A Deep Learning Framework for Spatiotemporal Feature Extraction and Characterization of Synchrotron X-Ray Computed Tomography

Thomas Ciardi
Department of Computer and Data Sciences

Committee Members

Prof. Roger H. French	(Chair): EMSE, CWRU
Prof. Yinghui Wu	(Chair): CSDS, CWRU
Prof. Jing Ma	CSDS, CWRU

CENTER OF EXCELLENCE

DE-NA0004104
MDS^3, SDLE Research Center, Roger H. French © 2023 https://mds3-coe.com http://sdle.case.edu

Stress Corrosion Cracking and Synchrotron Science

HOME LATEST NEWS NEWSLETTERS TOPICS DOCUMENTS ABOUT

INSIDE DEFENSE

Thursday, January 9, 2025

Key Issues SCO Investment strategy DOD's two-front conflict plan Chinese military companies list

Frigates, LCS problems addressed

Navy Report: Cruiser Superstructure Cracking To Cost $270M To Repair

/ April 13, 2012 at 8:54 PM

Post Share

Naval Sea Systems Command is telling Congress that superstructure cracking in several classes of surface combatants is being addressed, but is in some cases proving costly. Cracking problems on the CG-47 Ticonderoga-class cruisers "appears to be the most pervasive as it extends to all ships of the class," according to the March 5 document, "Report to Congress: Surface Combatant Topside Superstructure Cracking," which was recently reviewed by *Inside the Navy*. In addition to facing fatigue cracks, "stress corrosion cracking...

Stress corrosion cracking (SCC):

- Failure mechanism in marine AlMg
- Tensile stress, material, corrosive environment
- Slow cracking -> catastrophic failure

Synchrotron X-ray Tomography:

- High-resolution spatiotemporal resolution
- Enable materials degradation studies
- Huge throughput: 30GB/s

Experimental Background

AlMg plates from HMCS Iroquois:

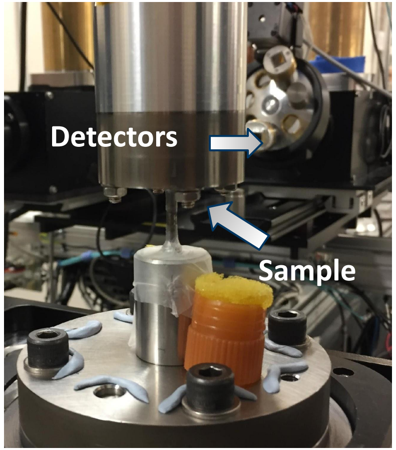
- Decommissioned Navy destroyer
- 1972 to 2014 in Gulf theatre, Domalia, and Caribbean Sea
- Aluminum: 5XXX rolled plates

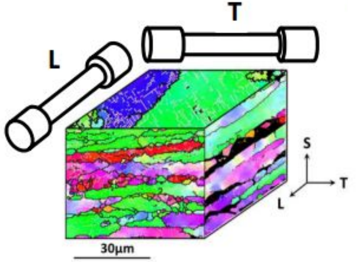
Sample Processing:

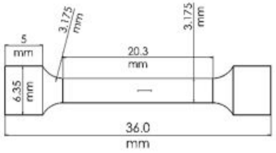
- Plate N (6 mm thick)
- High sun exposure
- T orientation

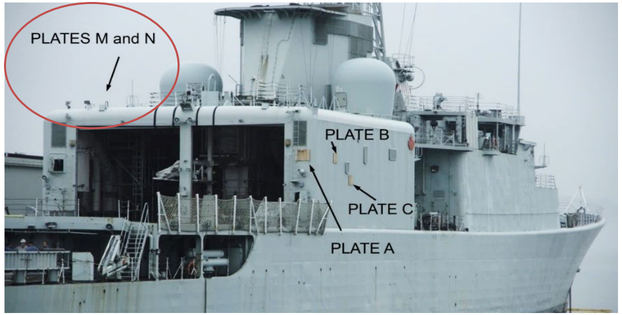
Slow strain-rate tension test:

- Synchrotron at Diamond Light Source (Didcot, UK)
- Intermittent holds on load to scan









[1] Burnett, T.L., Holroyd, N.J.H., Lewandowski, J.J., Ogurreck, M., Rau, C., Kelly, R., Pickering, E.J., Daly, M., Sherry, A.H., Pawar, S., Slater, T.J.A., and Withers, P.J. (2017). "Degradation of Metallic Materials Studied by Correlative Tomography", in 38 th Riso International Symposium on Materials Science – IOP Conf. Series: Materials Science and Engineering, 219(1), 012001.
[2] Gudla, V.C., Garner, A., Storm, M., Gajjar, P., Carr, J., Palmer, B.C., Lewandowski, P.J., Holroyd, N.J.H., and Burnett, T.L. (2019). "Initiation and Short Crack Growth Behavior of Environmentally Induced Cracks in AA5083-H131 Investigated Across Time and Length Scales" Corrosion Reviews, 37(5), pp. 469-481.

Research Challenge: Analyzing Massive 4D Synchrotron Datasets

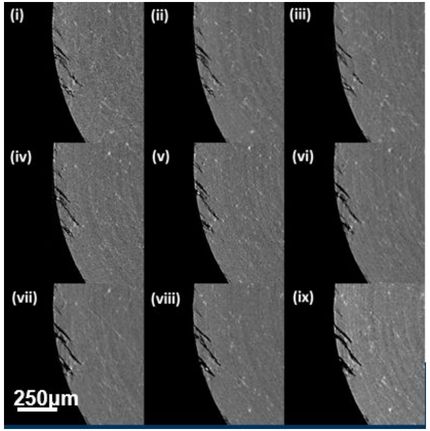
Scale of Dataset:

- 3TB imaging dataset
- Total: 342,000 high-resolution images

Goal: Detect all microstructural features in the 4D XCT dataset and quantify their properties to enable stress corrosion cracking studies

Framework Contributions:

1. **Domain-informed diversity sampling** strategy to select which images from the dataset are **most informative** for training
2. **Scalable** machine learning **spatiotemporal feature extraction** and characterization framework
3. Application to AlMg dataset to detect and quantify over **5 million microstructural defects**



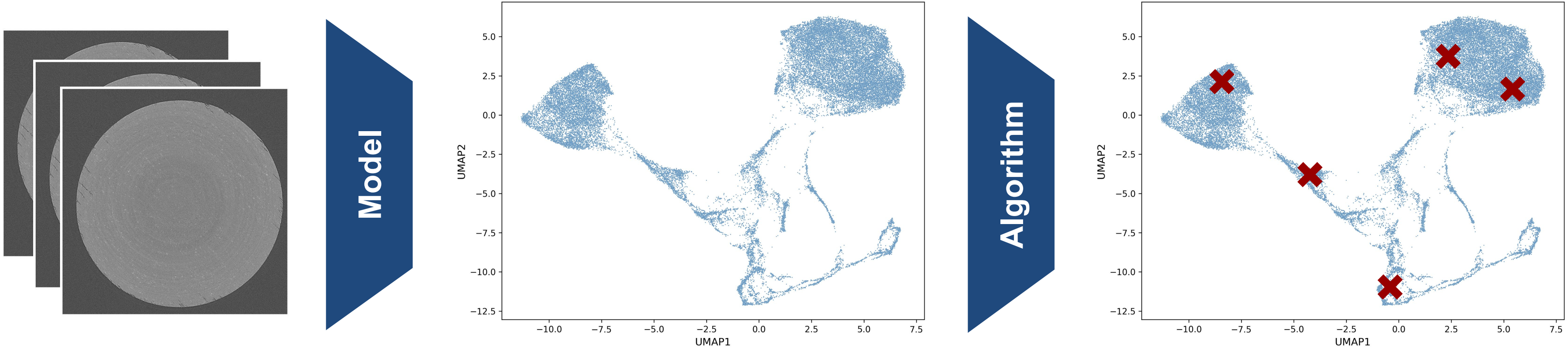
Part I: Domain-Informed Diversity Sampling

Crash Course: Diversity Sampling

General Form: Given a set of unlabeled items U and a budget B, select a subset S to maximize M

Our Case: Given 342,000 images, select the best 95 images to label for segmentation training

General Workflow:



Cold start problem: classical challenge in machine learning where a system struggles to make accurate predictions or recommendations due to a lack of initial data

Domain-Informed Diversity Sampling

Problems: Extending to scientific imaging

- Pretrained encoders typically trained on out of distribution (OOD) data
- Diversity sampling is designed for an iterative active learning loop: “typically diverse”

Solution: Domain-informed diversity sampling (DIDS)

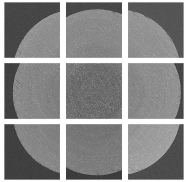
1. Incorporates domain information for diversity to close the OOD gap
2. Designed for “one-shot” setting by sampling both “typical” and “atypical” samples

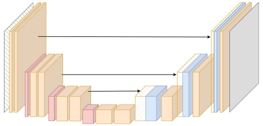
Algorithm 1 Domain-Informed Diversity Sampling (DIDS)

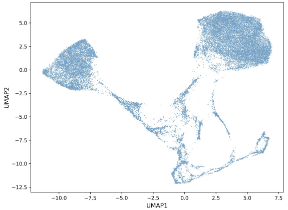
Require: Embeddings $X = \{x_1, ..., x_n\}$, displacements $D = \{d_1, ..., d_n\}$, clusters $\{C_1, ..., C_K\}$, target samples N, metric type m, α
Ensure: Selected sample indices S

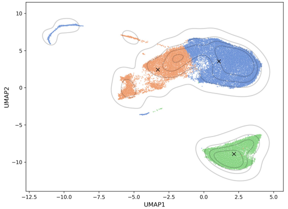
```
Normalize displacements:
$\hat{d}_i \leftarrow \frac{d_i - \min(D)}{\max(D) - \min(D)}$
Compute adaptive cluster allocation $\{n_1, ..., n_K\}$ ensuring:
$\sum_k n_k = N$ and $n_{\min} \leq n_k \leq n_{\max}$
for all clusters $C_k$ do
    Compute pairwise distances $D_{\text{emb}}$ based on metric $m$
    if $m$ = cosine then
        $D_{\text{emb}}(i,j) \leftarrow 1 - \frac{x_i \cdot x_j}{||x_i||||x_j||}$
    else if $m$ = euclidean then
        $D_{\text{emb}}(i,j) \leftarrow \frac{||x_i - x_j||}{\max_{p,q} ||x_p - x_q||}$
    end if
    Compute local densities:
    $\rho_i \leftarrow 1 - \frac{1}{|C_k|} \sum_{j \in C_k} D_{\text{emb}}(i,j)$
    Initialize $S_k$ with highest density point
    while $|S_k| < n_k$ do
        for all candidates $i \in C_k \setminus S_k$ do
            Compute displacement differences:
            $D_{\text{disp}}(i,j) \leftarrow |\hat{d}_i - \hat{d}_j|$
            Combine distances:
            $D(i,j) \leftarrow \alpha \cdot D_{\text{emb}}(i,j) + (1-\alpha) \cdot D_{\text{disp}}(i,j)$
            $D_{\min}(i) \leftarrow \min_{j \in S_k} D(i,j)$
        end for
        $i^* \leftarrow \arg\max_{i \in C_k \setminus S_k} D_{\min}(i)$
        Add $i^*$ to $S_k$
    end while
end for
return $S = \bigcup_k S_k$
```

DIDS Sampling Workflow: Displacement Informed









Image

262,144

Encoder

500-5,000

Reduce

50-100

Cluster

DIDS Sampling Algorithm:

For a given cluster

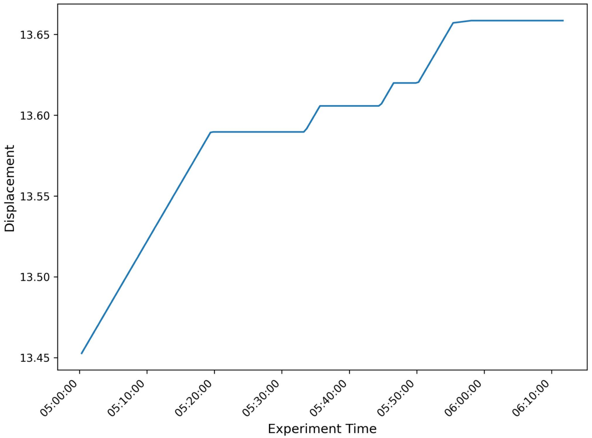
Get pairwise embedding distances of all points

Get pairwise displacement differences of all points

Get a pairwise combined distance / difference score

Select the sample with the highest density

Iteratively select points that have the largest minimum score



We consider displacement as a proxy for “material degradation”

Avoids pitfalls of simple spatiotemporal stratification

Quality Assessment: Diversity Score

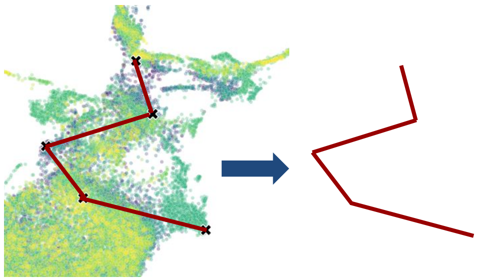
Problem: Evaluating diversity sampling requires post-hoc annotation analysis

- Sample baseline -> annotate -> train model -> measure difference

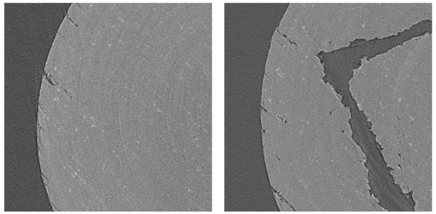
Solution: Composite diversity score that evaluates multiple diversity factors

$$D_{\text{total}} = \boxed{w_{\text{lat}}\text{LatentSpread}} + \boxed{w_{\text{vis}}\text{LPIPS}} + \boxed{w_{\text{disp}}\text{Displacement}}$$

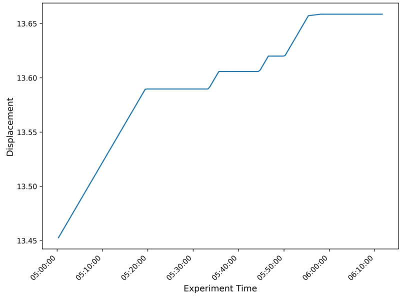
Coverage: Minimum spanning tree total and average edge distance



Visual similarity: Learned perceptual image patch similarity (LPIPS)



Domain similarity: Range of material degradation state captured



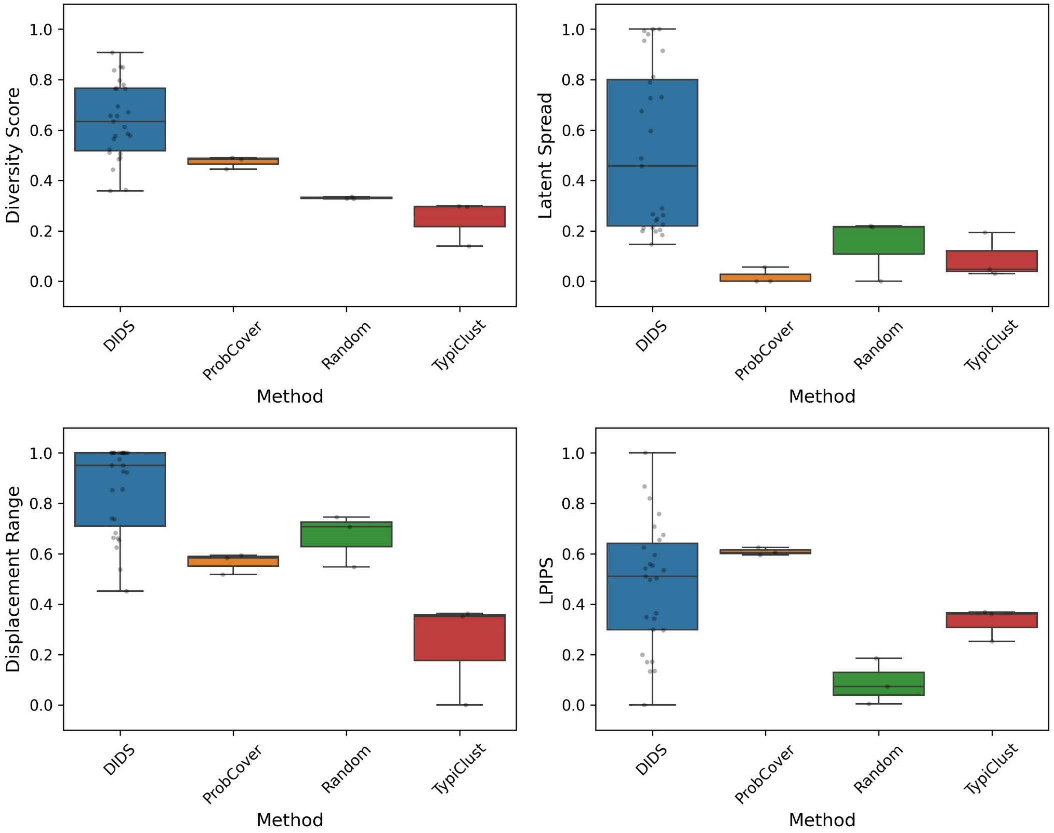
Diversity Sampling Metrics

Four Methods:

- Random sampling
- TypiClust: *typicality sampling from k-means embeddings*
- ProbCover: *probabilistic coverage maximization of embeddings*
- DIDS (Ours)

Three Encoders:

- CLIP
- ResNet50
- VGG-19

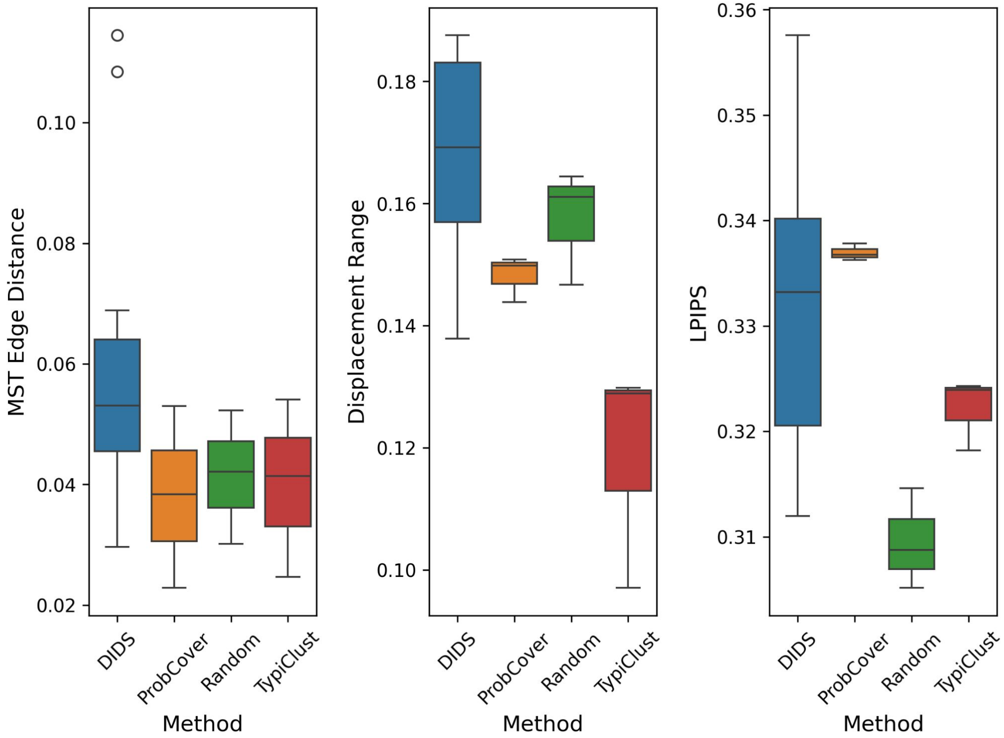


Real Value Diversity Sampling Metrics

Our diversity metric is biased towards DIDS due to including a displacement measurement

Coverage / Perceptual Metrics:

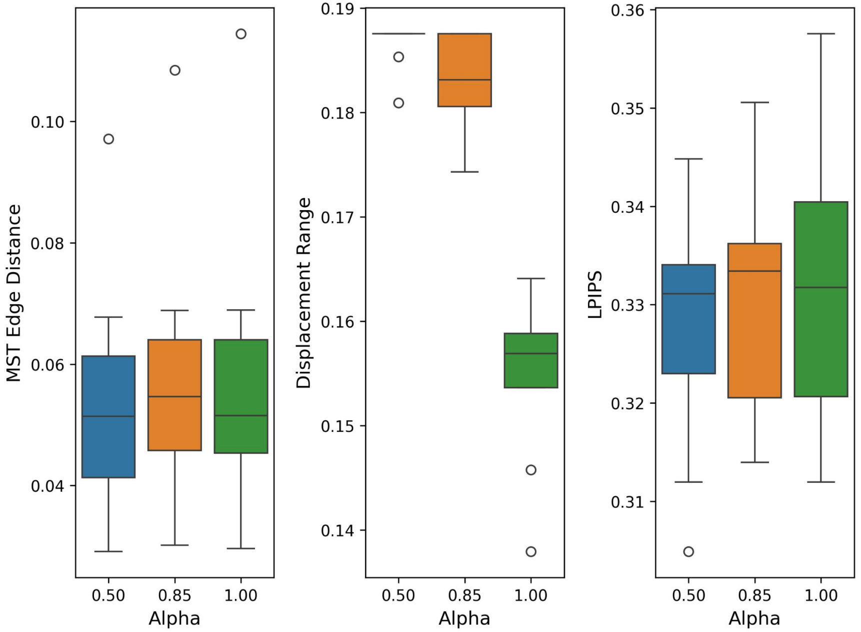
- DIDS demonstrates has larger MST coverage values: max min sampling strategy
- LPIPS differences cover a very small range: differences in tiles small compared to AlexNet's training base

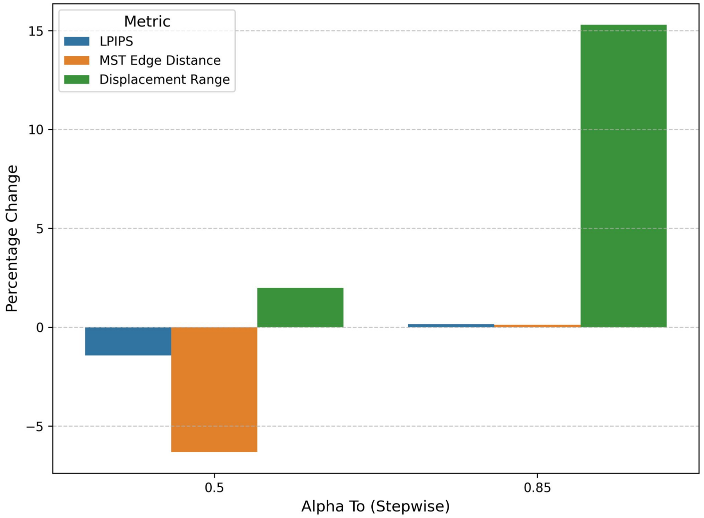


DIDS Domain Information Ablation Study

How does setting alpha = 1 (no displacement information) effect diversity in DIDS?

$$D(i,j) = \alpha \cdot d_{\text{emb}}(i,j) \cancel{+ (1-\alpha) \cdot d_{\text{disp}}(i,j)}$$



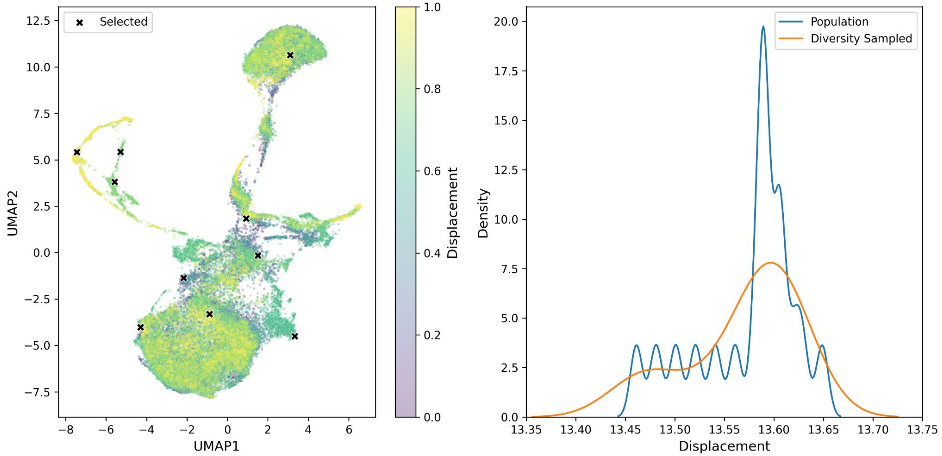


Alpha weighting can improve domain diversity without sacrificing other diversity metrics

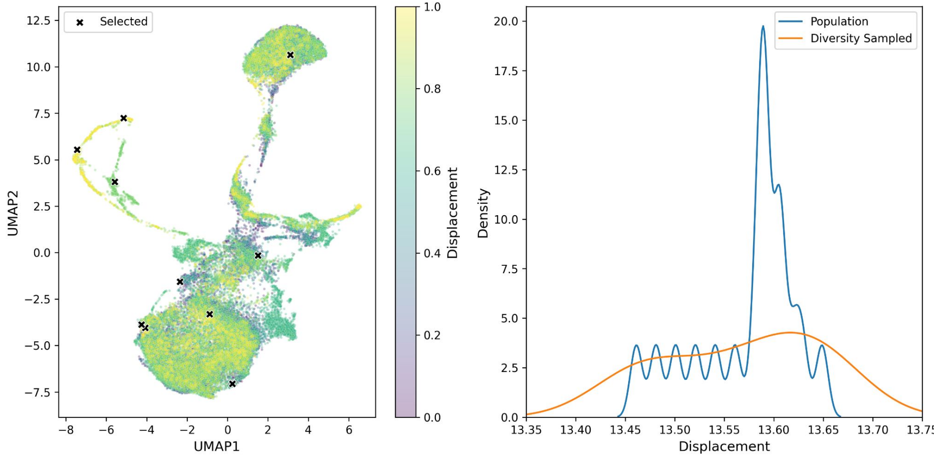
Study Case: Effect of Alpha

Effect of shifting alpha on our best performance method: DIDS Euclidean K-Means

Alpha = 1.0

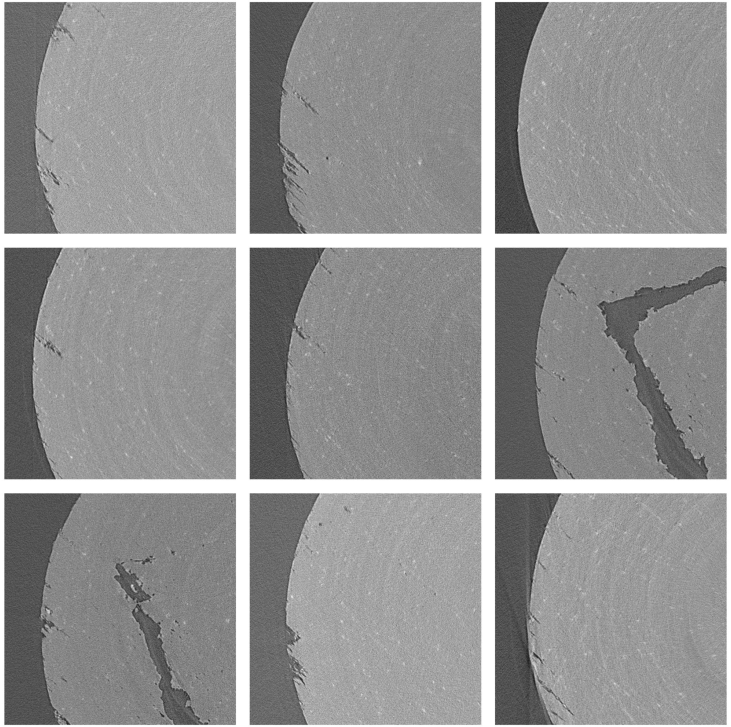


Alpha = 0.85

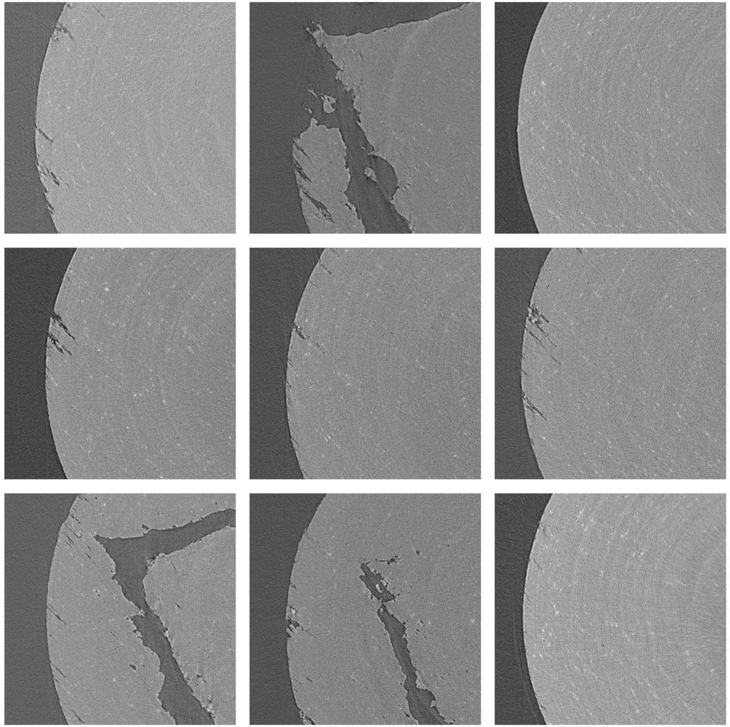


Study Case: Effect of Alpha

Alpha = 1.0



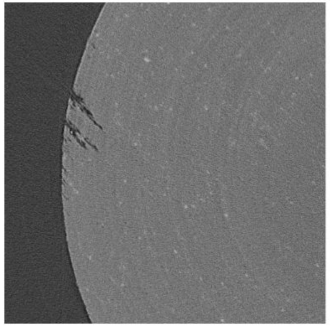
Alpha = 0.85



Part II: Scalable Spatiotemporal Feature Extraction and Characterization

Crash Course: Image Segmentation

Generate a pixel-wise mask of features of interest



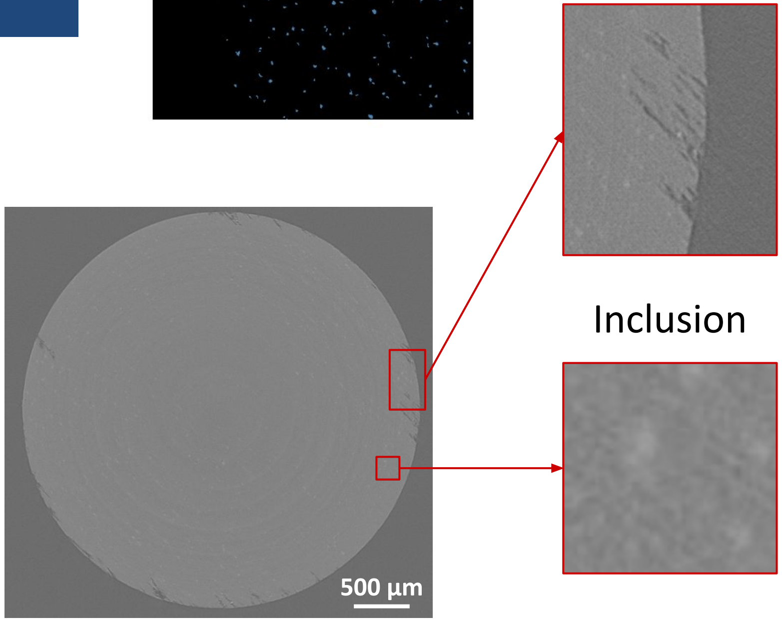
Machine Learning Model

Problems: Extending to scientific imaging

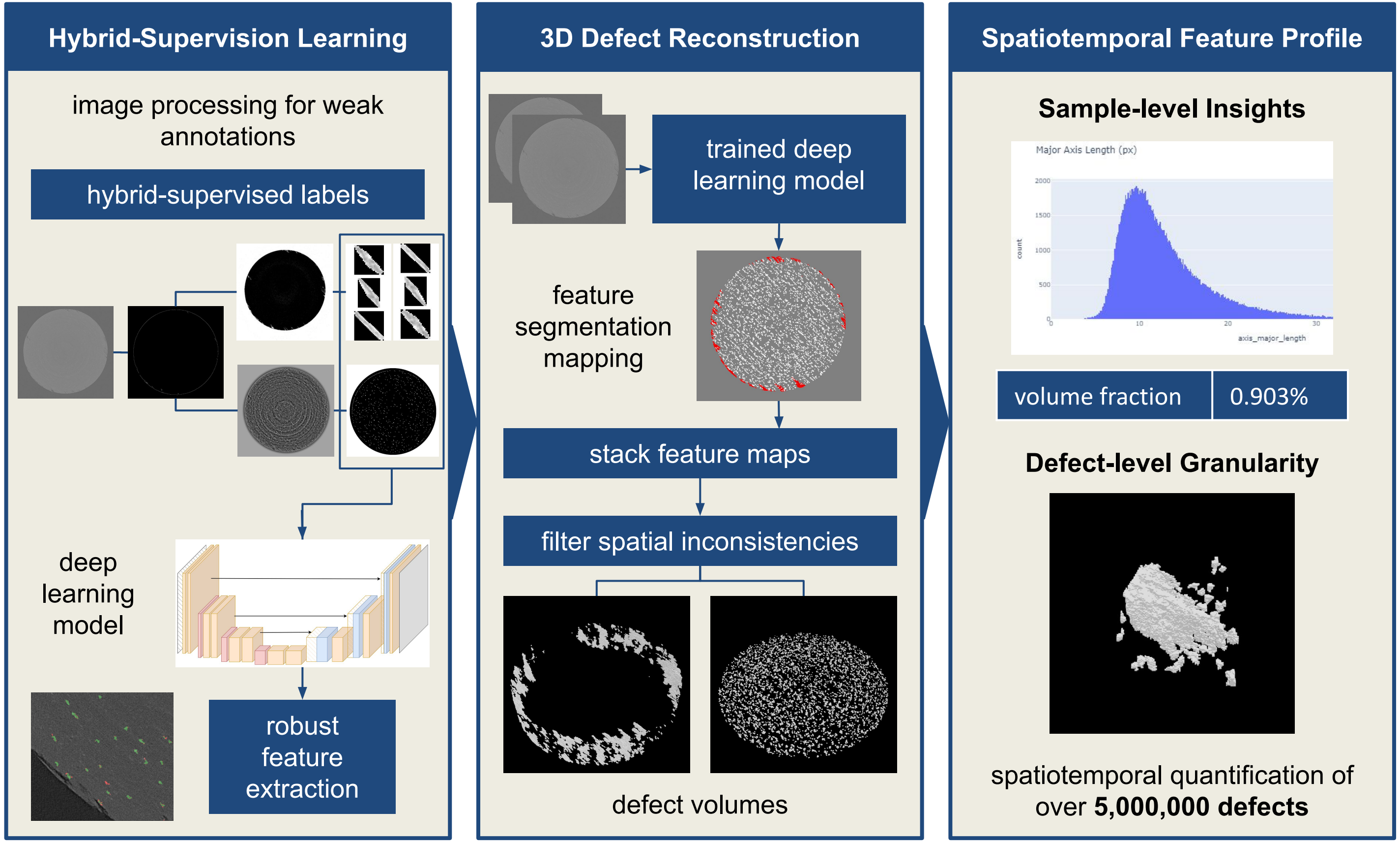
- In-situ XCT imaging generates **low resolution** detail due to strain
- Hundreds of sub-visible features per image

Solution:

- Scalable pipeline that leverages image processing for weakly supervised label generation

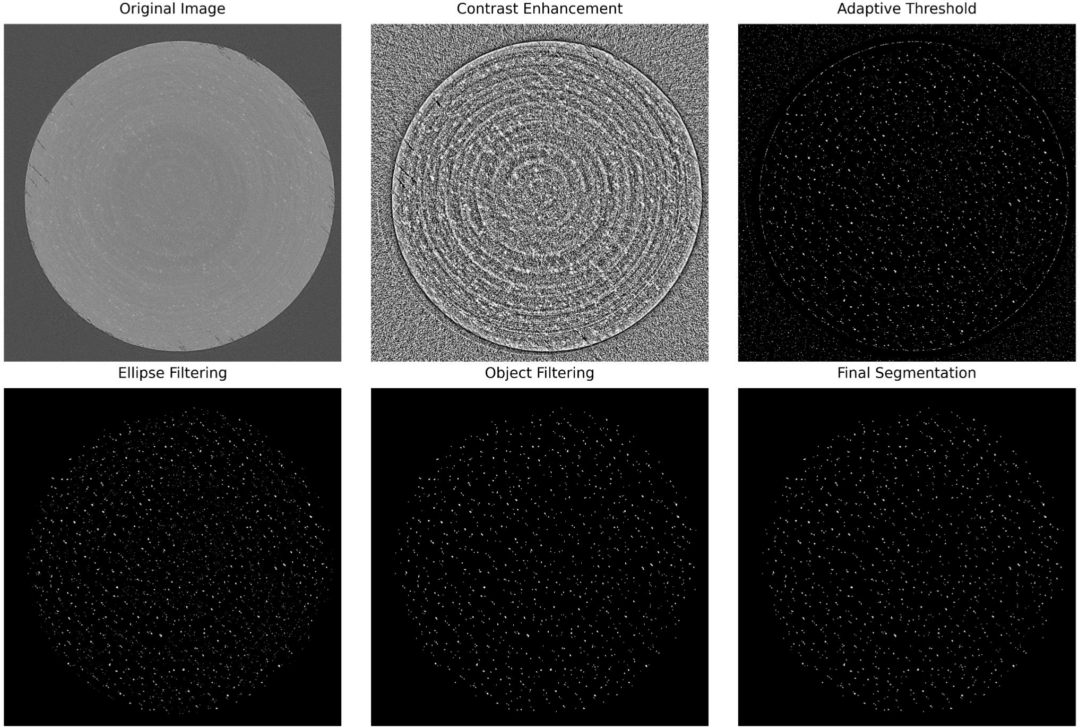


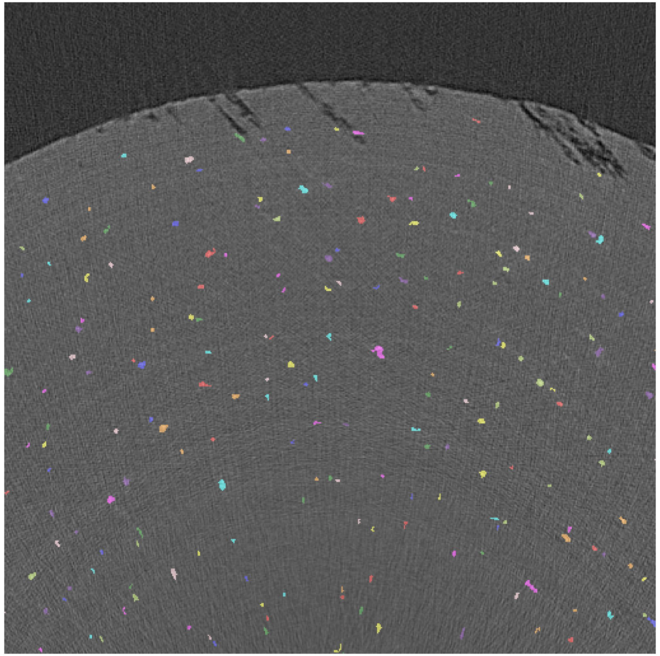
Spatiotemporal Feature Extraction Framework for Large-Scale Datasets



Weak Pseudo-label Generation for Inclusions

Classical image processing generates “rough” masks of sub-visible features **but fails to scale**





Segmentation Model Training

Model Architectures:

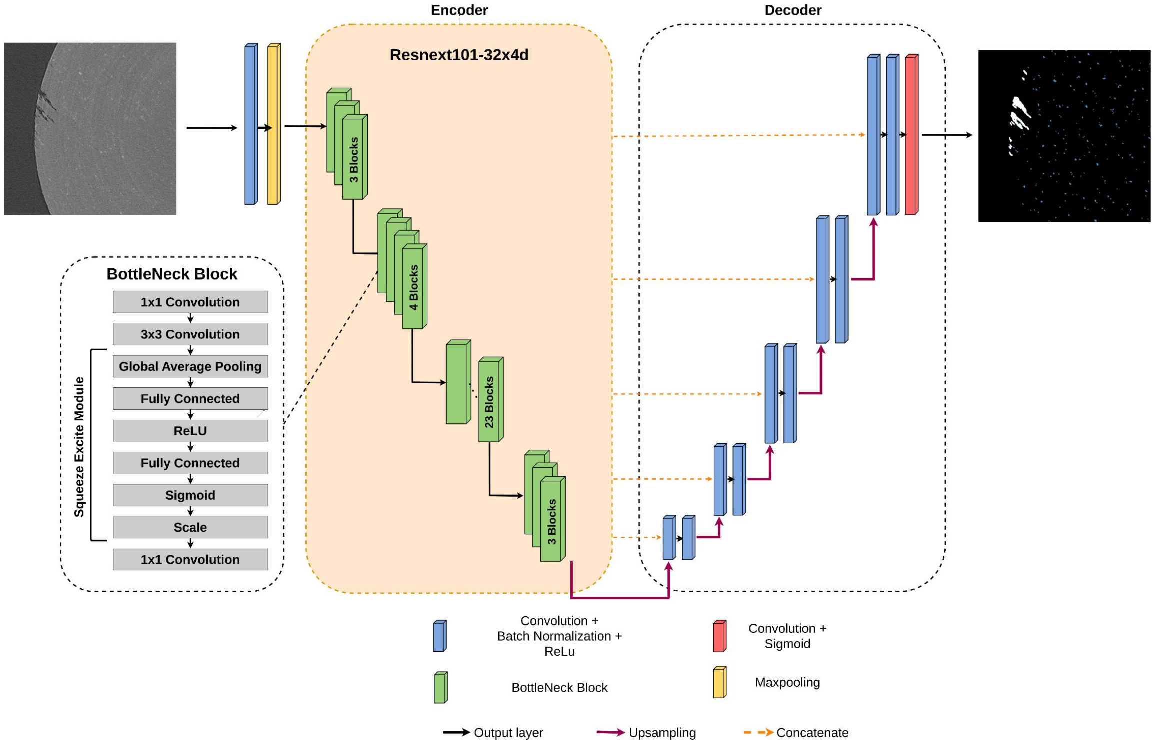
- U-Net
- UPerNet
- SegFormer

Data Parameters:

- 95 DIDS selected images
- 85/5/5 training/validation/test
- Spatial/contrast augmentations

Hyperparameters:

- ImageNet weights
- Dice loss
- 150 epochs
- Cosine LR scheduler



Segmentation Results: Quantitative

U-Net with Xception encoder has best performance on holdout test set (F1: 0.94, mIoU: 0.64)

Model	Encoder	Weights	Accuracy	Precision	Recall	F1-Score	mIoU
U-Net	None	None	0.865	1.000	0.865	0.926	0.588
U-Net	ResNet50	None	0.873	1.000	0.873	0.931	0.588
U-Net	ResNet50	ImageNet	0.881	1.000	0.881	0.935	0.612
U-Net	SE-ResNeXt101	None	0.879	1.000	0.879	0.935	0.600
U-Net	SE-ResNeXt101	ImageNet	0.896	0.999	0.896	0.944	0.627
U-Net	Xception	None	0.871	1.000	0.871	0.930	0.597
U-Net	Xception	ImageNet	0.911	0.999	0.911	0.949	0.646
UPerNet	ResNet50	None	0.755	1.000	0.755	0.849	0.432
UPerNet	ResNet50	ImageNet	0.808	0.997	0.808	0.884	0.481
UPerNet	SE-ResNeXt101	None	0.774	1.000	0.774	0.864	0.450
UPerNet	SE-ResNeXt101	ImageNet	0.802	0.997	0.802	0.880	0.479
UPerNet	Xception	None	0.772	0.999	0.772	0.860	0.430
UPerNet	Xception	ImageNet	0.923	0.997	0.923	0.928	0.480
SegFormer	MIT-B1	None	0.799	0.999	0.799	0.879	0.440
SegFormer	MIT-B1	ImageNet	0.827	0.999	0.827	0.900	0.421
SegFormer	MIT-B4	None	0.919	0.998	0.919	0.875	0.311
SegFormer	MIT-B4	ImageNet	0.798	0.998	0.798	0.884	0.481

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

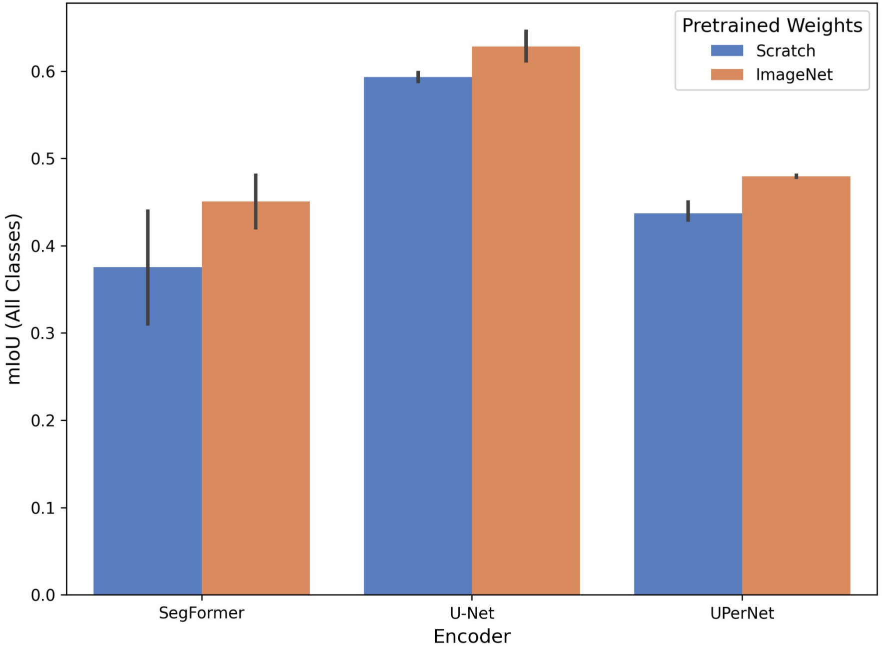
$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} = \frac{TP}{TP + FP + FN}$$

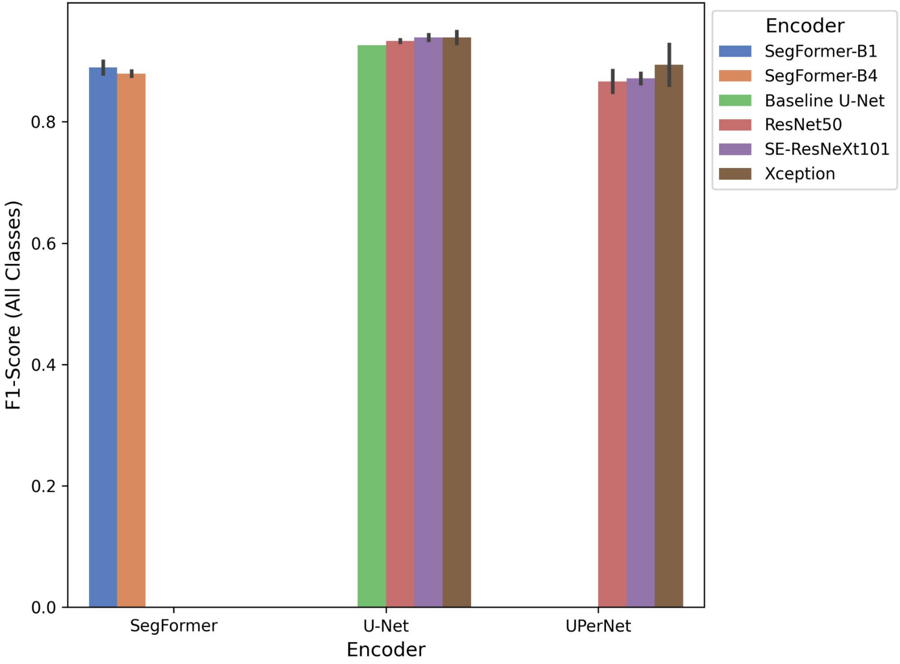
$$\text{mIoU} = \frac{1}{n}\sum_{i=1}^{n} \text{IoU}_i$$

Impact of Pretrained Weights and Encoders

Pretrained ImageNet weights consistently improve model performance

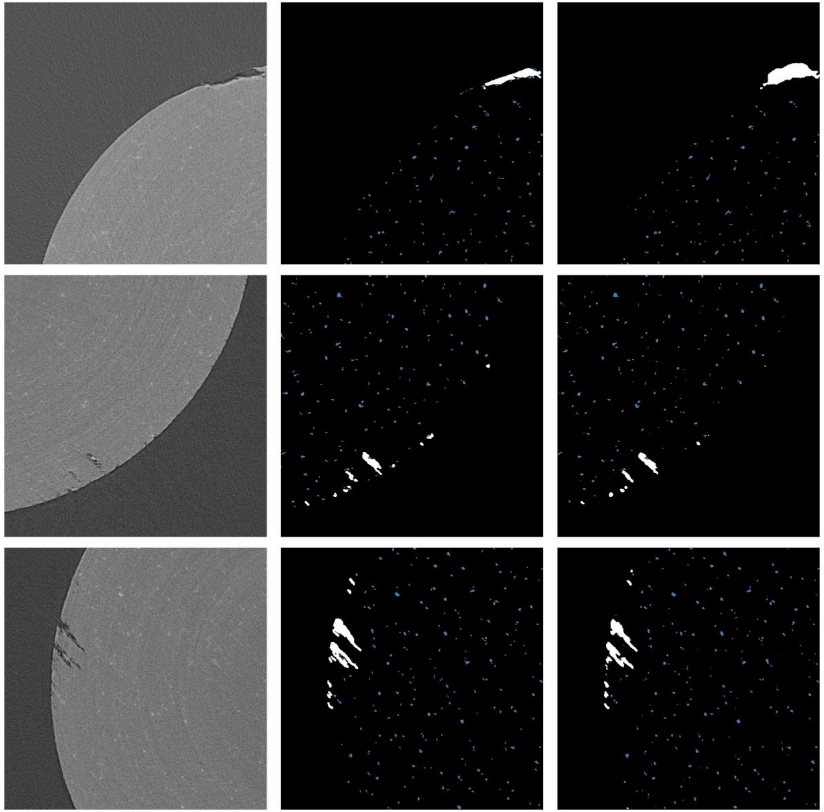


More consistent F1-scores suggest other architectures struggle more with boundaries

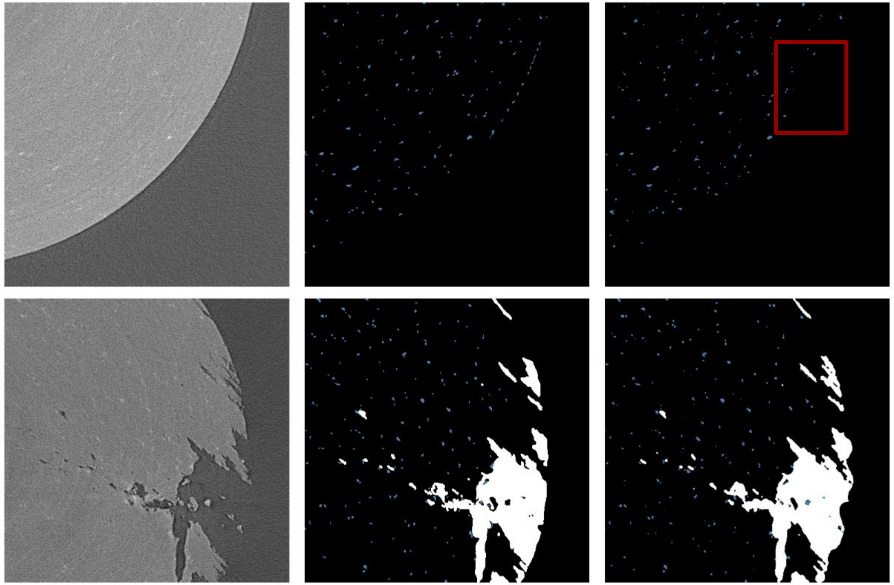


Predictions from UNet with Xception Encoder

Image Ground Truth Prediction

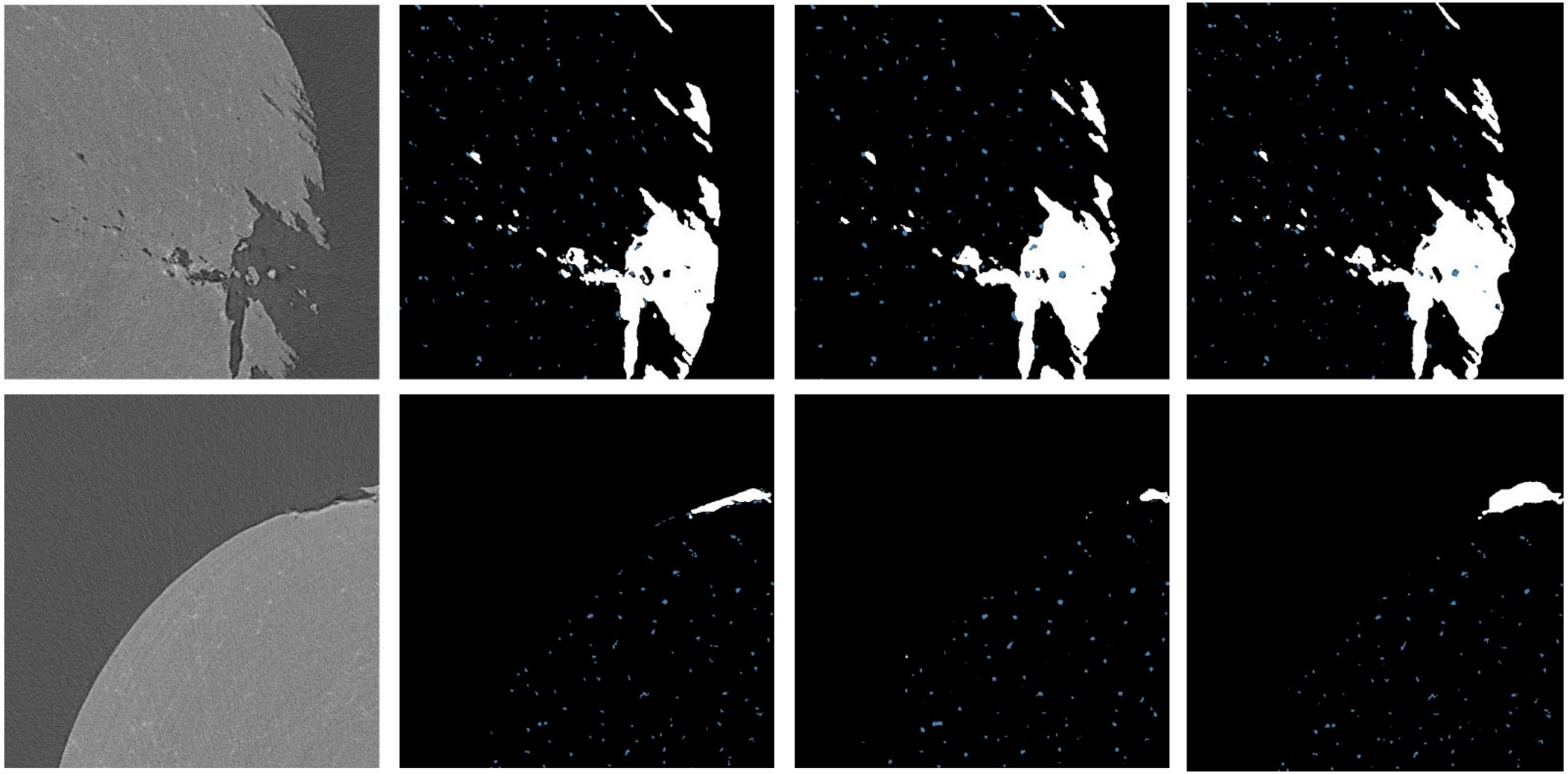


"Trickier" Examples



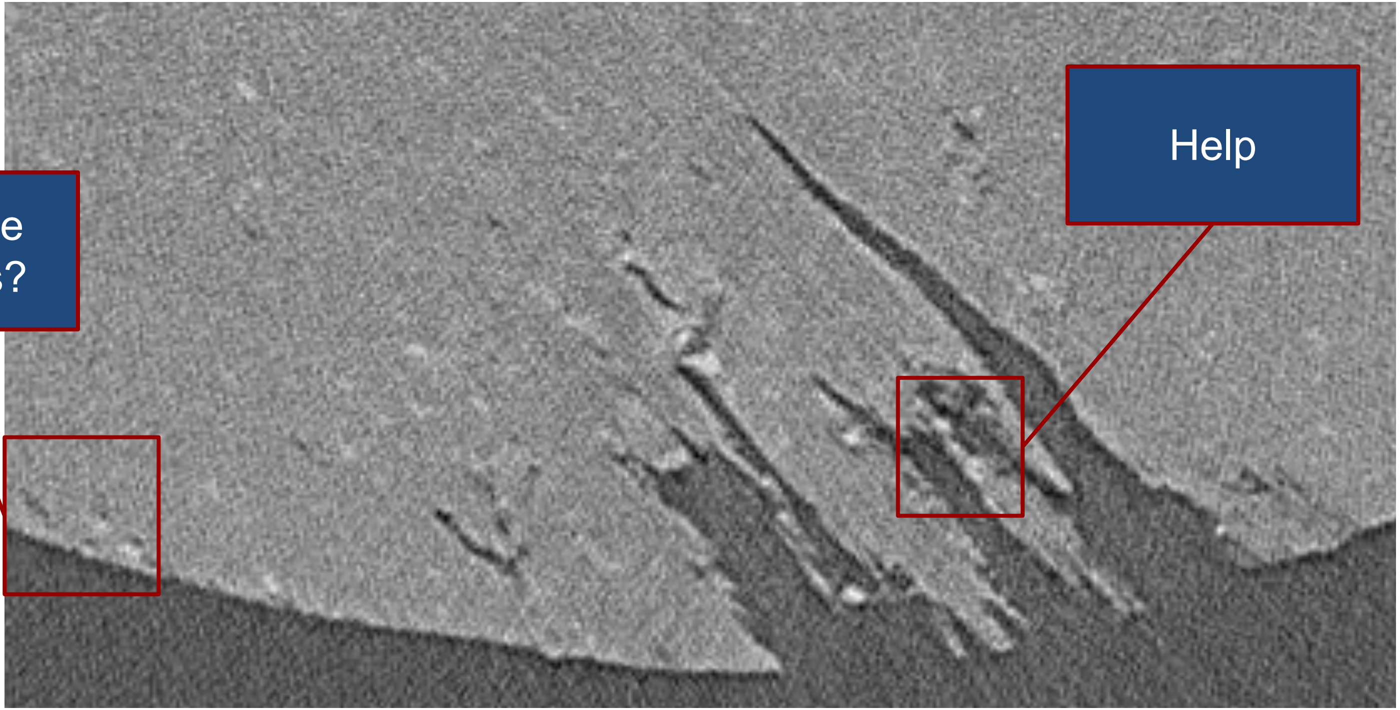
Are Transformers Really Underperforming?

Image	Ground Truth	SegFormer-B4	UNet Xception



Justification for Poor mIoU Scores

Inconsistent annotation may lead to poor quantitative metrics despite strong qualitative performance

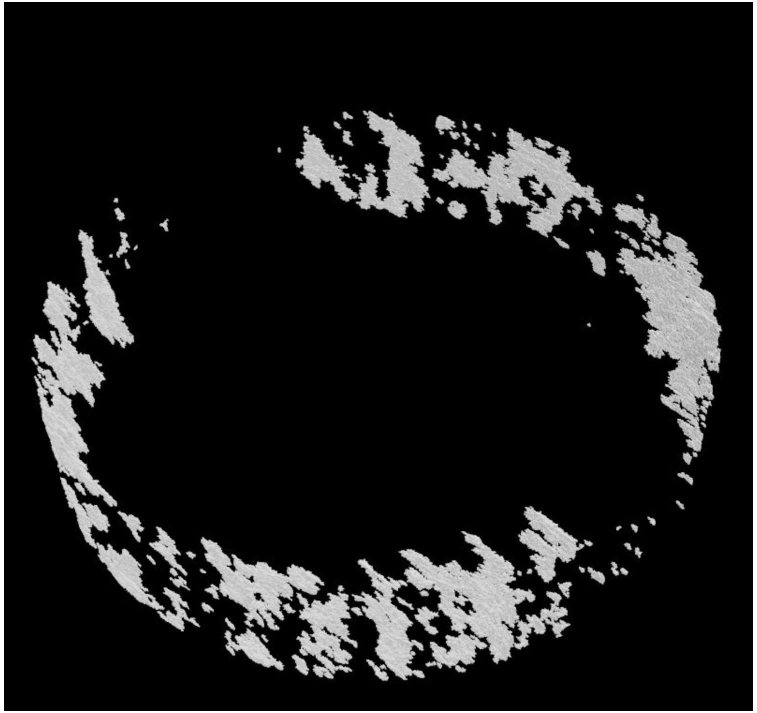


Are these fractures?

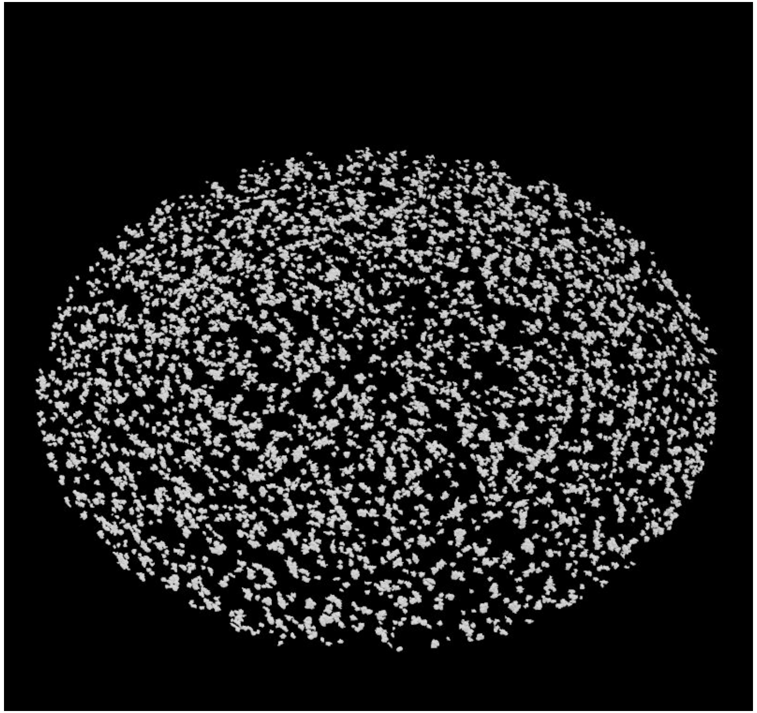
Help

3D Feature Reconstruction using Spatial Coherence

UNet Xception is used to generate predictions across the entire 342K image dataset



Fracture reconstruction: 250 slices

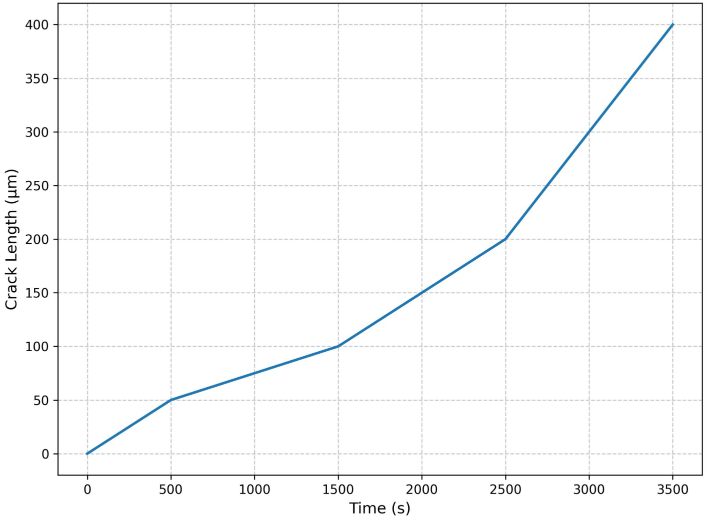


Inclusion reconstruction: 50 slices

Statistical Characterization: Feature Quantification

Sample-level Insights

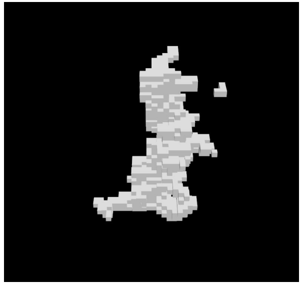
Inclusion Features	Value (pixel)	Value (μm)
Count	161,574	-
Average major axis	10.978 px	17.89 μm
Average volume	180.904 voxels	783.22 μm^2
Volume fraction	0.9%	0.9%



Defect-level Granularity

Query x feature that matches a set of attributes for thousands features detected

Largest inclusion at timestep 25 and attributes

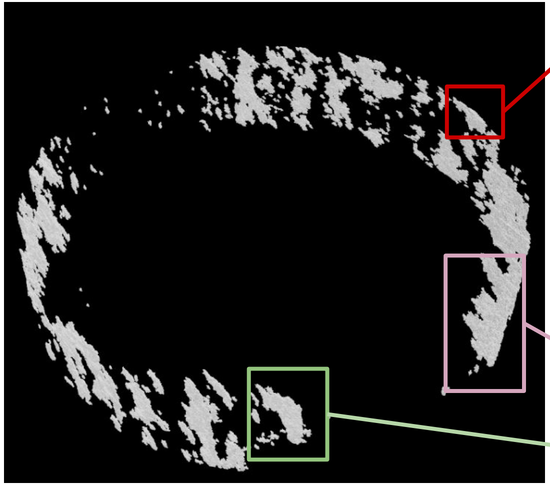


Inclusion Feature	Value
Major axis (px)	43.01
Volume (voxels)	1340

Automated extraction of 5 million+ total features throughout the 4D XCT dataset

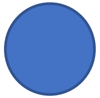
Summary Graph Generation

Translating 3D feature stacks and attributes into graphs

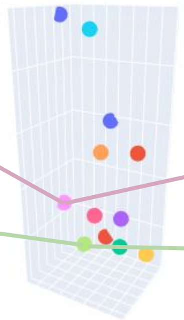


Fracture Features	Value
Major axis (px)	43.01
Volume (voxels)	1340
Orientation	intergranular

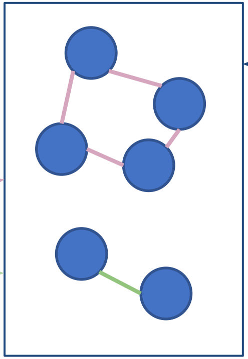
fracture embedded as node



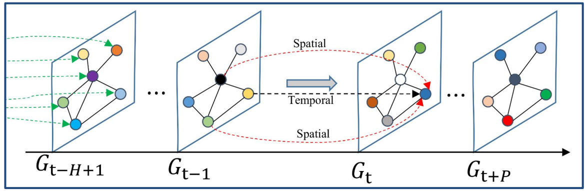
[43.01, 1340, 1]



Distance threshold or density clustering



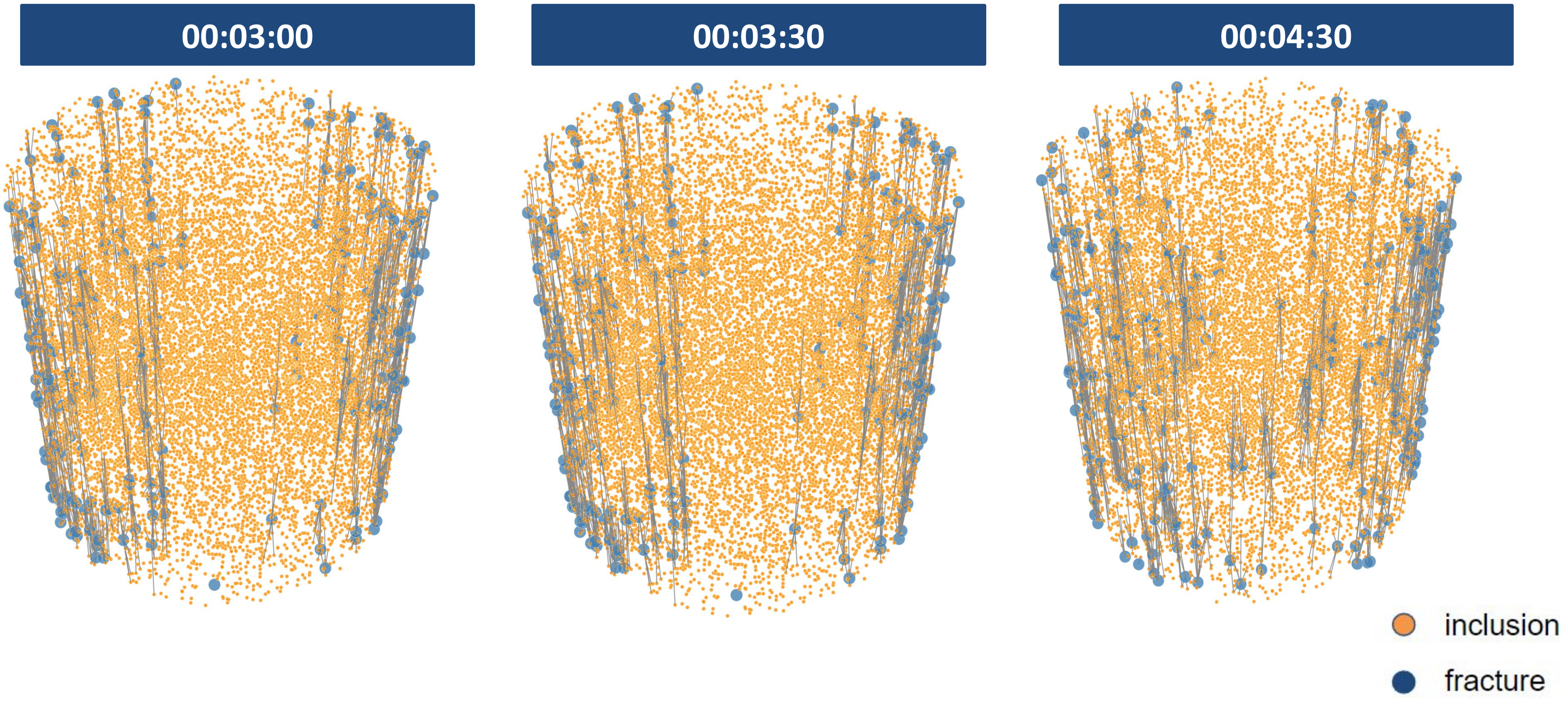
Graph representation



Graph neural network input

Summary Graph Representations

*Height scaled for visualization



Concluding Remarks

Conclusion

The following contributions were made through this research:

1. A **domain-informed diversity sampling strategy** designed for scientific data and "one-shot" annotations
2. **Scalable feature extraction** and segmentation framework that handles **sub-visible features**
3. Groundwork for **graph representations** for future degradation analysis
4. Code available under the XCTImage Python package

This framework is applied to an exemplar dataset of **342,000 XCT images** of stress corrosion cracking in AlMg to generate achieve a **0.94 F1-score** in segmentation only **labeling 0.03%** of the dataset and detect over **5 million microstructural defects**

Acknowledgements

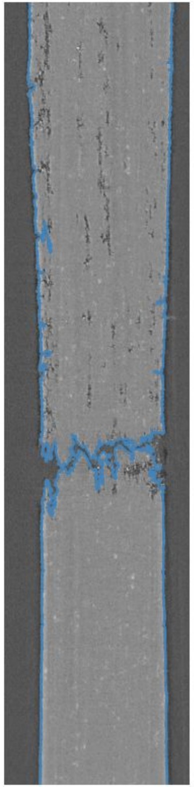
Thanks to all of the members of the SDLE Research Center!

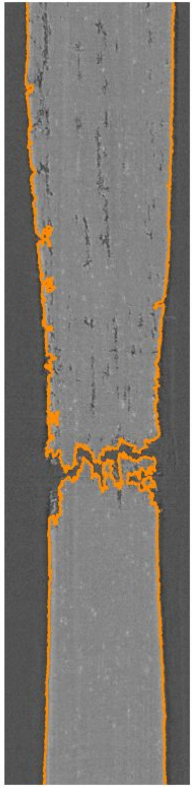
This work made use of the High Performance Computing Resource in the Core Facility for Advanced Research Computing at Case Western Reserve University.

This work was carried out with the support of the Diamond Light Source, on the Diamond-Manchester beamline I13-2 (proposal MT18165-1).

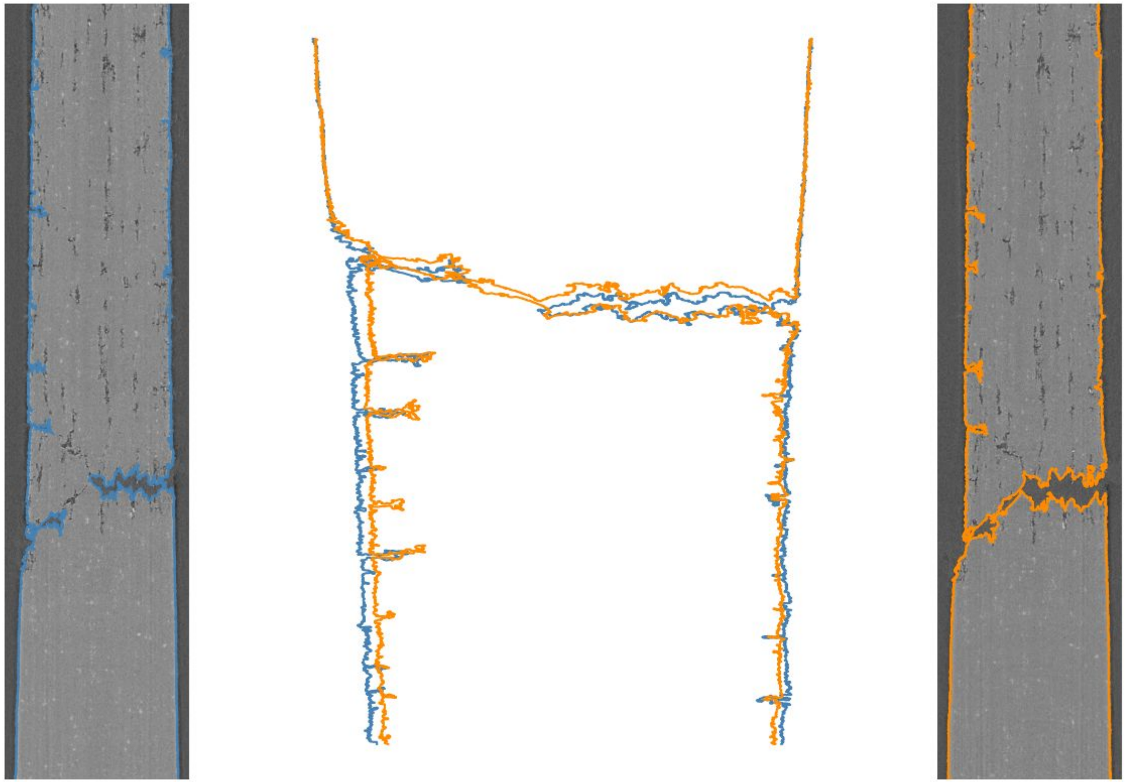
Appendix

Sample Distortion Over Time: X-axis

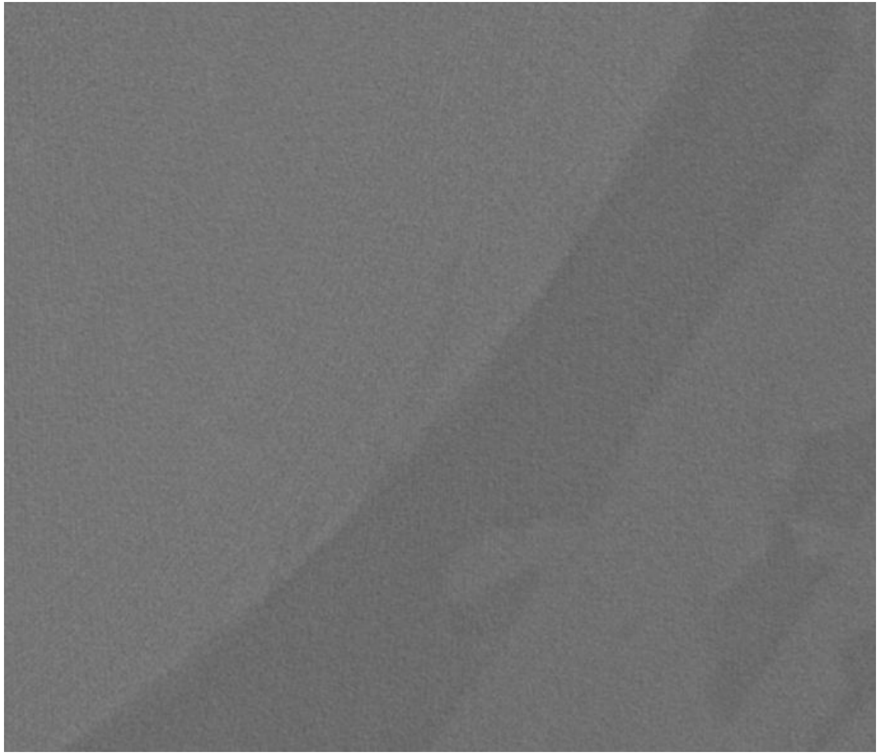




Sample Distortion Over Time: Y-axis



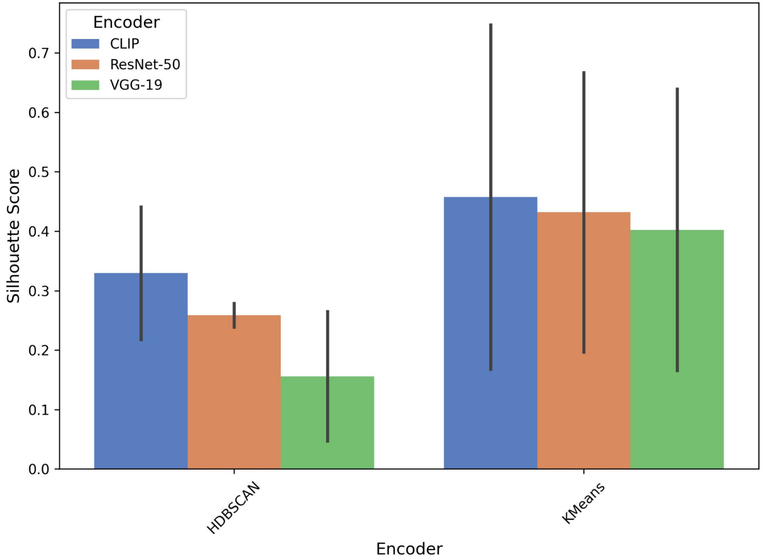
Dry Sample Resolution Issues

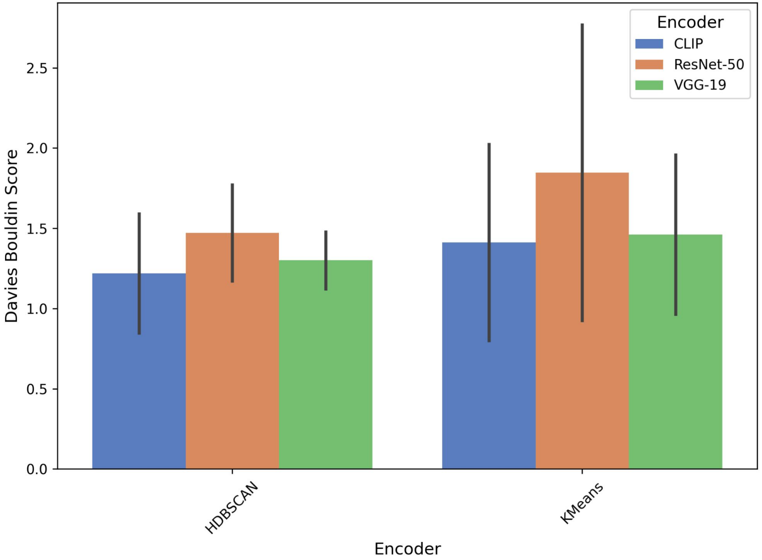


Results: Clustering Analysis

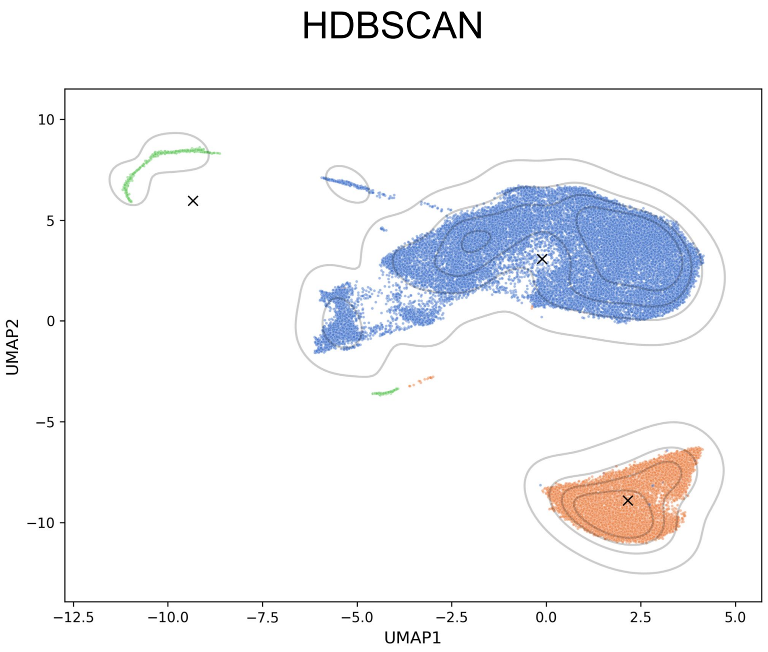
How well do we form clusters?

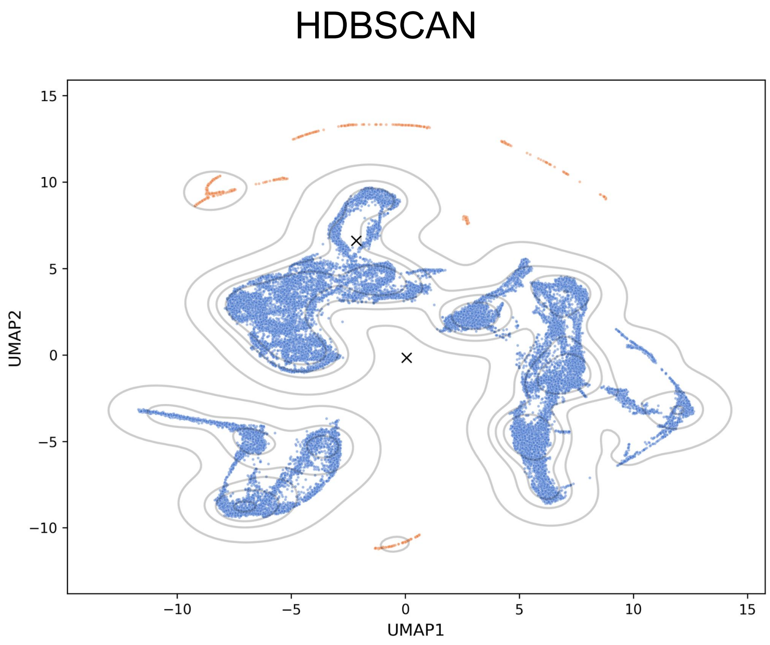
- Silhouette Score: inter-cluster evaluation
- Davies Bouldin: intra-cluster evaluation





Results: Clustering Examples





Clustering is difficult to parameterize across different encoders and techniques

Spatiotemporal Summary and Scene Graphs

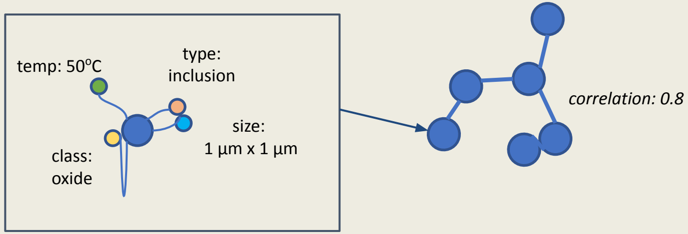
How can we ask more complex questions

- (E.g. do fractures tend to extend towards regions of higher defect density?)

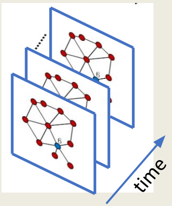
Generate scene graphs[1] for an interpretable full-scale microstructural and degradation analysis

Summary Graph Generation

Labeled features can be turned into nodes in a graph and then edges created between corresponding nodes

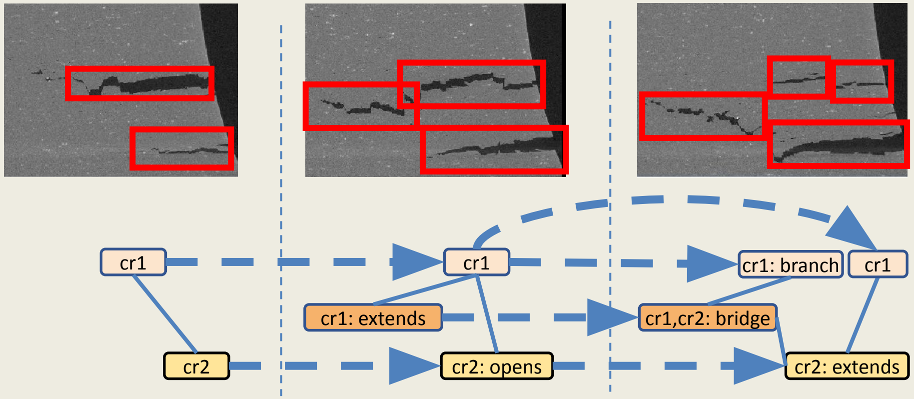


A single graph represents one point in time, multiple graphs can be stacked for temporal analysis



Spatiotemporal Scene Graph Generation

Scene graphs will be generated to label actions and relationships to identify what is occurring both spatially and temporally



[1] Ji, J., Krishna, R., Fei-Fei, L., & Niebles, J. C. (2020). Action genome: Actions as compositions of spatio-temporal scene graphs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10236-10247).