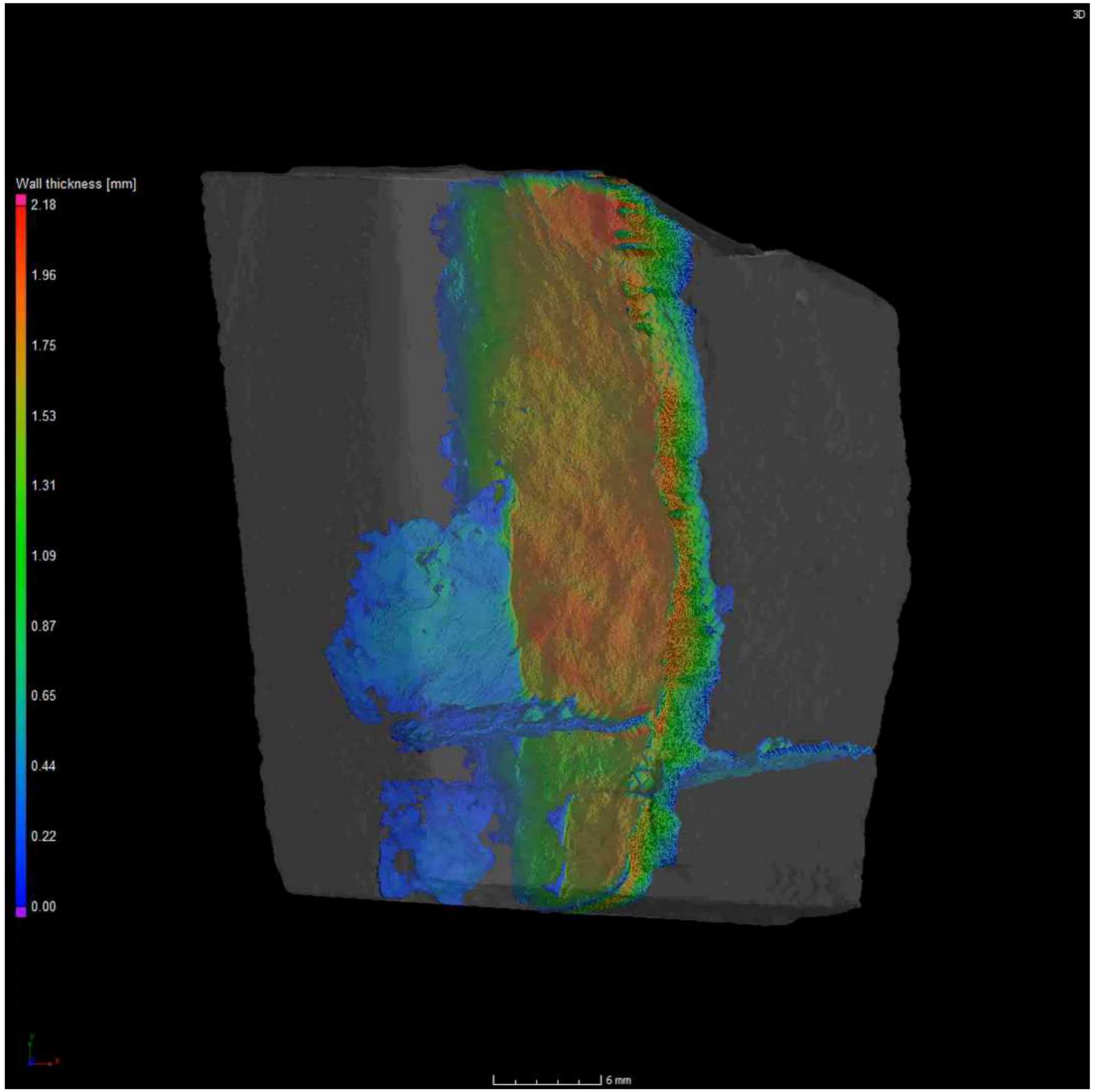




Reconstructed volumetric images of rings via computed tomography

Mitigating Challenges

- CT scans are large
  - Used 240x240x240 chunks of the volume
  - Optimized our model for GPU memory usage
- Class Imbalance
  - Adjusted loss function that guides training
- Artifacts and noise
  - Selected Convolutional Neural Network (CNN) architecture with strength in shape recognition



Rock with segmented opal veins

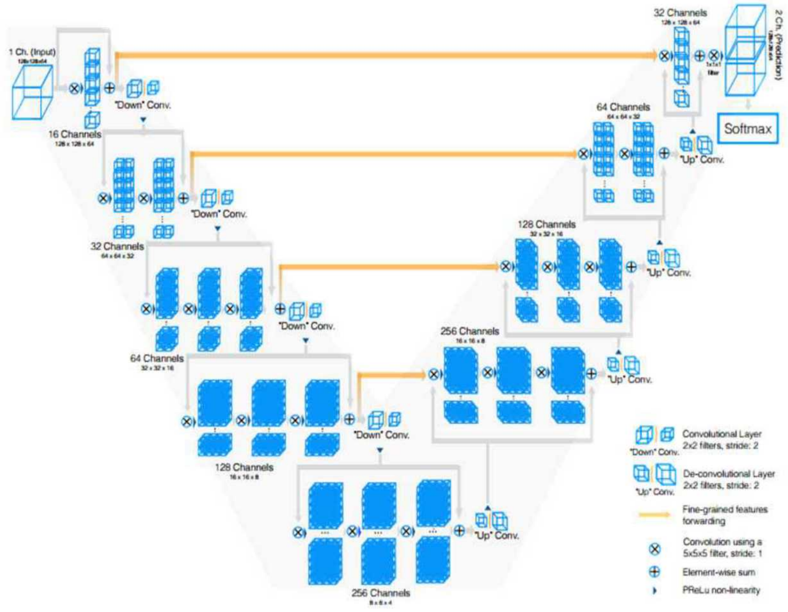
Deep learning is big data and large networks

"Deep learning is the first class of algorithms that is scalable... performance just keeps getting better as you feed them more data"

Andrew Ng, Founder of Google Brain

- U-net, a big advance in biomedical segmentation
  - Olaf Ronneberger, Philipp Fischer, Thomas Brox , "U-Net: Convolutional Networks for Biomedical Image Segmentation", in Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, LNCS, Vol.9351: 234--241, 2015

- V-net follows as a natural extension
  - F. Milletari, N. Navab, and S. A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in 2016 Fourth International Conference on 3D Vision (3DV), Oct 2016, pp.565–571

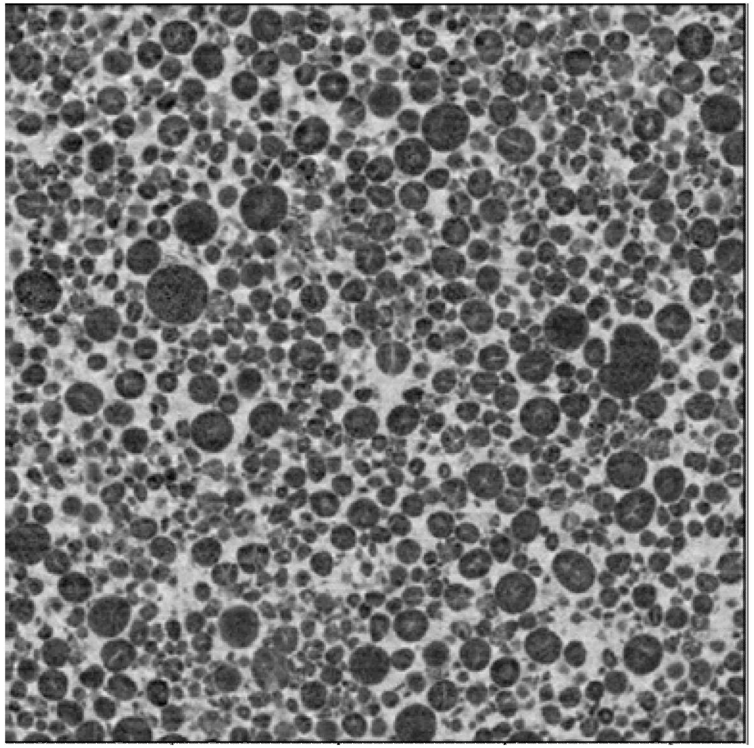


V-Net architecture for segmenting volumetric data (2016)

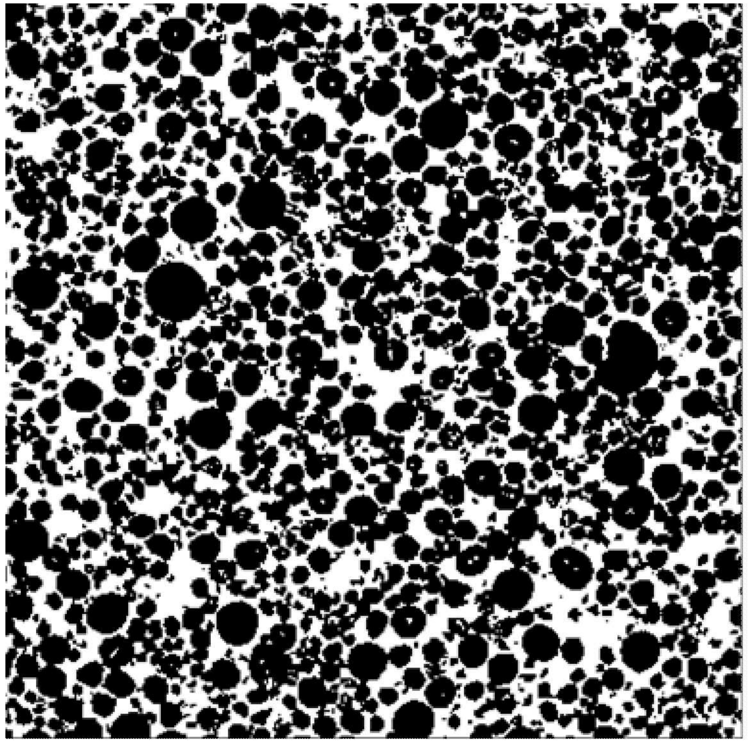
We started with a V-Net and made improvements as necessary

Deep learning model must learn to identify battery electrodes after training on human labeled examples

Slice of 3D Image

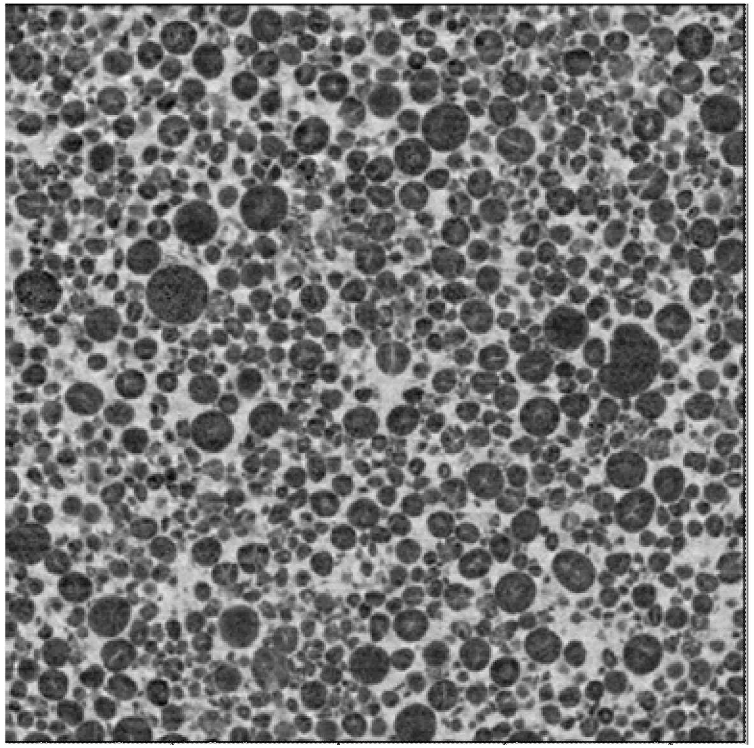


Human label

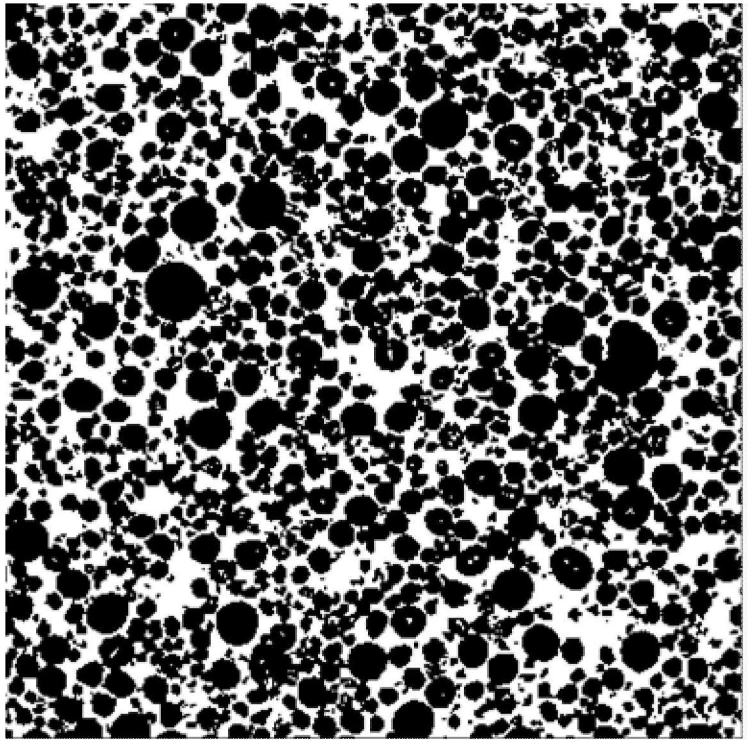


Volumetric battery segmentation achieves high accuracy compared to human labels

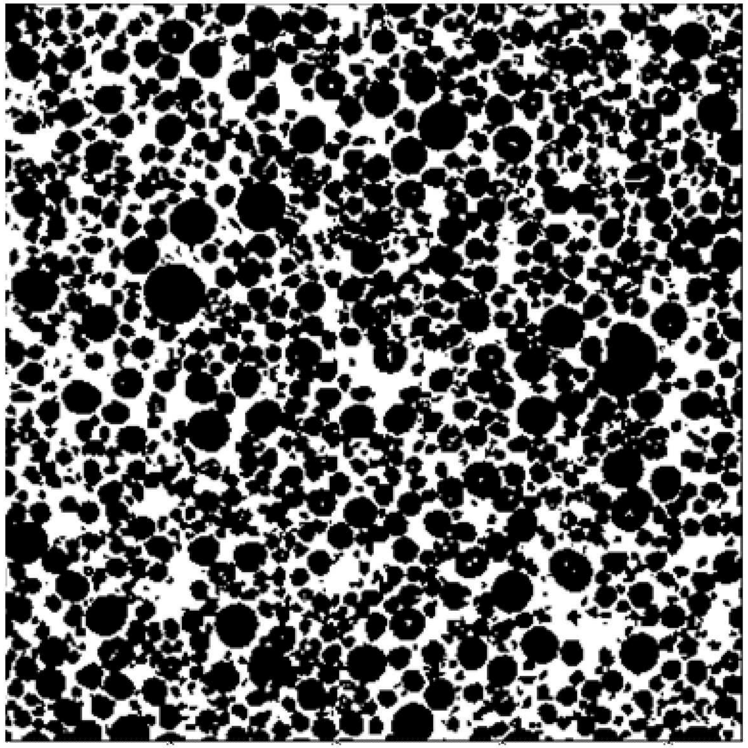
Slice of 3D Image



Human label



ML prediction

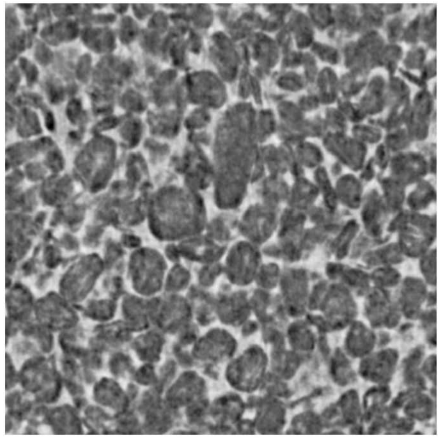


Averaged 99.7% accuracy over held out test set

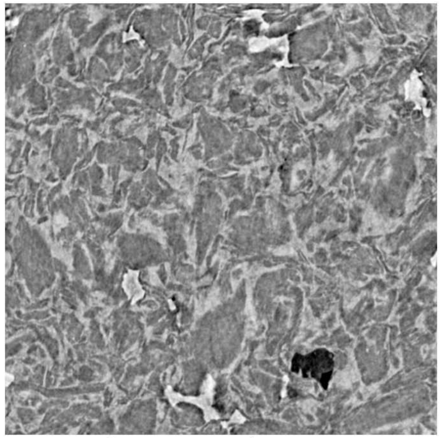
Extending segmentation model to different battery types seems challenging

DOMAIN NAME	ACCURACY
E35	0.984
Tesla	0.973
Litarion	0.966
25R6	0.955
Electrode_I_1	0.948
Electrode_III_1	0.945
GCA400	0.928
Electrode_IV_1	0.917
Electrode_II_2	0.902
GCA2000	0.900
Electrode_I_2	0.892
Electrode_III_2	0.773
Electrode_IV_3	0.748
Electrode_IV_2	0.745
Electrode_II_3	0.699
Electrode_III_3	0.668
Mean	**0.8714375**

TRAINING SET

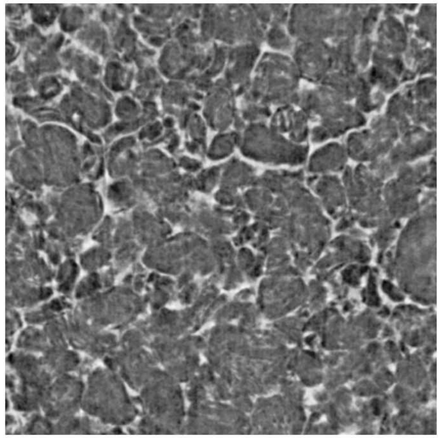


Litarion

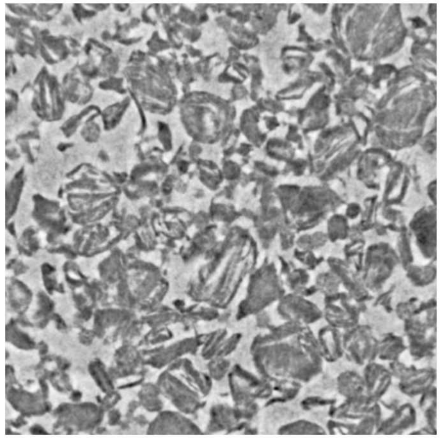


Electrode IV_1

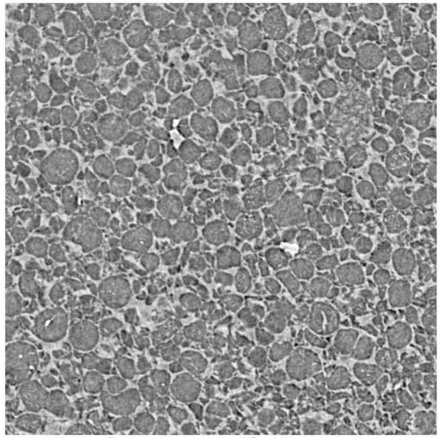
TEST SET



E35



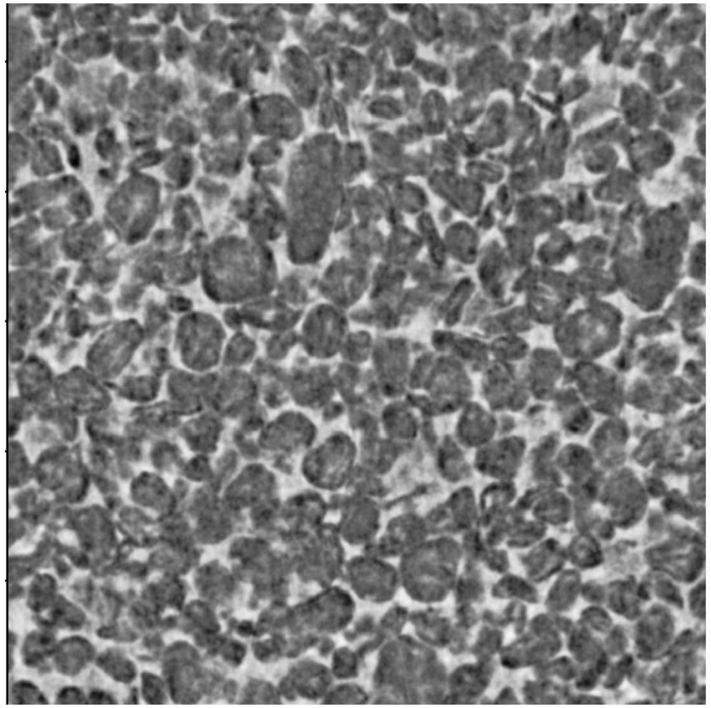
GCA400



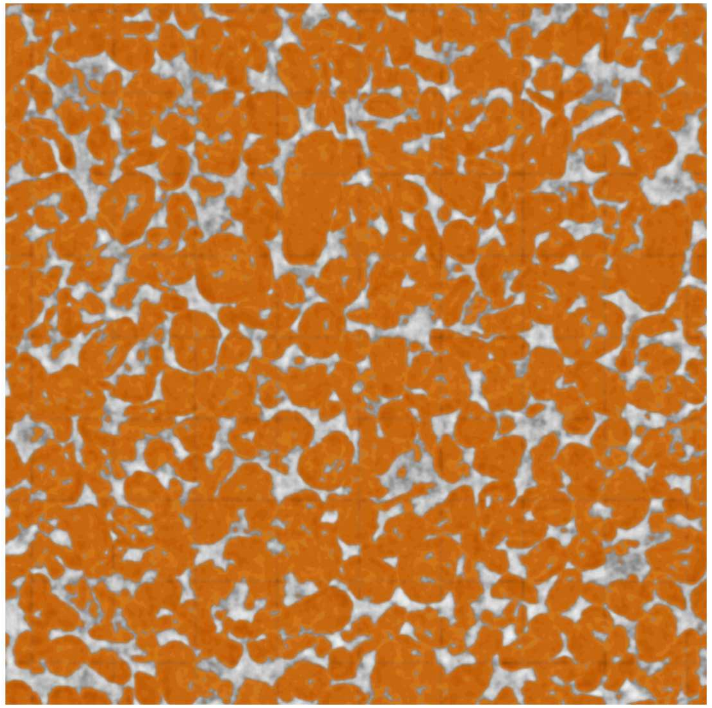
Electrode II_3

Inference results in training domain are as expected

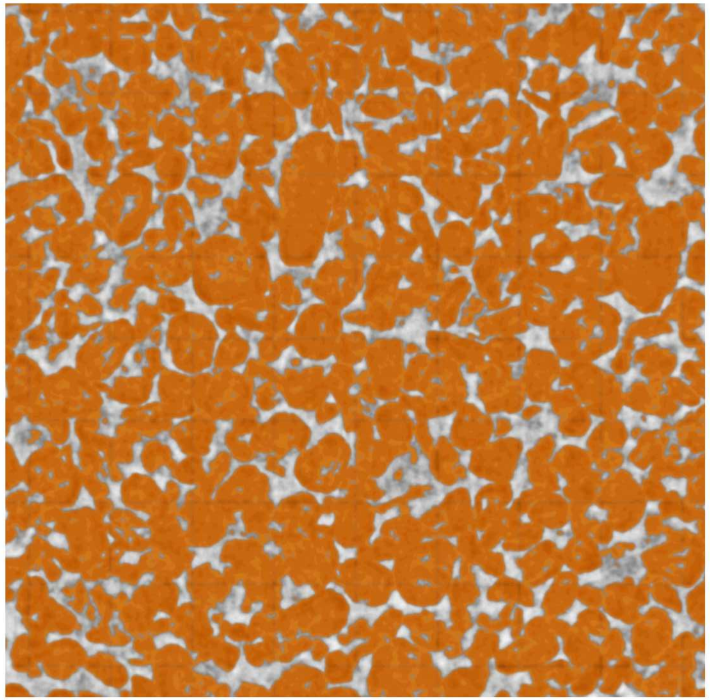
Litarion CT scan slice



Human label



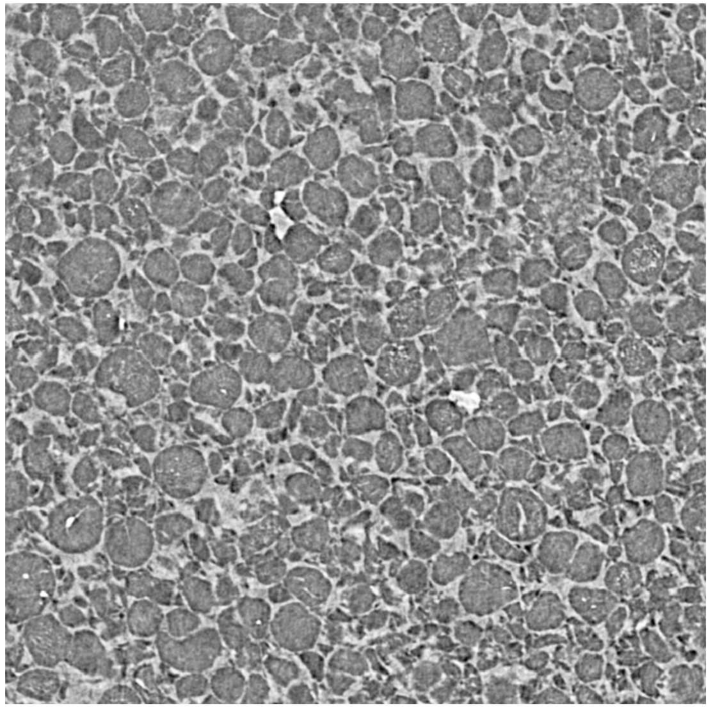
ML prediction



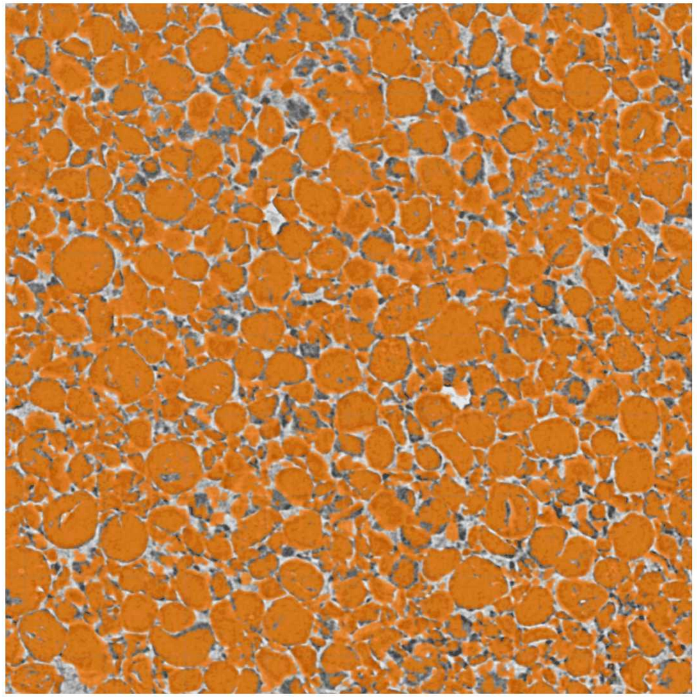
ML segmentation is 96.6% accurate to the human label

Inference results outside the training domain are qualitatively better than accuracy measurements indicate

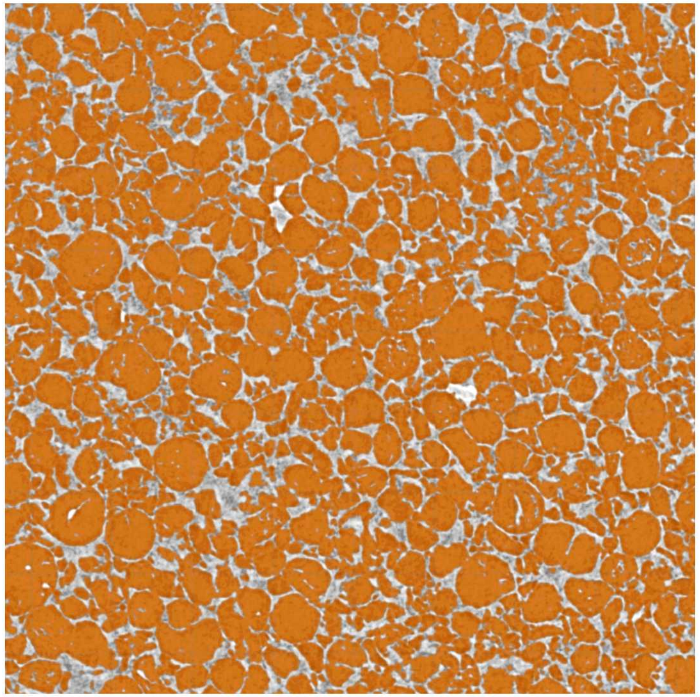
Electrode II_3 CT scan slice



Human label



ML prediction

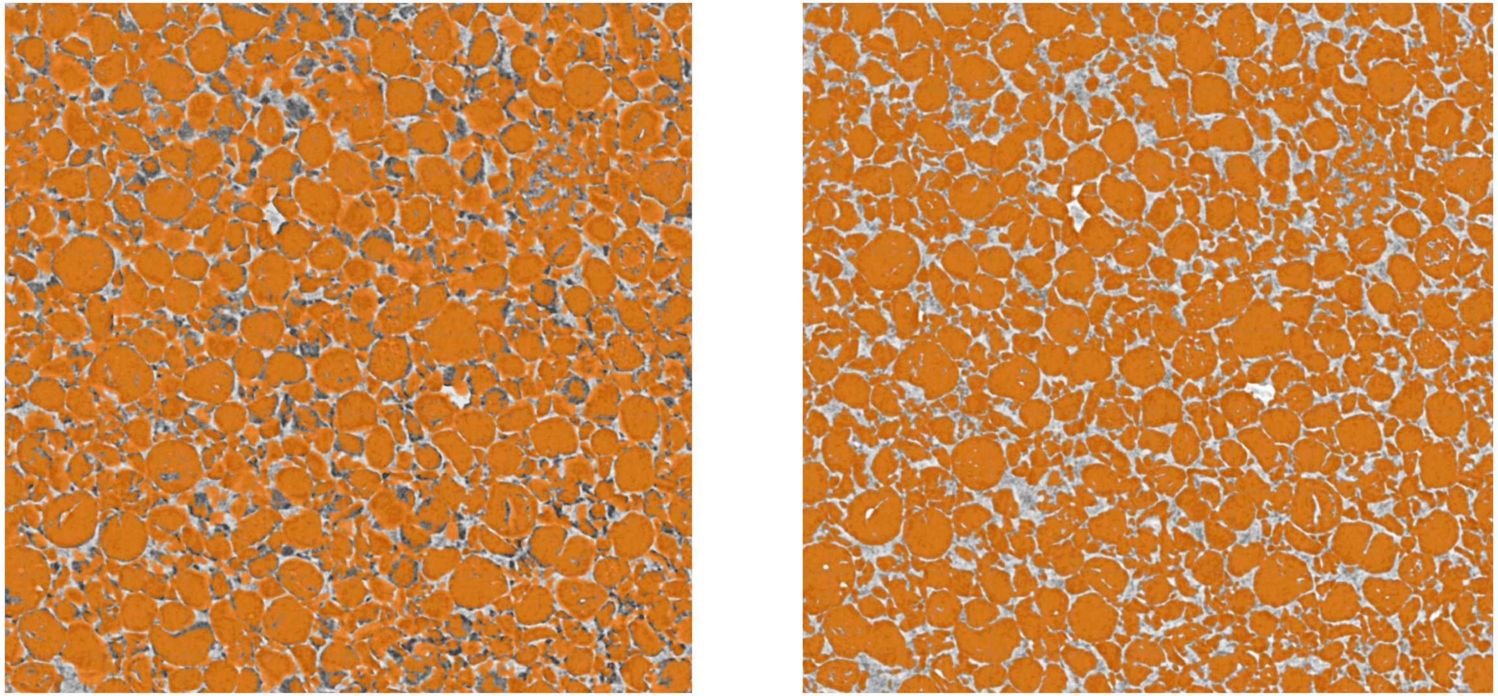


ML segmentation is 69.9% accurate to the human label…but looks qualitatively better

Measuring binary segmentation quality in the absence of precise labels

$$\text{VIABILITY SCORE} \stackrel{\text{def}}{=} mean(V_1) - mean(V_0)$$

where $V_1 = \{v_k | l_k = 1\}$, $V_0 = \{v_k | l_k = 0\}$,
v_k is the intensity of voxel k with label l_k



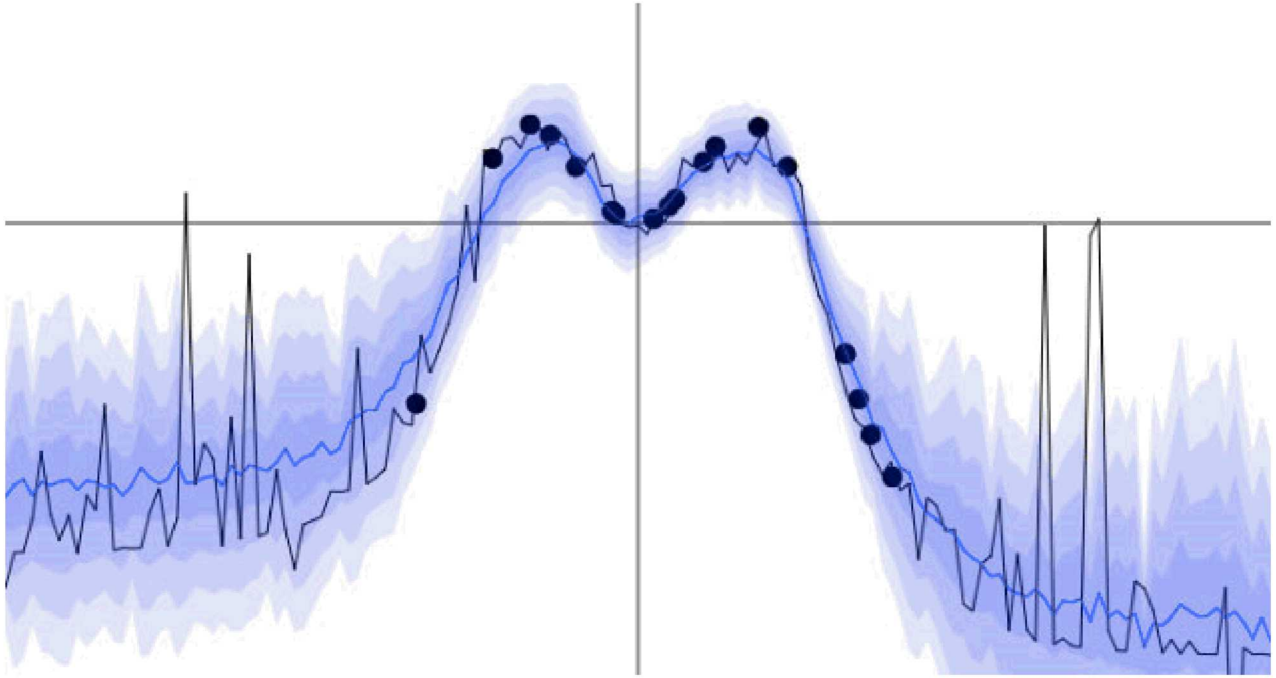
Human label

ML prediction

Robust ML model is capable of producing segmentations that are qualitatively better than human labels, even outside of the training domain

DOMAIN NAME	ACCURACY	HUMAN LABEL VIABILITY SCORE	ML LABEL VIABILITY SCORE
E35	0.984	**1.966**	1.945
Tesla	0.973	**2.383**	2.381
Litarion	0.966	2.105	**2.169**
25R6	0.955	2.363	**2.388**
Electrode_I_1	0.948	1.612	**1.613**
Electrode_III_1	0.945	1.676	**1.689**
GCA400	0.928	1.852	**1.876**
Electrode_IV_1	0.917	1.473	**1.554**
Electrode_II_2	0.902	1.508	**1.620**
GCA2000	0.900	1.894	**1.933**
Electrode_I_2	0.892	1.490	**1.573**
Electrode_III_2	0.773	1.001	**1.639**
Electrode_IV_3	0.748	0.868	**1.536**
Electrode_IV_2	0.745	0.782	**1.546**
Electrode_II_3	0.699	0.500	**1.685**
Electrode_III_3	0.668	0.475	**1.635**
Mean	0.8714375	1.497	**1.799**

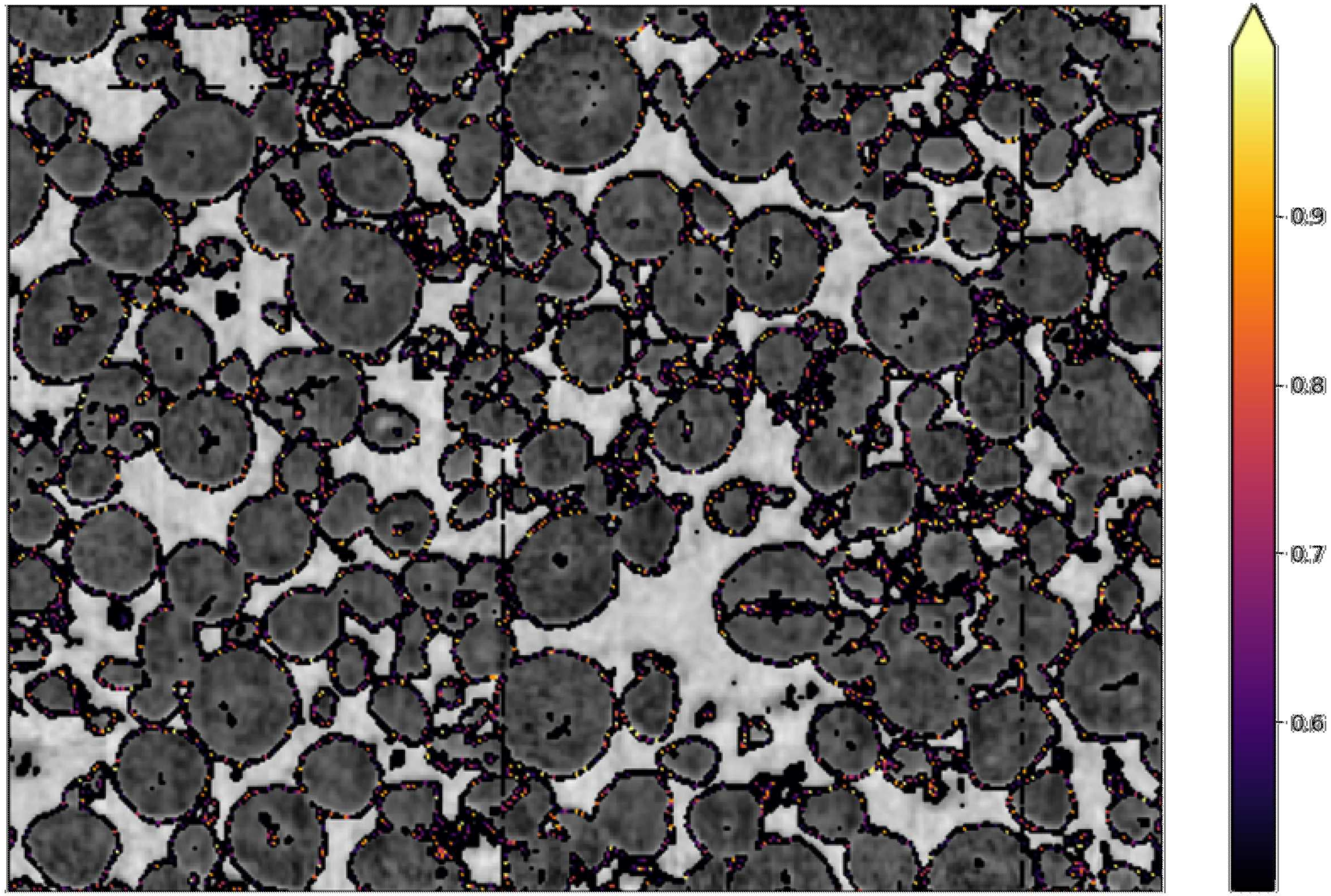
http://www.cs.ox.ac.uk/people/yarin.gal/website/blog_3d801aa532c1ce.html



Using dropout to estimate segmentation confidence

UNCERTAINTY QUANTIFICATION

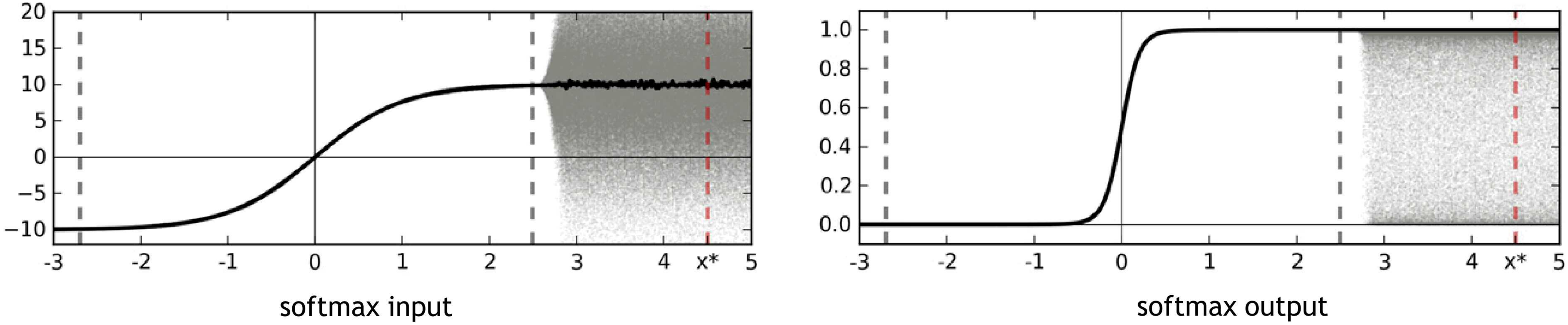
How can we understand geometric uncertainties in deep learning segmentations?



CT scan of battery with output of softmax layer overlaid

Neural network softmax layers are insufficient to characterize uncertainty outside of the training domain

- Softmax output is not always enough to determine model uncertainty



https://www.cs.ox.ac.uk/people/yarin.gal/website/blog_3d801aa532c1ce.html

- We can use dropout at inference time to approximate uncertainty

  Yarin Gal and Zoubin Ghahramani. "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning." International Conference on Machine Learning. 2016.

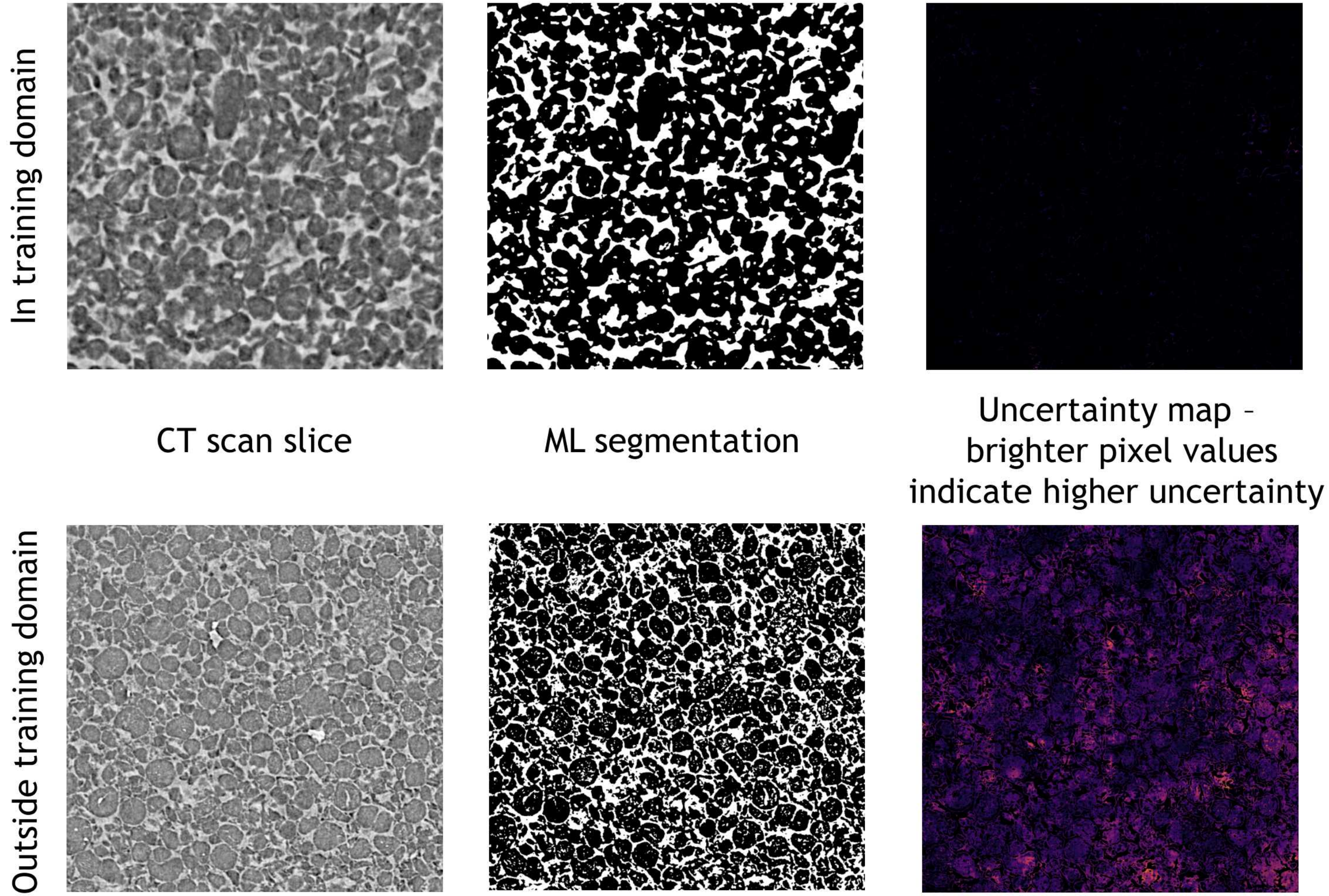
Uncertainty quantification allows us to add error bars to our deep learning models

Using a deep learning technique called dropout, we sample segmentation results from the trained model.

For each pixel, we quantify the level of uncertainty in our model, and we can make decisions about the model's credibility on a particular task.

The trained model has less confidence in segmentations of inputs that fall outside of the training distribution.

Work in progress: Use uncertainty maps to bound variance in geometries of as-built parts for use in simulations



QUESTIONS?

Sandia Interdisciplinary Machine Learning Research Team

simlr@sandia.gov