

# A Deep Learning Framework for Spatiotemporal Feature Extraction and Characterization of Synchrotron X-Ray Computed Tomography in Stress Corrosion Cracking of AlMg

THOMAS CIARDI

Thesis

Submitted in Partial Fulfillment of the Requirements for the Degree of  
Master of Science

Thesis Advisor: Dr. Yinghui Wu & Dr. Roger H. French

Department of Computer and Data Sciences

CASE WESTERN RESERVE UNIVERSITY

January, 2025

# **A Deep Learning Framework for Spatiotemporal Feature Extraction and Characterization of Synchrotron X-Ray Computed Tomography in Stress Corrosion Cracking of AlMg**

Case Western Reserve University  
Case School of Graduate Studies

We hereby approve the thesis<sup>1</sup> of

**Thomas Ciardi**

for the degree of

**Master of Science**

Dr. Yinghui Wu

---

Committee Chair, Advisor  
Department of Computer and Data Sciences

01.09.2025

Dr. Roger H. French

---

Committee Chair, Advisor  
Department of Materials Science and Engineering

01.09.2025

Dr. Jing Ma

---

Committee Member  
Department of Computer and Data Sciences

01.09.2025

---

<sup>1</sup>We certify that written approval has been obtained for any proprietary material contained therein.



*Dedicated to Progress in  
Science and Engineering*

## Table of Contents

List of Tables	vi
List of Figures	viii
Acknowledgements	xi
ABSTRACT	1
Chapter 1. Introduction	2
High-level Background	2
Research Objectives and Contributions	3
Chapter 2. Literature Review	5
Materials Domain	5
Data Science Domain	7
Materials Data Science	17
Chapter 3. Methods	20
Experimental Methods	20
Data Preprocessing	23
Diversity-Sampling Framework	25
Feature Extraction and Characterization Framework	34
Chapter 4. Results	43
Diversity Sampling Framework	43
Segmentation Performance Analysis	51
Microstructural Feature Characterization	56
Discussion	61
Chapter 5. Conclusions	71
Chapter 6. Appendix	73
Equations	73
Tables	75

Figures	77
References	86

## List of Tables

3.1	Chemical composition of AA5083-H131 plates (wt.%)	20
3.2	Mechanical testing and environmental parameters	22
3.3	Synchrotron XCT imaging parameters	23
3.4	Inclusion segmentation image processing pipeline parameters	37
3.5	Segmentation model architectures and training parameters	39
4.1	Number of components required to capture 95% variance in embeddings	44
4.2	Clustering performance metrics across pre-trained model and embedding spaces	47
4.3	Normalized diversity metrics and overall diversity score across different sampling strategies	49
4.4	Gamma sensitivity analysis by pre-trained model	51
4.5	Segmentation model performance metrics (all classes)	54
4.6	Segmentation model performance metrics (fracture class)	55
4.7	Segmentation model performance metrics (inclusion class)	55
4.8	Continuous crack growth characteristics during slow strain tension test	59
4.9	Segmented inclusion characterization metrics for sample scan	59
6.1	Pretrained model preprocessing parameters for embedding extraction	75
6.2	PCA parameters for embedding reduction	75
6.3	UMAP parameters for embedding reduction	75
6.4	K-means parameters for embedding clustering	75
6.5	DBSCAN parameters for embedding clustering	76
6.6	HDBSCAN parameters for embedding clustering	76

6.7	TypiClust sampling parameters	76
6.8	ProbCover sampling parameters	76

## List of Figures

3.1	Aluminum-magnesium (AlMg) dogbone cylinder sample prepared for slow strain tension test.	21
3.2	Experimental configuration inside the I13-2 beamline for in-situ X-ray computed tomography scans of slow strain tension test.	23
3.3	Image preprocessing pipeline. From left to right: 1) original XCT slice 2) center cropped XCT slice 3) final tiled images produced from sliding window over XCT slice.	24
3.4	High-level view of the diversity sampling pipeline. Images are reduced using a pre-trained encoder and dimensionality reduction. Subsequent embeddings are then clustered and diversity sampled for annotation.	26
3.5	Sample XCT slice and examples of fracture and inclusion features.	35
3.6	Spatiotemporal feature extraction and characterization pipeline which includes hybrid annotation generation, deep learning-based segmentation, feature reconstruction, and defect characterization.	36
3.7	Transformations for each step inside the image processing inclusion segmentation pipeline.	38
3.8	Inclusion pseudo-label example using weak supervision.	39
3.9	Architecture of U-Net with Xception as the encoder backbone.	40
4.1	2D UMAP embeddings of bottom center image tiles across different pre-trained models.	45
4.2	Examples of strong and poor clustering convergence for embeddings.	46
4.3	Distribution of unnormalized diversity metrics for different sampling methods.	48

4.4	Distribution of normalized diversity metrics for different sampling methods.	50
4.5	Distribution of raw diversity metrics for different $\gamma$ values.	52
4.6	Average change in metric stepping from $\gamma$ 1.0 to $\gamma$ 0.85 and then $\gamma$ 0.85 to $\gamma$ 0.5.	53
4.7	Segmentation mean Intersection over Union metrics across different architectures and pre-training strategies.	56
4.8	Sample predictions from U-Net with Xception backbone on images in test set. Fractures in white and inclusions in blue.	57
4.9	Example reconstruction segmentation predictions from model for fractures (left) and inclusions (right) on subset of volume.	58
4.10	Average fracture and growth rate evolution over experiment slow strain tension test.	60
4.11	Impact of displacement informed diversity sampling. Left is the scatter plot of UMAP embeddings with x marking selected samples. Right is the distribution of displacement values of those respective samples.	63
4.12	Diversity sampled images with $\gamma = 1.0$ .	64
4.13	Diversity sampled images with $\gamma = 0.75$ .	65
4.14	Segmentation F1 metrics across different architectures and encoders.	66
4.15	Sample predictions from SegFormer with MIT-B4 on images in test set. Fractures in white and inclusions in blue.	67
4.16	Summary graph representation and evolution of characterized microstructural features from segmentation pipeline.	69
6.1	Measured displacement of sample over time during the slow strain tension test.	77
6.2	CLIP embeddings PCA component cumulative variance.	78

6.3	ResNet50 embeddings PCA component cumulative variance.	79
6.4	VGG-19 embeddings PCA component cumulative variance.	80
6.5	Clustering silhouette scores compared by pre-trained model.	81
6.6	Clustering silhouette scores compared by dimensionality reduction technique.	82
6.7	Clustering Davies-Bouldin scores compared by pre-trained model.	83
6.8	Clustering Davies-Bouldin scores compared by dimensionality reduction technique.	84
6.9	LPIPS distribution shift comparing random sampling to diversity sampling.	85



## Acknowledgements

I would like to thank the funding sources that made this work possible. This material is based upon work supported in accordance with the provisions of NSF Solicitation: NSF 20-570 Industry-University Cooperative Research Centers Program. This material is also based upon research in the Materials Data Science for Stockpile Stewardship Center of Excellence (MDS3-COE), and supported by the Department of Energy's National Nuclear Security Administration under Award Number(s) DE-NA0004104. Computational analysis for this work made use of the High Performance Computing Resource in the Core Facility for Advanced Research Computing at Case Western Reserve University and supported by the U.S. National Science Foundation, Office of Advanced Cyberinfrastructure (OAC), Major Research Instrumentation, under Award Number 2117439.

This work would not have been possible without the guidance and support of many individuals. I would like to thank my professors and committee members: Dr. Yinghui Wu, Dr. Roger French, and Dr. Jing Ma. Their insights and encouragement have been invaluable in shaping both this research and my development as a scientist. A special thank you to Professor John Lewandowski for procuring the dataset used in this work and being a constant support system for answering my materials science questions. Finally, I would like to thank the lab members of SDLE research for their continued support, suggestions, and friendship.

## ABSTRACT

# A Deep Learning Framework for Spatiotemporal Feature Extraction and Characterization of Synchrotron X-Ray Computed Tomography in Stress Corrosion Cracking of AlMg

Thomas Ciardi

Spatiotemporal studies of material degradation have advanced significantly with the advent of high-resolution imaging techniques like synchrotron X-ray computed tomography (XCT). However, the massive Terabyte-scale datasets generated by these systems pose substantial challenges for analysis. Using environmental-induced cracking in marine-grade aluminum alloys as a case study, we introduce an integrated framework combining domain-informed diversity sampling and weak supervision strategies to enable automated microstructural analysis. Our novel diversity evaluation metric unifies embedding space coverage, perceptual similarity, and physical state representation, demonstrating our Domain-Informed Diversity Sampling (DIDS) achieves higher feature space coverage (average sample cosine distance improvement of +35-45% above baselines), broader displacement ranges (degradation state coverage of +10-45% above baselines), while maintaining superior perceptual diversity (DIDS: 0.358, ProbCover: 0.338, TypiClust: 0.324, Random: 0.315). Through weak supervision of only 90 images (<0.03% of the dataset), we achieve automated segmentation mean F1 scores of 0.949 (mean IoU of 0.646) in identifying over 5 million features with one sub-visible feature class being only 12 pixels in area on average. Applied to stress corrosion cracking analysis across 300,000+ images, this framework demonstrates unprecedented scalability in microstructural characterization. Our approach provides a general template for transforming massive materials characterization datasets into systematic, automated analyses while minimizing manual annotation requirements.

# 1 Introduction

## 1.1 High-level Background

Aluminum-magnesium (AlMg) alloys are crucial in marine applications due to their combination of high strength, low density, and strong corrosion resistance in saline environments and low temperatures<sup>1</sup>. AA5xxx series alloys, in particular AA5083, are widely used as a marine-grade material ship building material<sup>2</sup>. These material systems, however, suffer structural integrity issues with prolonged maritime service<sup>3,4</sup>. Specifically environmentally induced cracking (EIC) phenomenon composed of a spectra of stress corrosion cracking (SCC) and hydrogen embrittlement (HE) mechanisms<sup>5</sup>. EIC can occur at stress regimes below the yield strength of the material and result in catastrophic failures, posing significant safety risks to maritime structures and vessels<sup>6</sup>.

Historically, studying EIC has been difficult due to lack of instrumentation to capture and characterize these microscale mechanisms. Advancements in experimentation techniques and imaging technologies, however, have made it possible to study these degradation phenomena at appropriate time and length scales<sup>7</sup>. For example, synchrotron X-ray computed tomography (XCT) provides unprecedented capabilities for non-destructive, time-resolved imaging of crack initiation and propagation<sup>8</sup>. This allows for detailed tracking of crack networks, their morphology, and interaction with microstructural features to understand EIC<sup>9</sup>. The capture rates of these systems, however, generate a new problem; data at scale. Synchrotrons can capture information on the order of GB/s, presenting massive

data storage and analysis challenges<sup>10</sup>. Extracting meaningful quantitative information from these large 4D datasets to characterize the complex spatiotemporal behavior remains an open challenge.

This problem is not unique to characterization of EIC in AlMg, however. Materials science as a discipline has evolved to consistently generate vast amounts of data through experimentation. Effectively leveraging this scale of information requires scalable data engineering, machine learning, and high performance computing. The ability to successfully integrate materials science domain knowledge and data science together, enables the ability to study the mechanisms and science behind materials at new scales.

## 1.2 Research Objectives and Contributions

This research addresses fundamental challenges in characterizing environmental-induced cracking (EIC) in AlMg alloys through synchrotron XCT. While modern characterization techniques enable high-resolution 4D data collection, two critical challenges emerge in extracting meaningful insights from these massive datasets.

First, we confront the cold-start problem in spatiotemporal materials characterization, where effective sampling must capture both spatial features and their temporal evolution without prior knowledge of the degradation process. Modern synchrotron XCT experiments generate hundreds of thousands of images spanning multiple temporal states, but conventional sampling approaches struggle to identify representative subsets that adequately cover both the feature space and degradation progression. This challenge is particularly acute in materials science, where physical state evolution must be captured alongside spatial characteristics.

Second, we address the challenge of characterizing subvisible features in low-resolution tomography images, where critical microstructural elements exist at the limits of pixel-level detection. Intermetallic inclusions, crucial to understanding crack initiation and propagation, manifest as features merely 12 pixels in area, with hundreds present in each image. This creates an extreme annotation challenge

where traditional manual approaches become infeasible, particularly given the need to identify and characterize millions of features across our dataset.

Our integrated framework addresses these challenges through four key contributions:

- (1) A novel composite quality assessment framework that unifies embedding space coverage, perceptual similarity, and physical state representation into a single metric for evaluating sampling effectiveness in spatiotemporal datasets.
- (2) Domain-Informed Diversity Sampling (DIDS) that achieves higher feature space coverage (average sampled cosine distance improvement of +35-45% above baselines), broader displacement ranges (degradation state coverage of +10-45% above baselines), while maintaining superior perceptual diversity (DIDS: 0.358, ProbCover: 0.338, TypiClust: 0.324, Random: 0.315) within selected samples.
- (3) A scalable weak supervision pipeline achieving 0.949 mean F1 score and 0.646 mean IoU using only 90 annotated images (<0.03%) from a 300,000+ image dataset.
- (4) Automated detection and characterization of over 5 million microstructural features, including subvisible intermetallic inclusions averaging 12 pixels in area, enabling comprehensive analysis of material degradation at unprecedented scale.

## 2 Literature Review

This chapter summarizes the theoretical basis for relevant materials science and data science concepts within this work. The chapter also surveys the intersection of machine learning with materials science and its specific applications to synchrotron X-ray computed tomography (XCT).

### 2.1 Materials Domain

#### 2.1.1 5000 Series Aluminum Alloys

5000 series aluminum alloys constitute a fundamental class of non-heat-treatable materials that derive their strength primarily through solid solution hardening with magnesium<sup>11</sup>. The addition of magnesium, typically ranging from 4-5 weight percent, not only provides significant strengthening through solid solution effects but also maintains excellent weldability and formability characteristics<sup>2</sup>. These alloys, particularly AA5083, have found extensive use in marine and naval applications due to their exceptional combination of high strength, low density, and good corrosion resistance in saline environments<sup>1</sup>.

This class of alloys, however, faces a critical challenge regarding microstructural stability. At temperatures as low as 40 °C, commonly encountered in operational conditions, magnesium exhibits a tendency to precipitate from the solid solution<sup>3</sup>. This precipitation occurs preferentially at grain boundaries, resulting in the formation of  $\beta$ -phase ( $\text{Al}_3\text{Mg}_2$ )<sup>12,13</sup>. This phase is highly anodic compared to the aluminum matrix and creates conditions favorable for localized corrosion<sup>14-17</sup>.

The amount of  $\beta$ -phase precipitation at grain boundaries, quantified as the degree of sensitization (DoS), significantly influences the alloy's susceptibility to various forms of corrosion and environmental cracking<sup>18–20</sup>.

### 2.1.2 Stress Corrosion Cracking

Stress corrosion cracking (SCC) is a complex failure mechanism that emerges from the interaction between a material, tensile stress, and a corrosive environment<sup>5</sup>. In 5000 series aluminum alloys, this phenomenon manifests primarily as intergranular stress corrosion cracking (IGSCC), where crack propagation follows grain boundaries<sup>21</sup>. The crack propagation pattern in these alloys often shows complex branching behavior, with multiple crack paths developing simultaneously<sup>15</sup>.

The mechanism involves multiple interacting processes: anodic dissolution of the sensitized grain boundaries, hydrogen generation and uptake at crack tips, and mechanical loading effects<sup>13</sup>. The interaction between mechanical loading and environmental factors creates distinct stages of crack evolution: initiation, stable growth, and final failure<sup>22</sup>. The process typically initiates at surface defects or corrosion sites and can propagate at stress levels well below the material's yield strength<sup>3</sup>. This makes SCC particularly dangerous as failures can occur without warning under normal service loads<sup>6</sup>.

### 2.1.3 Synchrotron X-Ray Computed Tomography

X-ray computed tomography (XCT) represents a non-destructive imaging technique for studying material behavior at multiple length scales. XCT has been extensively used to image corrosion in AlMg alloys to better understand the degradation mechanisms<sup>7,9</sup>. Synchrotron XCT offers even higher-resolution capabilities to examine electrochemically and mechanically driven events in SCC. The technique utilizes high-energy X-rays generated by electron acceleration in a synchrotron facility to create three-dimensional representations of material structure<sup>23</sup>.

The key advantage of synchrotron XCT over conventional laboratory sources lies in its high flux and coherence, enabling rapid acquisition of high-resolution

data. The high temporal resolution allows for time-resolved imaging during mechanical testing, while the spatial resolution can capture fine details of crack networks and microstructural features<sup>8,24</sup>. The technique provides sufficient contrast to distinguish between different phases, voids, and cracks within the material, making it particularly suitable for studying the relationship between microstructural features and crack evolution<sup>25</sup>.

Synchrotron XCT, however, generates large volumes of data, with each scan typically producing several gigabytes (GB) of information. Experiments can produce hundreds of scans, generating datasets that are consistently on the order of terabytes (TB) in size<sup>26</sup>. This data richness provides unprecedented opportunities for studying material behavior but also presents significant challenges in data management, processing, and analysis<sup>10</sup>.

## 2.2 Data Science Domain

### 2.2.1 Dimensionality Reduction

Dimensionality reduction techniques transform high-dimensional data into lower-dimensional representations while preserving essential characteristics. This enables computational analysis at larger scales with increased efficiency by addressing the curse of dimensionality and removing uninformative noise.

Principal component analysis (PCA), is a foundational linear technique that identifies orthogonal axes of maximum variance in the data<sup>27</sup>. The transformation is defined by:

$$X_{reduced} = XW$$

where  $W$  contains the top  $k$  eigenvectors of the covariance matrix  $XX^T$ .

t-distributed Stochastic Neighbor Embedding (t-SNE) is a staple non-linear dimensionality reduction technique that aims to retain local structures<sup>28</sup>. The approach models similarities between points as conditional probabilities:



$$p_{j|i} = \frac{\exp(-|x_i - x_j|^2/2\sigma_i^2)}{\sum_{k \neq i} \exp(-|x_i - x_k|^2/2\sigma_i^2)}$$

where the similarity of data point  $x_i$  to  $x_j$  is the conditional probability  $p_{j|i}$  that  $x_i$  would be the neighbor of  $x_j$  if selected in proportion to their probability density under a Gaussian centered at  $x_i$ .

Uniform Manifold Approximation and Projection (UMAP) provides a theoretically grounded alternative to t-SNE based on Riemannian geometry and algebraic topology<sup>29</sup>. It constructs a fuzzy topological representation of high-dimensional data and optimizes a low-dimensional layout that aims to preserve both global and local structure through:

$$\sum_{i,j} \left[ v_{ij} \log \left( \frac{v_{ij}}{w_{ij}} \right) + (1 - v_{ij}) \log \left( \frac{1 - v_{ij}}{1 - w_{ij}} \right) \right]$$

where  $v_{ij}$  and  $w_{ij}$  represent high and low-dimensional similarities respectively.

### 2.2.2 Clustering Algorithms

Unsupervised learning represents a fundamental paradigm in machine learning where algorithms discover inherent patterns and structure in data without labeled examples. Within this domain, clustering serves as a core technique that aims to partition data points into groups (clusters) such that points within the same cluster are more similar to each other than to those in other clusters, typically measured through some distance or density metric.

K-means clustering represents one of the most simple and widely adopted partitioning methods<sup>30</sup>. The algorithm iteratively assigns points to the nearest of  $k$  centroids and updates these centroids based on cluster membership, minimizing the within-cluster sum of squares:

$$J = \sum_{i=1}^k \sum_{x \in C_i} |x - \mu_i|^2$$

where  $C_i$  represents the  $i$ -th cluster and  $\mu_i$  its centroid. Despite its simplicity, k-means has proven remarkably effective across diverse applications, though it

assumes spherical cluster shapes and requires prior specification of the number of clusters.

Density-based clustering methods emerged to address these limitations, with DBSCAN (Density-Based Spatial Clustering of Applications with Noise) representing a significant advance<sup>31</sup>. DBSCAN defines clusters as dense regions separated by areas of lower density, using two key parameters: the neighborhood radius ( $\epsilon$ ) and the minimum number of points required to form a dense region ( $MinPts$ ). The density around a point  $p$  is estimated through:

$$N_\epsilon(p) = \{q \in D \mid dist(p, q) \leq \epsilon\}$$

where  $D$  is the dataset and  $dist$  is a distance metric. A point  $p$  is considered a core point if  $|N_\epsilon(p)| \geq MinPts$ . This approach effectively identifies arbitrary-shaped clusters and naturally handles noise, though parameter selection can significantly impact results.

HDBSCAN (Hierarchical DBSCAN) further advances density-based clustering by eliminating the need for a fixed density threshold<sup>32</sup>. The algorithm constructs a hierarchical cluster tree and extracts a flat clustering that optimizes cluster stability across different density levels. This adaptation provides robust performance across varying density scales while maintaining DBSCAN's capability to identify arbitrary-shaped clusters. By converting DBSCAN into a hierarchical clustering algorithm, HDBSCAN effectively handles clusters of varying densities within the same dataset.

### 2.2.3 Machine Learning

Machine learning encompasses computational methods that enable systems to improve their performance on a task through experience. At its core, supervised learning involves finding a function  $f$  that maps inputs  $x$  to outputs  $y$  by minimizing a loss function  $L(y, f(x))$  over a training dataset.

The fundamental building block, a neural network layer, performs the operation:

$$h = \sigma(Wx + b)$$

where  $W$  represents learnable weights,  $b$  is a bias term, and  $\sigma$  is a non-linear activation function<sup>33</sup>. Modern networks stack multiple such layers, with each subsequent layer learning increasingly abstract features. The learning process employs backpropagation, where gradients of the loss function are computed with respect to parameters:

$$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial h} \frac{\partial h}{\partial W}$$

#### 2.2.4 Deep Learning

Deep learning is a subset of machine learning that extends early foundations of artificial neural networks to provide state of the art performance in various tasks including natural language processing, image classification, and generative modeling. Here we describe several key modern architectures and their underlying structures.

Convolutional Neural Networks (CNNs) revolutionized computer vision by introducing architecturally-enforced spatial structure through convolution operations<sup>34</sup>:

$$(f * g)(p) = \sum_{s+t=p} f(s)g(t)$$

where  $f$  is the input (e.g., image or feature map),  $g$  is the convolution kernel (learnable filter),  $p$  represents the output position,  $s$  and  $t$  are positions in the input and kernel respectively, and  $*$  denotes the convolution operation. The convolution operation learns local feature detectors that scan across the input space, providing translation invariance and parameter efficiency.

Modern CNN architectures combine three essential components for effective visual processing. Convolutional layers form the backbone, using learnable filters to detect hierarchical patterns at different scales, with each layer building upon features learned by previous ones through non-linear activation functions. Pooling

operations provide spatial downsampling by selecting maximum values within local neighborhoods, reducing complexity while maintaining key features and providing translation invariance. Batch normalization stabilizes training by normalizing layer activations using mini-batch statistics, enabling faster convergence through learnable scale and shift parameters. Together, these components create a powerful hierarchy of feature detection, spatial summarization, and activation normalization that has revolutionized computer vision tasks.

The introduction of attention mechanisms, through the Transformer architecture, marked a fundamental advance in sequence modeling and neural architecture design<sup>35</sup>. At its core, the Transformer relies on self-attention operations that compute weighted relationships between all pairs of features in a sequence:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where  $Q$ ,  $K$ , and  $V$  represent query, key, and value matrices respectively, derived from learned linear projections of the input sequence:  $Q = XW^Q$ ,  $K = XW^K$ ,  $V = XW^V$ . The query matrix represents features doing the attending, while keys represent features being attended to, and values contain the features to be aggregated. The scaling factor  $\sqrt{d_k}$  prevents extremely small gradients in the deeper layers of the network.

The Transformer extends this mechanism through multi-head attention, which enables parallel attention computations across different representation subspaces. Multiple sets of  $Q$ ,  $K$ ,  $V$  projections operate independently before being concatenated and projected to the final output. This multi-head structure allows the model to simultaneously capture different types of dependencies and relationships at various scales and semantic levels, significantly enhancing the model's ability to process complex sequential data.

### 2.2.5 Image Segmentation

Semantic segmentation, the pixel-wise classification of images, represents a fundamental technique for detailed visual analysis across diverse application domains.

Contemporary approaches predominantly employ encoder-decoder architectures, where the encoder extracts hierarchical feature representations while the decoder reconstructs detailed segmentation maps from these features. This architectural paradigm enables both efficient feature extraction and precise spatial reconstruction.

The effectiveness of segmentation models is commonly evaluated using the Intersection over Union (IoU) metric:

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|}$$

where  $A$  represents the predicted segmentation mask and  $B$  represents the ground truth mask. This metric quantifies the overlap between predicted and actual segmentations, providing a robust measure of segmentation accuracy.

### 2.2.6 Deep Learning Segmentation Architectures

Deep learning architectures for semantic segmentation have evolved to address the fundamental challenge of simultaneously maintaining global context and fine spatial detail in dense pixel-wise predictions. This evolution spans from CNN-based encoder-decoder structures to recent transformer-based approaches, each introducing novel mechanisms for multi-scale feature processing and information flow.

U-Net pioneered a symmetric encoder-decoder architecture that revolutionized medical image segmentation through its distinctive skip connections between corresponding encoder and decoder layers:

$$f_{\text{decoder}}^l = \text{Up}(f_{\text{decoder}}^{l-1}) \oplus f_{\text{encoder}}^l$$

where  $\oplus$  denotes concatenation, Up represents upsampling, and  $l$  indicates the layer level. The contracting path captures contextual information through successive convolution and pooling operations, while the expansive path enables precise localization through upsampling and feature concatenation<sup>36</sup>.

UPerNet advances these capabilities by introducing a unified framework for multi-scale feature processing, combining a Pyramid Pooling Module (PPM) for global context capture with a Feature Pyramid Network (FPN) for bidirectional information flow across resolution levels. This multi-scale fusion strategy, coupled with a flexible backbone network, enables effective handling of complex scene parsing tasks across varying scales and semantic categories<sup>37</sup>.

SegFormer represents a transformer-based approach that achieves superior performance while maintaining computational efficiency. Its hierarchical structure eliminates the need for positional encodings, progressively reduces sequence lengths while increasing channel dimensions, and employs a lightweight All-MLP decoder for efficient multi-level feature aggregation<sup>38</sup>. This design bridges the gap between global context modeling and fine-grained localization inherent in segmentation tasks.

### 2.2.7 Deep Learning Backbones

Feature extraction backbones have evolved from simple convolutional architectures to sophisticated designs incorporating residual connections, attention mechanisms, and transformer blocks. Each innovation addresses specific challenges in deep network training and efficiency, while providing increasingly powerful feature representations. These backbones can be leveraged in deep learning architectures such as U-Net and UPerNet to improve task performance.

VGG19 established fundamental principles for deep CNN architecture design through its systematic use of small (3×3) convolutional filters and deep layer stacking. This simple yet effective architecture demonstrated that network depth is crucial for learning hierarchical features, though its dense structure leads to high computational demands<sup>39</sup>.

ResNet50 revolutionized deep network training by introducing residual learning through skip connections:

$$y = F(x, W_i) + x$$

where  $F$  represents residual mapping and  $x$  is the identity mapping. This elegant solution to the vanishing gradient problem enabled effective training of substantially deeper networks<sup>40</sup>.

SE-ResNeXt101 enhances the ResNet architecture with Squeeze-and-Excitation (SE) blocks that perform dynamic channel-wise attention:

$$s = F_{scale}(F_{squeeze}(F_{transform}(x)))$$

where  $F_{squeeze}$  captures global channel dependencies and  $F_{scale}$  performs adaptive feature recalibration, enabling the network to emphasize informative features while suppressing less useful ones<sup>41</sup>.

Xception advances computational efficiency through depthwise separable convolutions:

$$y = F_{pointwise}(F_{depthwise}(x))$$

This factorization of standard convolutions into channel-wise and point-wise operations significantly reduces computational complexity while still maintaining model capacity<sup>42</sup>.

CLIP introduced a powerful vision encoder trained through natural language supervision on 400 million image-text pairs. Its visual backbone employs a modified ResNet or Vision Transformer (ViT) architecture, learning robust visual representations that align with semantic concepts through contrastive learning against text encodings<sup>43</sup>.

MiT-Bx (Mix Transformer) variants provide hierarchical transformer-based encoding with progressive reduction in sequence length and increase in channel dimensions across stages. These backbones offer flexible capacity-efficiency trade-offs through varying model sizes (B0 to B4), enabling deployment across different computational constraints while maintaining the advantages of self-attention for feature extraction<sup>38</sup>.

### 2.2.8 Learned Perceptual Image Similarity

Visual diversity metrics have evolved from simple pixel-wise comparisons metrics to learned perceptual measures that better align with human judgments. Learned Perceptual Image Patch Similarity (LPIPS) leverages deep neural network features to approximate human perceptual differences, typically using AlexNet as its backbone network<sup>44</sup>. AlexNet's architecture pioneered modern deep learning through its effective use of ReLU activations, deep convolutional structure, and dropout regularization<sup>45</sup>.

The LPIPS distance between two images is computed using features extracted from multiple layers of the network:

$$d_{\text{LPIPS}}(x, y) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} |w_l \odot (F_l^x(h, w) - F_l^y(h, w))|_2^2$$

where  $x$  and  $y$  are the input images,  $F_l^x$  and  $F_l^y$  are features extracted at layer  $l$ , and  $w_l$  are learned channel weights that determine each feature's importance. The features are extracted through AlexNet's five convolutional layers, each followed by ReLU activation and, in early layers, max pooling. This hierarchical structure allows LPIPS to capture perceptual differences at multiple scales and levels of abstraction, from low-level edges to higher-level patterns.

### 2.2.9 Active Learning

Active learning optimizes dataset creation by strategically selecting the most informative samples for annotation. The approach encompasses two primary sampling paradigms: uncertainty-based and diversity-based selection, each addressing different aspects of the sample selection challenge.

Uncertainty-based sampling identifies samples where the model exhibits low confidence, quantified through various information-theoretic metrics:

$$\text{Entropy: } H(x) = - \sum_i p_i(x) \log(p_i(x))$$



$$\text{Variance: } \sigma^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \mu)^2$$

$$\text{Mutual Information: } I(X; Y) = H(E[p(Y|X)]) - E[H(p(Y|X))]$$

where  $p_i(x)$  represents class probabilities,  $y_i$  are model predictions, and  $\mu$  is the mean prediction.

Diversity-based sampling ensures comprehensive feature space coverage by selecting samples that maximize representational variation. Two prominent approaches have emerged for implementing this strategy:

TypiClust combines representation learning with clustering-based diversity sampling<sup>46</sup>. Given an unlabeled pool  $U$ , the algorithm first learns embeddings  $f(x)$  for each sample, then partitions the embedding space into  $B$  clusters (where  $B$  is the annotation budget):

$$C_k = \{x_i \in U : k = \arg \min_j |f(x_i) - \mu_j|^2\}$$

For each cluster, it selects the most typical example using a density-based typicality score:

$$\text{typicality}(x) = \frac{1}{|N_\epsilon(x)|} \sum_{y \in N_\epsilon(x)} \exp(-|f(x) - f(y)|^2 / \sigma^2)$$

where  $N_\epsilon(x)$  represents the  $\epsilon$ -neighborhood of  $x$  in the embedding space.

ProbCover implements a graph-based diversity sampling strategy<sup>47</sup>. It constructs a directed graph  $G = (V, E)$  where:

$$E = \{(u, v) : \|f(u) - f(v)\| \leq \delta\}$$

The coverage score for each vertex is computed as:

$$\text{score}(v) = |\{u \in V : (v, u) \in E \text{ and } u \text{ not covered}\}|$$

The algorithm iteratively selects vertices maximizing:

$$v^* = \arg \max_{v \in V} \text{score}(v)$$

This greedy optimization ensures maximum coverage while maintaining computational efficiency through sparse graph representations.

Hybrid approaches implement a two-stage process combining both strategies. First, diversity sampling establishes initial coverage, then uncertainty-based refinement targets model weaknesses<sup>48</sup>. This combination leverages the complementary strengths of both paradigms: comprehensive feature space exploration and focused refinement of decision boundaries.

### 2.2.10 Weak Supervision

Weak supervision addresses the challenge of limited labeled data by leveraging automated methods to generate initial training labels. Rather than requiring extensive manual annotation, this approach uses various sources of weak signals to create approximate labels that can bootstrap model training<sup>49</sup>.

Traditional image processing techniques provide one approach to weak supervision, utilizing domain knowledge encoded in classical algorithms. For example, Hernandez et al. leveraged classical transformations such as contrast enhancements and thresholding to generate weak labels of melt pools in laser powder bed fusion images<sup>50</sup>. While these methods may not achieve the accuracy of deep learning approaches, they can generate reasonable initial masks by exploiting known image characteristics like feature brightness or shape.

Pre-trained models offer another powerful source of weak supervision, particularly through transfer learning and zero-shot approaches. Models trained on large-scale datasets like ImageNet or domain-specific collections can provide rich feature representations that generalize to new tasks<sup>51,52</sup>. For instance, in biological imaging, pre-trained models can identify cellular structures in new experimental contexts, even without task-specific fine-tuning<sup>53,54</sup>. These masks, while imperfect, capture sufficient structural information to train deep learning models that eventually surpass the accuracy of the original weak supervision signal.

## 2.3 Materials Data Science

### 2.3.1 Deep Learning in Materials Science

The integration of deep learning into domain sciences has transformed traditional research methodologies across multiple disciplines. This has resulted in novel scientific breakthroughs in topics such as protein folding predictions, cryo-EM particle picking, and climate change modeling<sup>55–57</sup>. Materials science, in particular, has seen the significant adoption of these techniques due to the increasing complexity and volume of experimental data<sup>26,58</sup>.

Early applications of deep learning in materials science included simple classification and regression tasks but recent developments address complex challenges such as microstructure characterization, property prediction, and process optimization<sup>52,59,60</sup>. A key development has been the emergence of domain-adapted architectures that incorporate physical constraints and domain knowledge into neural network designs<sup>61</sup>. This hybridization of deep learning with scientific principles has led to models that not only predict but also provide insights into underlying physical phenomena<sup>62</sup>.

### 2.3.2 Deep Learning in Synchrotron X-Ray Computed Tomography

Synchrotron XCT data presents a particularly unique challenge and opportunity for deep learning applications. The technique generates massive datasets with complex 3D structures, temporal evolution, and multiple contrast mechanisms<sup>10,26</sup>.

Deep learning has been employed in synchrotron XCT to address a variety of challenges. Strohmman et al. leveraged CNN's for scalable, robust segmentation of microstructural features in synchrotron XCT scans of Al-Si alloys<sup>63</sup>. Karamov et al. has demonstrated the use of Generative Adversarial Networks (GANs) for the super-resolution of synchrotron XCT images of unidirectional composites to improve fibre break analysis<sup>64</sup>. Hendriksen et al. improved denoising strategies for reconstructed images using deep learning approaches<sup>65</sup>. Scan reconstruction time and accuracy has been improved through deep learning architectures in the work of Shin et al.<sup>66</sup>.

These applications have enabled new capabilities in materials characterization. However, significant challenges remain. The high dimensionality of XCT data requires specialized architectural considerations. The need for accurate quantification demands robust uncertainty estimation. Additionally, the variety of experimental conditions and imaging modes requires models that can generalize across different experimental setups and material systems.

## 3 Methods

### 3.1 Experimental Methods

#### 3.1.1 Material and Sample Preparation

The experimental material comprised commercially available AA5083-H131 alloy plates (29 mm thickness), with composition determined via optical emission spectroscopy as detailed in Table 3.1. Intergranular corrosion (IGC) susceptibility was evaluated using the ASTM G67 Nitric Acid Mass Loss Test (NAMLT), yielding a degree of sensitization (DoS) of  $8.6 \text{ mg cm}^{-2}$  for as-received material<sup>67</sup>.

The microstructure exhibited typical rolling-induced pancake-shaped grains with dimensions of  $150 \mu\text{m}$  (L),  $80 \mu\text{m}$  (T), and  $35 \mu\text{m}$  (S)<sup>68</sup>. Mechanical properties in the short transverse direction were characterized by a yield strength (YS) of 260 MPa, ultimate tensile strength (UTS) of 375 MPa, and plane-strain fracture toughness ( $K_{IC}$ ) of  $31 \text{ MPa}\sqrt{\text{m}}$ <sup>68</sup>. Cylindrical tensile specimens were machined in the short-transverse orientation from the received plates to minimize crack tortuosity<sup>69</sup>. Samples had a gauge length of 12.7 mm and diameter of 3.2 mm<sup>4</sup>.

The specimen preparation involved several sequential steps. First, mechanical polishing was performed using P4000 SiC abrasive paper, followed by ultrasonic cleaning in ethanol and warm air drying. The samples then underwent a

Table 3.1. Chemical composition of AA5083-H131 plates (wt.%)

Element	Mg	Mn	Fe	Cr	Si	Zn	Ti	Ni	Sn	Pb	Al
wt.%	4.46	0.62	0.12	0.08	0.07	0.02	0.02	<0.01	<0.01	0.01	Bal.

sensitization treatment at 80 °C for 225 h, resulting in a degree of sensitization of approximately 40 mg cm<sup>-2</sup>. Next, specimens were pre-exposed in a 0.6 mol NaCl solution at room temperature for 180 h to introduce surface intergranular corrosion and facilitate hydrogen charging. Finally, post-exposure processing included thorough rinsing and careful drying of the specimens, followed by masking of the gauge length extremities using Parafilm™ (Bemis Company, Inc.). The Parafilm™ masking served dual purposes for prevention of shoulder region failure and environmental exposure limitation to match X-ray imaging field of view (FOV). A depiction of the prepared sample can be found in Figure 3.1.

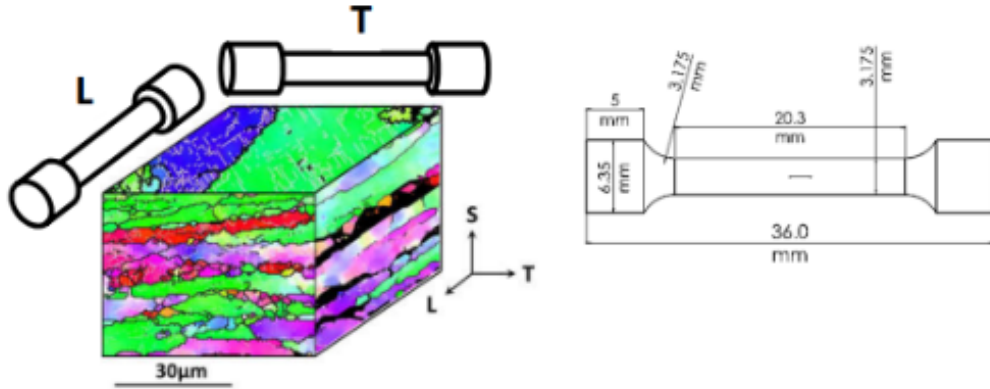


Figure 3.1. Aluminum-magnesium (AlMg) dogbone cylinder sample prepared for slow strain tension test.

### 3.1.2 Slow Strain Tension Test

The slow strain rate testing (SSRT) was conducted using a Deben CT5000 testing system with 5 kN capacity (Deben, UK). Samples were deformed at a nominal strain rate of  $4 \times 10^{-5} \text{ s}^{-1}$  until reaching the ultimate tensile strength (UTS). Upon reaching UTS, displacement was held constant until sample failure occurred. To prevent motion artifacts in the tomographic imaging, the straining was temporarily paused during each XCT acquisition (duration  $\sim 90 \text{ s}$ ) in the pre-UTS regime, resulting in a stepwise loading profile.

Environmental conditions were controlled using a custom vitreous glassy carbon chamber. Humidity was maintained at approximately 70% relative humidity

Table 3.2. Mechanical testing and environmental parameters

Parameter	Value
Testing Machine	Deben CT5000
Load Capacity	5 kN
Nominal Strain Rate	$4 \times 10^{-5} \text{ s}^{-1}$
Loading Profile	continuous to UTS, then fixed displacement
Scan Interruption Time	$\sim 90 \text{ s}$
Relative Humidity	$\sim 70\%$
Temperature	room temperature
Sponge Distance from Sample	20 mm

at room temperature by positioning wet sponges approximately 20 mm below the sample gauge length. Detailed parameters for the SSRT are contained in Table 3.2

### 3.1.3 Synchrotron XCT Scanning

The experiment was conducted at Diamond Light Source's I13-2 imaging beamline. To achieve optimal imaging conditions, a pink beam configuration was employed which maximized flux at roughly 28 keV<sup>70</sup>. The imaging setup provided a ( $1.63 \mu\text{m}^3$ ) voxel resolution over a field of view measuring  $4.2 \text{ mm} \times 3.5 \text{ mm}$ . Each tomographic acquisition comprised 1200 projections with 0.035 s exposure time per projection, resulting in approximately a 90 s total acquisition time per volume.

Throughout the in-situ mechanical testing, which spanned 13,260 s (3 h and 41 min), 43 tomograms were collected. The acquisition protocol included a high-quality reference scan (0.09 s exposure) before loading, followed by regular interval scans until yield strength was reached. Upon approaching the ultimate tensile strength (UTS), the setup was transitioned to continuous scanning. A final high-quality scan (0.09 s exposure) was performed post-test. Volume reconstruction was implemented using the Gridrec algorithm within TomoPy, integrated into a Savu processing pipeline<sup>71</sup>. The experimental setup is depicted in Figure 3.2 and synchrotron parameters in Table 3.3.

Table 3.3. Synchrotron XCT imaging parameters

Parameter	Value
Beamline	I13-2, Diamond Light Source
Beam Configuration	pink beam
Nominal Energy	28 keV
Field of View	$4.2 \times 3.5$ mm
Voxel Size	$1.63 \mu\text{m}^3$
Projections per Scan	1200
Standard Exposure Time	0.035 s/projection
Reference Scan Exposure	0.09 s/projection
Acquisition Time per Volume	$\sim 90$ s
Total Number of Tomograms	43
Total Test Duration	13,260 s

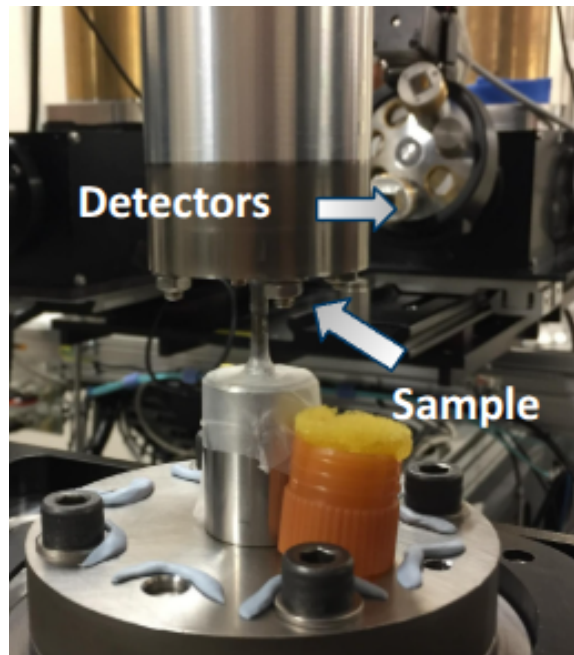


Figure 3.2. Experimental configuration inside the I13-2 beamline for in-situ X-ray computed tomography scans of slow strain tension test.

## 3.2 Data Preprocessing

The XCT volumes consist of 2,110 sequential TIFF images in a stack, each a resolution of  $2510 \times 2510$  pixels. While these high-resolution images capture fine details,



their large dimensionality poses memory challenges for deep learning applications. To address this, we implement a preprocessing pipeline that preserves the small features needed for segmentation while making the data more manageable.

The preprocessing workflow consists of two main steps: 1) apply center cropping with a 615-pixel padding in each direction to remove unnecessary background around the sample 2) employ a sliding window technique with a 384-pixel stride, which divides each cropped image into nine 512×512 tiles. This tiling approach ensures that we maintain sufficient resolution to detect small features while creating appropriately sized inputs for our deep learning models. Figure 3.3 depicts an example of an original and preprocessed image.

The research initially collected 43 tomographic scans, but only final 20 were selected for detailed analysis due to significant vertical shifting in early scans that would have compromised temporal analysis. From this subset of 20 scans, two were further excluded due to poor image resolution. The final pipeline was applied to 18 high-quality tomograms, ultimately generating a comprehensive dataset of 341,820 tiled images for in-depth examination.

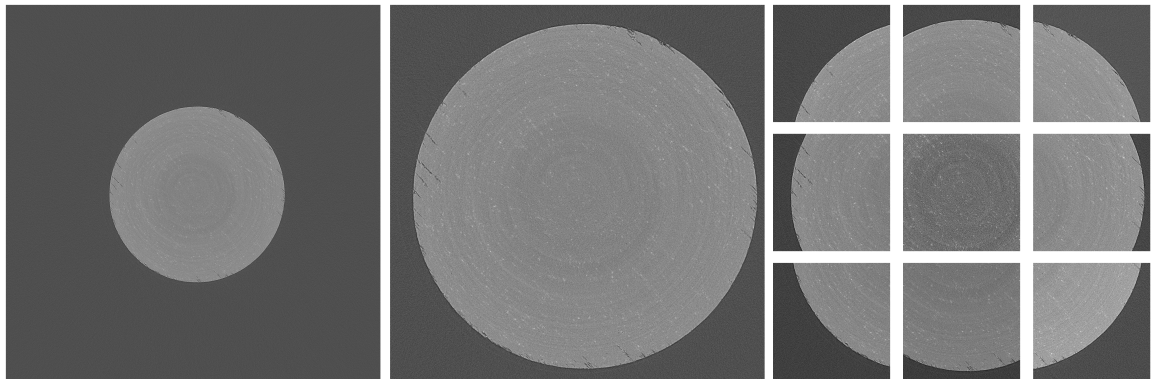


Figure 3.3. Image preprocessing pipeline. From left to right: 1) original XCT slice 2) center cropped XCT slice 3) final tiled images produced from sliding window over XCT slice.

### 3.3 Diversity-Sampling Framework

To address the cold start problem in annotation-constrained environments, we develop a comprehensive framework for analyzing microstructural evolution in XCT images. Our framework consists of three main components: 1) feature extraction and representation, 2) sampling strategy selection, and 3) quality assessment. Figure 3.4 showcases the high-level steps in the sampling pipeline.

The feature extraction pipeline leverages transfer learning through pre-trained vision models to capture the visual representations of material structures. These high-dimensional embeddings undergo dimensionality reduction and clustering to identify distinct structural patterns. The extracted representations serve as input for both baseline sampling methods and our novel Domain-Informed Diversity Sampling (DIDS) approach.

Our framework specifically targets the challenge of building initial training sets that comprehensively represent complex material variations with minimal labeling effort. By integrating perceptual metrics with domain-specific physical measurements, we enable precise feature extraction while dramatically reducing annotation requirements. The effectiveness of different sampling strategies is evaluated through a multi-metric quality assessment framework that considers latent space coverage, visual diversity, and physical degradation progression.

#### 3.3.1 Embedding Extraction using Pre-trained Models

To capture representations of visual variations, we extract perceptual embeddings from three ImageNet pre-trained models to enable diverse sampling of our tiled XCT dataset<sup>51</sup>. The CLIP Vision Transformer (ViT-Base/32) generates 512 dimensional embeddings from its final projection layer, capturing high-level semantic concepts through attention mechanisms<sup>72</sup>. ResNet50 generates 2048 dimensional embeddings from its global average pooling layer, offering deep residual representations<sup>40</sup>. VGG-19 provides 4096 dimensional features extracted from its first fully connected layer, representing hierarchical convolutional features<sup>39</sup>.

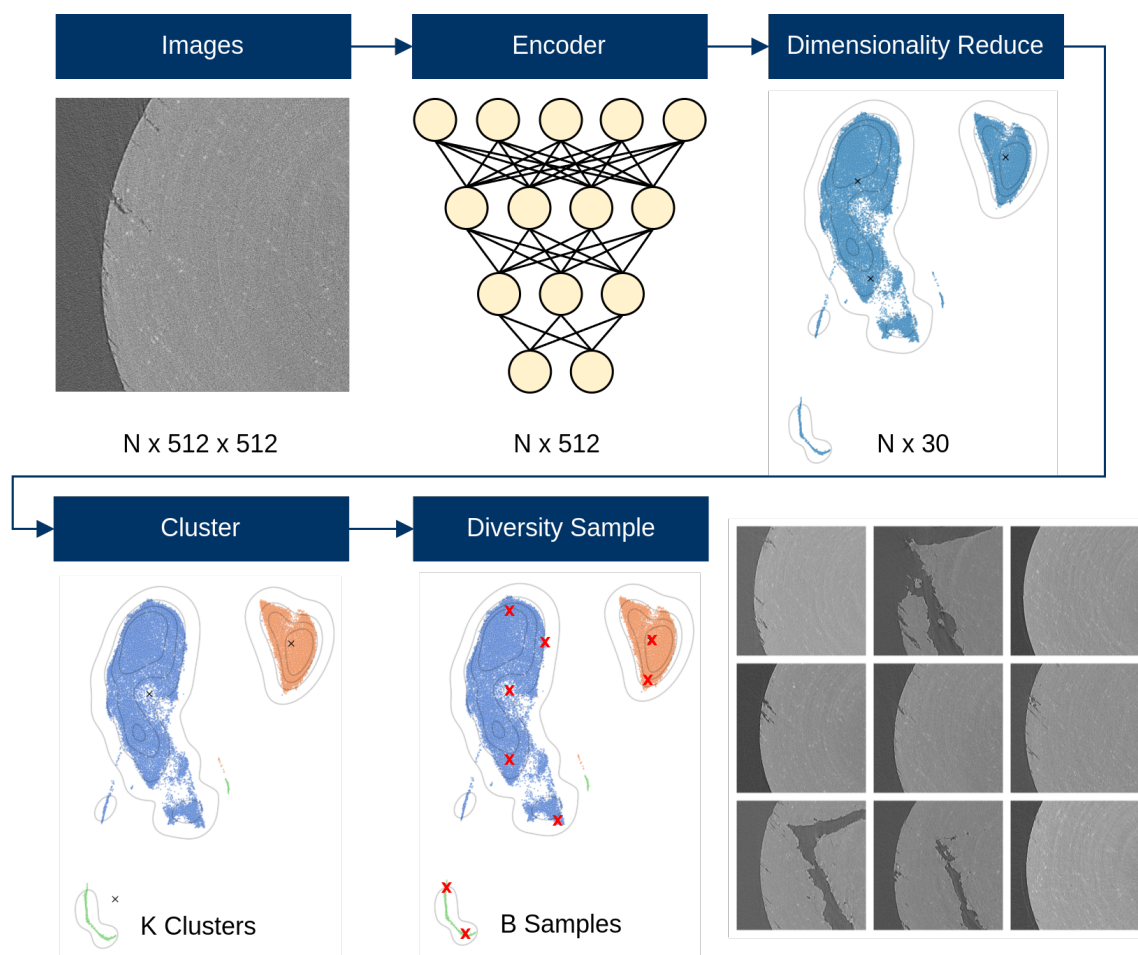


Figure 3.4. High-level view of the diversity sampling pipeline. Images are reduced using a pre-trained encoder and dimensionality reduction. Subsequent embeddings are then clustered and diversity sampled for annotation.

Image preprocessing follows model-specific requirements, with grayscale microscopy images converted to RGB and normalized according to each model's ImageNet training distribution<sup>51</sup>. Detailed preprocessing information can be found in Table 6.1 in Appendix 6.2.

### 3.3.2 Dimensionality Reduction

The high-dimensional embeddings from CLIP, VGG-19, and ResNet50, ranging from 512-4096 dimensions, pose challenges for analysis and visualization. To mitigate the curse of dimensionality while preserving visual representations, we compare two complementary dimensionality reduction approaches.

In our first approach, we apply standard scaling to the raw embeddings followed by Principal Component Analysis (PCA), retaining components that maintain 95% of the variance to capture the most significant axes of variation<sup>27</sup>. As an alternative approach, we L2 normalize the raw embeddings before applying Uniform Manifold Approximation and Projection (UMAP) to generate a 30-dimensional representation, using cosine distance<sup>29</sup>. Detailed parameters for both can be found in Tables 6.2 and 6.3 in Appendix 6.2.

These techniques offer different perspectives for downstream analysis: PCA provides an interpretable linear projection optimizing variance retention, while UMAP captures non-linear relationships in the data through local and global structure preservation. By applying both methods to the set of tiles at each of the nine tile locations, we ensure the preserved information predominantly captures visual features of interest rather than location in the original image the tile was sampled from.

### 3.3.3 Structural Pattern Analysis

After dimensionality reduction, we employ multiple clustering strategies to identify distinct structural patterns in our image embeddings. Three clustering algorithms are implemented: K-means for basic partition-based clustering, DBSCAN for density-based clustering, and HDBSCAN for hierarchical density-based clustering that can adapt to varying density regions<sup>30–32</sup>. RAPIDS cuML library was used for accelerated GPU processing, enabling efficient large-scale clustering<sup>73</sup>.

Each algorithm is configured to capture different perspectives on structural relationships. K-means is set with  $k = 3$  to provide a baseline partitioning of the structural space. DBSCAN set with  $\epsilon = 0.1$  and requires a minimum of 5 samples to

form a core point. HDBSCAN is set with  $\epsilon = 0.1$ , a minimum of 5 samples for core point consideration, and a minimum cluster size of 50 samples. All parameters for each approach can be found in Tables 6.4, 6.5, and 6.6 in Appendix 6.2.

To evaluate inter- and intra-clustering performance, we compute two metrics across these configurations. The Silhouette score measures point similarity within and between clusters, while the Davies-Bouldin index assesses the average similarity ratio of each cluster with its most similar cluster<sup>74,75</sup>. Practical metrics including cluster count, number of noise points, and computational runtime were also recorded. This comprehensive approach enables us to compare the effectiveness of different clustering methods and identify the most informative combinations of pre-trained model, dimensionality reduction, and clustering techniques.

### 3.3.4 Baseline Sampling Strategies

We implement three established baseline sampling methods to benchmark against our proposed DIDS approach:

**Random Sampling:** As a control baseline, we implement random sampling across the dataset. While simple, random sampling provides a reference point for evaluating more sophisticated approaches. It provides an approximate uniform distribution across both spatial dimensions (slice locations within the volumes) and temporal progression (experimental timesteps), serving as a basic spatiotemporal stratification baseline.

**TypiClust:** This approach combines representation learning with clustering-based diversity sampling<sup>46</sup>. TypiClust leverages the pre-trained embeddings to establish a semantic feature space where the data is partitioned into clusters using K-means. It selects the most typical examples from the B largest uncovered clusters, where B is the allocated budget. Typicality is measured by the density of neighboring points in the embedding space. This ensures both diversity across the feature space and representativeness within each cluster.

**ProbCover:** This method implements a graph-based diversity sampling strategy that operates in the embedding space<sup>47</sup>. ProbCover constructs a directed graph where vertices represent samples and edges connect points within a specified

distance threshold  $\delta$ . The algorithm iteratively selects samples with the highest out-degree (most connections to uncovered regions) and updates the graph by removing edges to newly covered regions. This greedy optimization approach ensures maximum coverage of the feature space while maintaining computational efficiency through sparse graph representations.

Both TypiClust and ProbCover utilize the same pre-trained model embeddings described in Section 3.3.1, ensuring a fair comparison with our DIDS approach. TypiClust and ProbCover are both run with a budget size of 10 to sample 10 images from each tile location for a total dataset of 90 sampled images. ProbCover has  $\delta$  set to a value from 0.25-0.35 depending on the pre-trained model embeddings used, where the  $\delta$  was selected to maximize coverage. All parameters used for these two sampling strategies can be found in Tables 6.7 and 6.8 in Appendix 6.2.

### 3.3.5 Domain-Informed Diversity Sampling (DIDS)

The evolution of stress corrosion cracking in AlMg presents a complex, multiscale phenomenon where both spatial microstructural features and temporal progression play crucial roles. Basic spatiotemporal stratification becomes ineffective when critical events occur in spatially localized regions or follow non-uniform temporal progression, potentially missing rare but significant phenomena such as crack initiation points that may be confined to specific regions of the volume and emerge at irregular intervals during degradation. Alternatively, modern diversity sampling methods such as TypiClust and ProbCover often rely solely on visual embeddings, which may not capture critical domain-specific aspects of material degradation. To address this limitation, we propose Domain-Informed Diversity Sampling (DIDS), a dual-objective optimization strategy that integrates visual feature diversity with physical measurements of material degradation.

DIDS operates through a selection process that combines embedding-based distances with domain-specific metrics - in our case, displacement measurements that directly indicate degradation progression. The method first normalizes both visual embeddings and displacement values to ensure comparable scaling.

---

**Algorithm 1** Domain-Informed Diversity Sampling (DIDS)
 

---

**Require:**

- 1:  $X = \{x_1, \dots, x_n\}$  : Feature embeddings for each point
- 2:  $D = \{d_1, \dots, d_n\}$  : Displacement measurements
- 3:  $\{C_1, \dots, C_K\}$  : Set of clusters from previous clustering
- 4:  $N$  : Total number of samples to select
- 5:  $m$  : Distance metric type (cosine or euclidean)
- 6:  $\gamma$  : Weight between embedding and displacement distances

**Ensure:** Selected sample indices  $S$ 

```

7:
8:  $\hat{d}_i \leftarrow \frac{d_i - \min(D)}{\max(D) - \min(D)}$  for all  $i$ 
9:
10:  $n_k \leftarrow \max(\min(N \cdot \frac{|C_k|}{\sum_j |C_j|}, n_{\max}), n_{\min})$  for all  $k$ 
11: ensuring:  $\sum_k n_k = N$ 
12: for all clusters  $C_k$  do
13:
14:   if  $m = \text{cosine}$  then
15:      $D_{\text{emb}}(i, j) \leftarrow 1 - \frac{x_i \cdot x_j}{\|x_i\| \|x_j\|}$ 
16:      $\rho_i \leftarrow \frac{1}{|C_k|} \sum_{j \in C_k} \frac{x_i \cdot x_j}{\|x_i\| \|x_j\|}$ 
17:   else if  $m = \text{euclidean}$  then
18:      $D_{\text{emb}}(i, j) \leftarrow \frac{\|x_i - x_j\|}{\max_{p,q} \|x_p - x_q\|}$ 
19:      $\rho_i \leftarrow 1 - \frac{1}{|C_k|} \sum_{j \in C_k} \frac{\|x_i - x_j\|}{\max_{p,q} \|x_p - x_q\|}$ 
20:   end if
21:
22:    $S_k \leftarrow \{\arg \max_{i \in C_k} \rho_i\}$ 
23:   while  $|S_k| < n_k$  do
24:     for all candidates  $i \in C_k \setminus S_k$  do
25:
26:        $D_{\text{disp}}(i, j) \leftarrow |\hat{d}_i - \hat{d}_j|$ 
27:
28:        $D(i, j) \leftarrow \gamma \cdot D_{\text{emb}}(i, j) + (1 - \gamma) \cdot D_{\text{disp}}(i, j)$ 
29:
30:        $D_{\min}(i) \leftarrow \min_{j \in S_k} D(i, j)$ 
31:     end for
32:
33:      $i^* \leftarrow \arg \max_{i \in C_k \setminus S_k} D_{\min}(i)$ 
34:      $S_k \leftarrow S_k \cup \{i^*\}$ 
35:   end while
36: end for
37: return  $S = \bigcup_k S_k$ 

```

---

$$\hat{d}_i = \frac{d_i - \min(D)}{\max(D) - \min(D)}$$

where  $d_i$  is the specific displacement value corresponding to image  $i$  in the dataset and  $D$  is the set of all displacement values.

Within each assigned cluster of visual states, sampling begins from points of maximum local density, computed using Euclidean or cosine similarity between feature embeddings.

For cosine similarity:

$$\rho_i = \frac{1}{|C_k|} \sum_{j \in C_k} \cos(x_i, x_j)$$

For Euclidean distance:

$$\rho_i = 1 - \frac{1}{|C_k|} \sum_{j \in C_k} \frac{\|x_i - x_j\|}{\max_{p,q} \|x_p - x_q\|}$$

where  $\rho_i$  is the local density score for point  $i$ ,  $|C_k|$  is the size (number of points) in cluster  $k$ , and  $x_i, x_j$  are feature embeddings for points  $i$  and  $j$ . In the Euclidean version,  $\|x_i - x_j\|$  is the Euclidean distance between embeddings of points  $i$  and  $j$  and  $\max_{p,q} \|x_p - x_q\|$  is the maximum pairwise Euclidean distance between any two points  $p$  and  $q$  in the cluster (used for normalization).

The core selection criterion combines two distance metrics: a feature-based distance ( $d_{\text{emb}}$ ) capturing visual diversity and a displacement-based distance ( $d_{\text{disp}}$ ) measuring degradation progression:

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

where  $\gamma$  controls the balance between visual and degradation diversity. It can be thought of as a "grounding" factor for how much visual embeddings are "grounded" by a physical measurement. When  $\gamma = 1.0$  only the visual embedding space is leveraged, and as  $\gamma$  decreases, sampling is grounded more with domain knowledge. This combined metric drives an iterative selection process that maximizes the



minimum distance to previously selected points, ensuring both local and global diversity in the sample set.

To maintain representative coverage across different structural regimes, we implement an adaptive cluster-based allocation strategy. Each cluster receives a sample allocation proportional to its size, bounded by minimum and maximum thresholds to prevent both under- and over-representation of any particular regime. The complete selection process is detailed in Algorithm 1.

We evaluate three weighting configurations ( $\gamma = 1.0, 0.85, 0.5$ ) to assess different balances between visual and physical diversity. The  $\gamma = 1.0$  case focuses purely on visual diversity, while lower values increasingly incorporate temporal evolution information, allowing us to explore how this balance affects the quality and representativeness of selected samples.

A key advantage of DIDS is its flexibility - while we use displacement as our domain metric in this study, the framework can accommodate any relevant physical measurement (e.g. temperature, strain, load) or composite of domain metrics. This adaptability makes DIDS particularly valuable for scientific applications where multiple physical indicators may be relevant to the relevant process.

### 3.3.6 Quality Assessment Framework

To evaluate our sampling strategy, we develop a quality assessment framework that combines three complementary metrics: latent space coverage, visual perceptual diversity, and displacement range coverage. Each metric is normalized and weighted in our final diversity score for a configurable evaluation across different aspects of the sampling.

Latent space coverage measures how well the sampled set spans the underlying embedding space, ensuring we capture the full range of possible visual variations. A minimum spanning tree (MST) is a graph-theoretical concept that connects all points in a set with the minimum total edge weight, providing an efficient way to measure the spread and connectivity of points in high-dimensional spaces<sup>76</sup>. For latent space coverage, we compute two complementary minimum spanning tree (MST) metrics in the raw embedding space: the total MST length and the mean

edge length. The MST provides a measure of global connectivity by identifying the shortest possible path connecting all points, with larger values indicating better coverage of the feature space.

We compute these metrics using both Euclidean and cosine distances, then combine them into a normalized latent spread score:

$$\text{LatentSpread} = \frac{1}{4}(\hat{m}_{\text{euc}} + \hat{m}_{\text{cos}} + \hat{L}_{\text{euc}} + \hat{L}_{\text{cos}})$$

where  $\hat{m}$  represents normalized mean MST edge lengths and  $\hat{L}$  represents normalized total MST lengths.

Perceptual diversity quantifies how visually distinct the sampled images appear to the human eye, ensuring we capture meaningful variations rather than just subtle numerical differences. Visual diversity is assessed using Learned Perceptual Image Patch Similarity (LPIPS), which leverages deep neural network features to approximate human perceptual differences<sup>44</sup>. This metric provides a value between 0 and 1, where 0 indicates perceptually identical images and 1 indicates maximum perceptual difference. Visual similarity for a selected subset of data is measured as the mean LPIPS distance between all pairwise comparisons. We use the TorchMetrics implementation of LPIPS with AlexNet as the backbone network<sup>45,77</sup>.

In the domain of material degradation, displacement fields characterize how surface points move during deformation, making them a crucial metric for capturing the progression of material aging and damage. Degradation progression coverage is measured through the normalized displacement range:

$$\text{DisplacementRange} = \frac{R_{\text{disp}} - \min(R_{\text{disp}})}{\max(R_{\text{disp}}) - \min(R_{\text{disp}})}$$

where  $R_{\text{disp}}$  represents the range of displacement values in the selected samples. These normalized metrics combine into a unified diversity score:

$$D_{\text{total}} = w_l D_l + w_v D_v + w_d D_d$$

where:

- $D_l$  is the LatentSpread metric

- $D_v$  is the VisualSimilarity metric
- $D_d$  is the DisplacementRange metric

and  $w_l + w_v + w_d = 1$ . In our case we use  $w_l = 0.2$ ,  $w_v = 0.4$ , and  $w_d = 0.4$

This weighting scheme ensures dynamic consideration of latent space coverage, perceptual diversity, and physical degradation progression in evaluating sampling effectiveness. Normalization is performed within each encoder group for latent space metrics to account for different embedding dimensionalities, while LPIPS and displacement metrics are normalized globally across all methods.

### 3.4 Feature Extraction and Characterization Framework

In AlMg alloys, two microstructural features play critical roles in stress corrosion cracking: fracture networks and material inclusions. Fracture networks manifest as branching crack patterns that directly indicate material degradation and provide key insights into failure mechanisms. Material inclusions, primarily consisting of intermetallic Al-Fe-Mn particles and Mg-rich precipitates (typically 2-10  $\mu\text{m}$  in diameter), serve as both crack initiation sites and pathway guides for crack propagation. These inclusions significantly influence local stress distributions and chemical reactivity, making their distribution and characteristics crucial for understanding degradation progression. An example of each feature is depicted in Figure 3.5.

To enable detailed analysis of microstructural evolution in XCT volumes, we develop an integrated feature extraction framework that combines automated annotation generation with advanced segmentation techniques. A high-level view of the pipeline is depicted in Figure 3.6. Our framework addresses two key challenges: minimizing manual annotation requirements while maintaining high accuracy for critical feature detection, and ensuring consistent feature tracking across large temporal sequences. The framework consists of three main components: 1) a hybrid annotation pipeline that adapts its strategy based on feature characteristics, 2) deep learning models optimized for robust feature segmentation, and 3) automated quality assessment methods to validate detection.

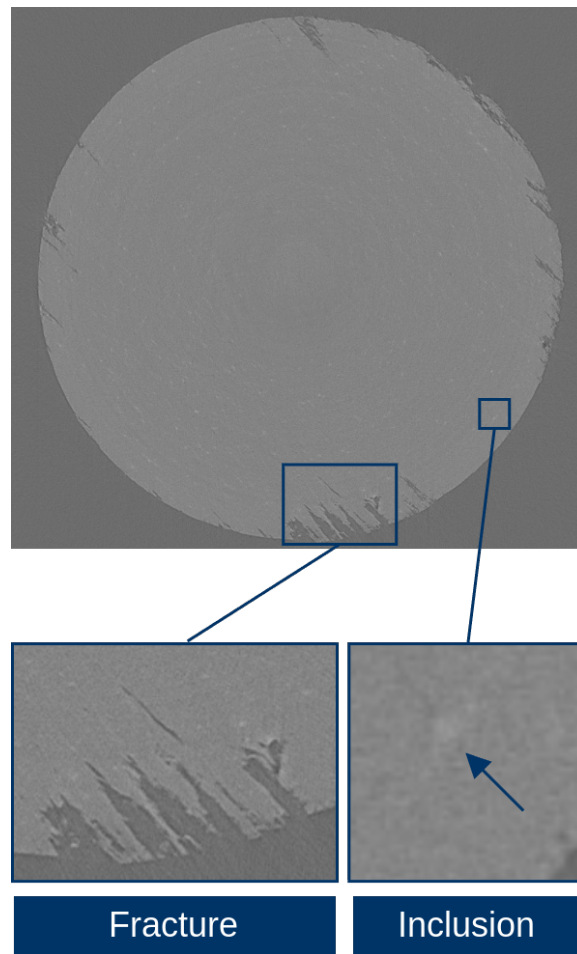


Figure 3.5. Sample XCT slice and examples of fracture and inclusion features.

### 3.4.1 Fracture Network Annotation

Using the best performing DIDS approach, we diversity sample a dataset of 90 images (10 from each tile location). Five additional images are hand selected as a holdout test set to capture a range of timesteps and fracture events. Distinct annotation strategies optimized for the unique characteristics of each microstructural feature type are then employed.

Fracture networks present as sparse (0-20 per image) but structurally complex features with high contrast properties. Given their critical importance in degradation analysis and relatively low frequency, we employ manual pixel-wise labeling

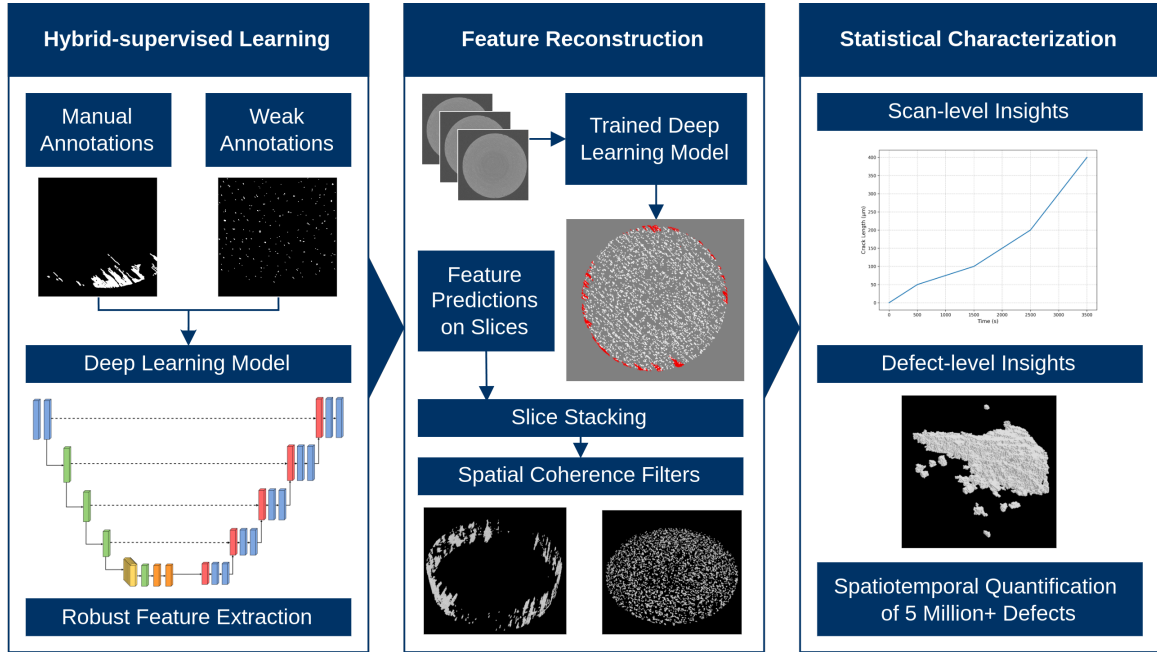


Figure 3.6. Spatiotemporal feature extraction and characterization pipeline which includes hybrid annotation generation, deep learning-based segmentation, feature reconstruction, and defect characterization.

through LabelStudio<sup>78</sup>. This approach, while time-intensive, ensures high-quality ground truth annotations for these complex structures.

### 3.4.2 Weakly Supervised Inclusion Annotation Pipeline

Material inclusions pose fundamentally different annotation challenges, characterized by high density (hundreds per image), small spatial scale (12 pixels in area), and low contrast. The high feature count and sub-visible characteristics makes manual annotation impractical, necessitating an automated processing pipeline. Our solution employs a sequential weakly supervised workflow:

- (1) Mechanical noise suppression via Gaussian kernel convolution
- (2) Local contrast enhancement through disk-based morphological operations
- (3) Adaptive thresholding at percentile of local intensity distributions

Table 3.4. Inclusion segmentation image processing pipeline parameters

Processing Step	Parameter	Value
Gaussian smoothing	sigma	1.0
Contrast enhancement	neighborhood disk size	30px
Adaptive threshold	percentile	0.98
Canny edge detection	sigma	2.0
Ellipse masking	dilation factor	0.99
Object filtering	minimum size	6
	maximum size	70
	max eccentricity	0.95

- (4) Material region identification using least squares ellipse fitting on Canny edge detection outputs
- (5) Feature refinement through geometric constraints on size and shape

Table 3.4 provides a comprehensive overview of the parameters for each image processing step. Parameters were empirically determined through iterative testing on our diversity-sampled dataset, ensuring robust performance across the full range of microstructural variations and deformation states. Figure 3.7 depicts the progressive transformation of the raw image through each stage, culminating in the final segmentation mask.

This automated approach provides a robust foundation for weakly supervised learning while dramatically reducing annotation time compared to manual methods. While it may introduce some segmentation imperfections, these are acceptable given the statistical nature of inclusion analysis and the fact manual annotation would likely be more biased and improbable due to these features being sub-visible for the human eye.

### 3.4.3 Model Selection and Implementation

Several state-of-the-art segmentation architectures were evaluated to identify optimal approaches for XCT microstructural feature extraction. We investigated three architectural families: U-Net for its symmetric encoder-decoder architecture with skip connections to preserved fine-grained details and global context, UperNet for

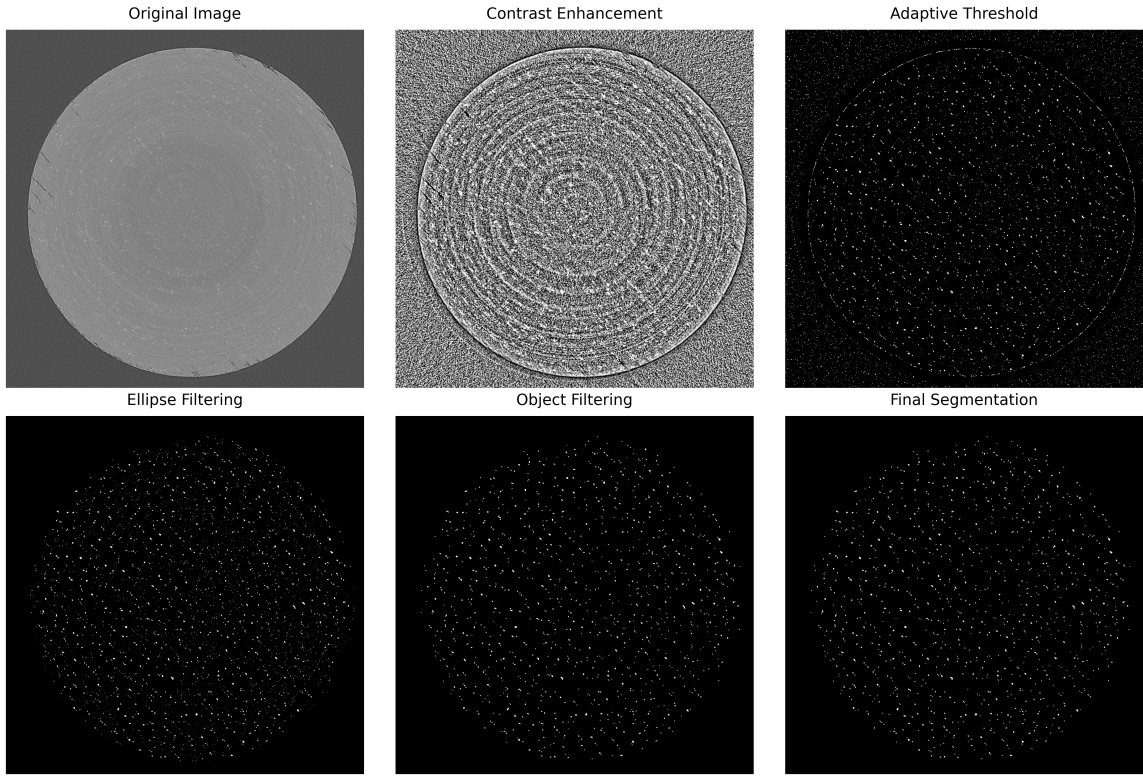


Figure 3.7. Transformations for each step inside the image processing inclusion segmentation pipeline.

its hierarchical feature pyramid processing, and SegFormer for its efficient transformer-based design<sup>36–38</sup>. For both U-Net and UperNet architectures, we evaluated three encoder backbones: ResNet50 for its gradient flow optimization through residual learning, Xception for computational efficiency via depthwise separable convolutions, and SE-ResNeXt101 for enhanced feature recalibration through channel-wise attention<sup>40–42</sup>. For SegFormer, we employed two Mix Transformer (MiT) variants: B0 for lightweight encoding and B4 for high-capacity feature extraction<sup>38</sup>. Each of the architecture encoder combinations was tested with initialization from ImageNet pre-trained weights and from scratch<sup>51</sup>. As a baseline, a traditional U-Net with no specialized backbone was implemented as well. Models were implemented using base PyTorch and Iakubovskii’s Segmentation Models library<sup>79,80</sup>. Figure 3.9 depicts the architecture of U-Net with Xception as the backbone (the best performing model).



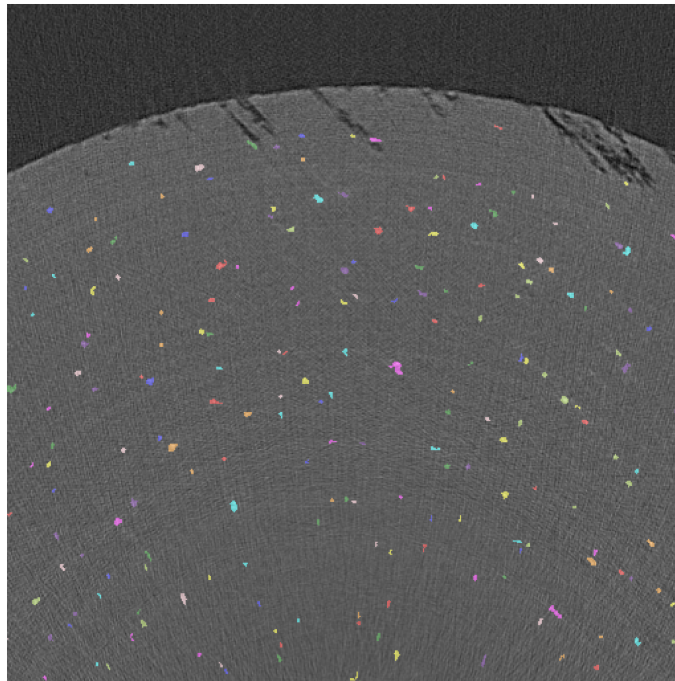


Figure 3.8. Inclusion pseudo-label example using weak supervision.

Table 3.5. Segmentation model architectures and training parameters

Model	Encoder	Trainable Parameters
U-Net	None	31,043,651
U-Net	ResNet50	32,521,395
U-Net	SE-ResNeXt101	55,919,779
U-Net	Xception	28,769,691
UPerNet	ResNet50	29,971,587
UPerNet	SE-ResNeXt101	53,369,971
UPerNet	Xception	27,096,427
SegFormer	MIT-B1	13,678,019
SegFormer	MIT-B4	61,369,283

#### 3.4.4 Training Strategy

Due to our constrained dataset size, we implement a careful training protocol to ensure model robustness. The dataset is partitioned into training (90 images), validation (5 images), and holdout test (5 images) sets. Input preprocessing follows two



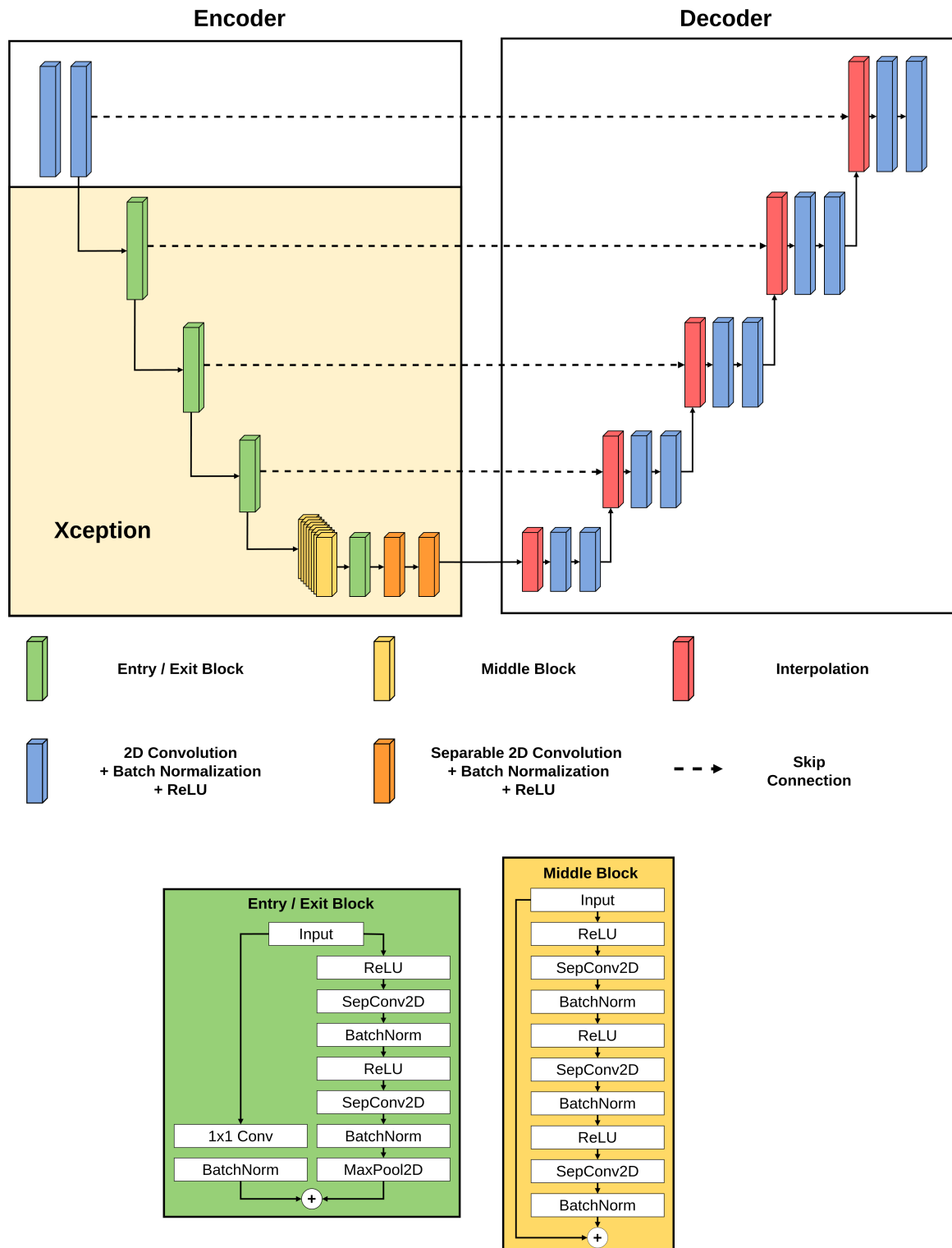


Figure 3.9. Architecture of U-Net with Xception as the encoder backbone.

paths: standard 0-1 min-max normalization for models trained from scratch, and ImageNet statistics adjustment (mean subtraction and standard deviation scaling) for pre-trained models<sup>51</sup>. We employ a targeted data augmentation strategy to improve model generalization, including random horizontal and vertical flips with probability 0.5 and contrast and brightness adjustments ( $\pm 0.1$ ) to enhance robustness to imaging variations while preserving feature characteristics.

Model training uses Adam optimization with an initial learning rate of  $1 \times 10^{-3}$  and cosine annealing schedule to a minimum of  $1 \times 10^{-6}$  over 150 epochs with a batch size of 16<sup>81,82</sup>. We employ the Dice loss function to address class imbalance inherent in sparse feature detection. All models process standardized 512×512 pixel inputs and are evaluated using a comprehensive metric suite including pixel-wise accuracy, precision, recall, F1-score, and mean Intersection over Union (mIoU). In-depth definitions for each metric are contained in Appendix 6.1. Model selection is based on best mIoU performance on the validation set, with final performance assessed on the independent holdout test set to ensure unbiased evaluation.

### 3.4.5 Feature Prediction and 3D Volume Reconstruction

Following model evaluation, we deployed U-Net with an Xception backbone for volumetric reconstruction using a tiled prediction strategy. Prediction was performed through a sliding window approach with overlapping tiles, where we implemented cosine-weighted blending to minimize boundary artifacts:

$$w(x) = \cos(\pi x / 2L)$$

where  $x$  represents the distance from tile center and  $L$  denotes the overlap width.

Quality verification forms a critical component of our reconstruction pipeline. We employ a two-stage verification process: visual inspection of randomly sampled 2D slices against ground truth annotations and consistency checks across adjacent slices to ensure 3D continuity. This multi-level verification ensures reliable feature extraction while maintaining computational efficiency.

For crack networks, we focus on physically meaningful metrics including total crack volume, approximate surface area, major branch lengths, and basic connectivity through intersection points. These properties provide direct insight into degradation progression while avoiding computational complexity. Similarly, inclusion analysis centers on count and volume distributions, basic shape metrics such as aspect ratio, and nearest distances to crack paths.

This targeted approach to feature characterization enables practical analysis of degradation mechanisms while maintaining interpretability. The verification steps combined with focused metrics provide a robust framework for analyzing stress corrosion cracking progression in large-scale XCT volumes.

## 4 Results

### 4.1 Diversity Sampling Framework

#### 4.1.1 Embedding Characteristics

Table 4.1 denotes the number of PCA components required to capture 95% of the variance from the pre-trained model embeddings. The CLIP embeddings demonstrate remarkable dimensionality efficiency, requiring only 92 components to capture 95% of the variance. In contrast, ResNet50 requires substantially more components (1,193) to achieve the same variance threshold, indicating its embeddings contain more distributed information across dimensions. VGG-19 shows intermediate efficiency with 178 components, performing better than ResNet50 but not achieving the same compact representation as CLIP. The significant difference in required components between these architectures has important implications for our diversity sampling pipeline. CLIP’s efficient representation suggests it may be more computationally tractable for large-scale XCT analysis, while still capturing relevant features. The higher dimensional representations from ResNet50 may capture more fine-grained details but at the cost of increased computational overhead during the diversity sampling process. More detailed information on variance stratified at the tile-location resolution can be found in Figures 6.2-6.4 in Appendix 6.3.

Figure 4.1 demonstrates a qualitative visualization of how using different pre-trained models impacts the embedding representation of the same set of images. The figure depicts how CLIP divides samples into more distinct clusters while

Table 4.1. Number of components required to capture 95% variance in embeddings

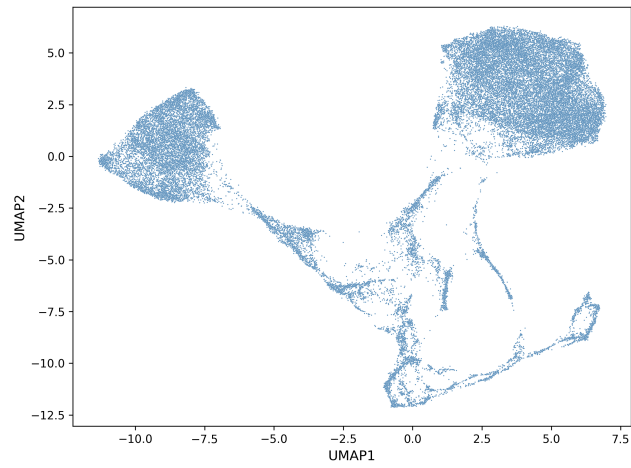
Pre-trained Model	Original Dimensions	Number of Components
CLIP	512	92
ResNet50	2048	1193
VGG-19	4096	178

ResNet50 and VGG-19 provide a more dispersed and smooth transition over the embedding space.

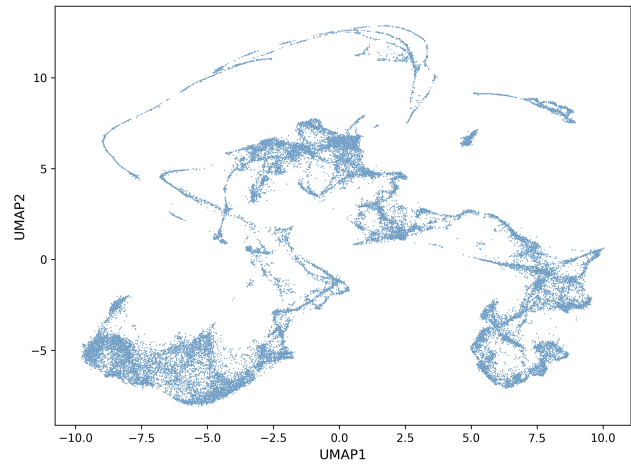
#### 4.1.2 Clustering Performance Evaluation

The clustering evaluation across different pre-trained model embeddings is shown in Table 4.2. We evaluated clustering performance across our two dimensionality reduction approaches: PCA retaining 95% variance and the 30-component UMAP. The CLIP-UMAP embeddings clustered using K-means achieved the highest overall Silhouette score of 0.747 and the lowest Davies-Bouldin index of 0.800, indicating well-separated and balanced clusters. Across all models, K-means with UMAP-reduced embeddings consistently outperformed HDBSCAN, with ResNet50 and VGG-19 achieving Silhouette scores of 0.667 and 0.639 respectively. DBSCAN consistently failed to converge and is not included in the analysis as a result. Figures 6.5-6.8 in Appendix 6.3 demonstrate clustering metrics stratified by pre-trained model and reduction technique.

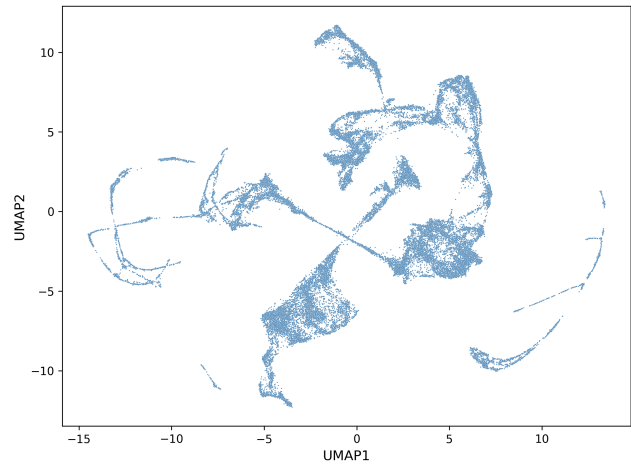
Figure 4.2 illustrates the generally inconsistent performance of clustering approaches, however. While some instances show clear delineation of distinct regions with evident boundaries between visual domains, there are also cases of sub-optimal clustering where algorithms either over-segment the feature space or fail to identify coherent structural patterns. These variations highlight a fundamental challenge: effectively parameterizing clustering across different pre-trained models, embedding spaces, and distance metrics remains a non-trivial problem.



(a) CLIP embeddings.

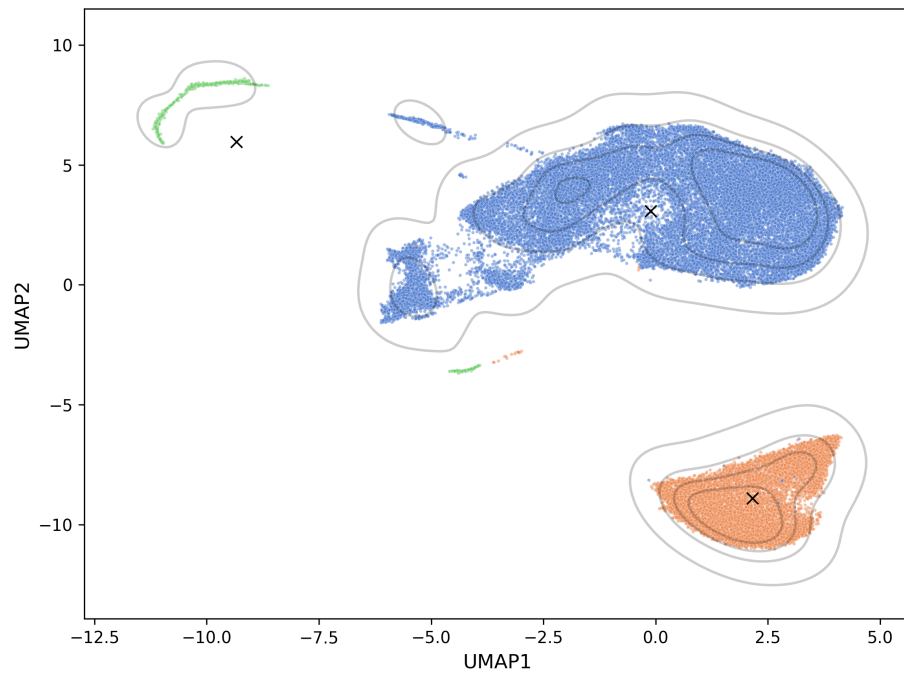


(b) ResNet50 embeddings.

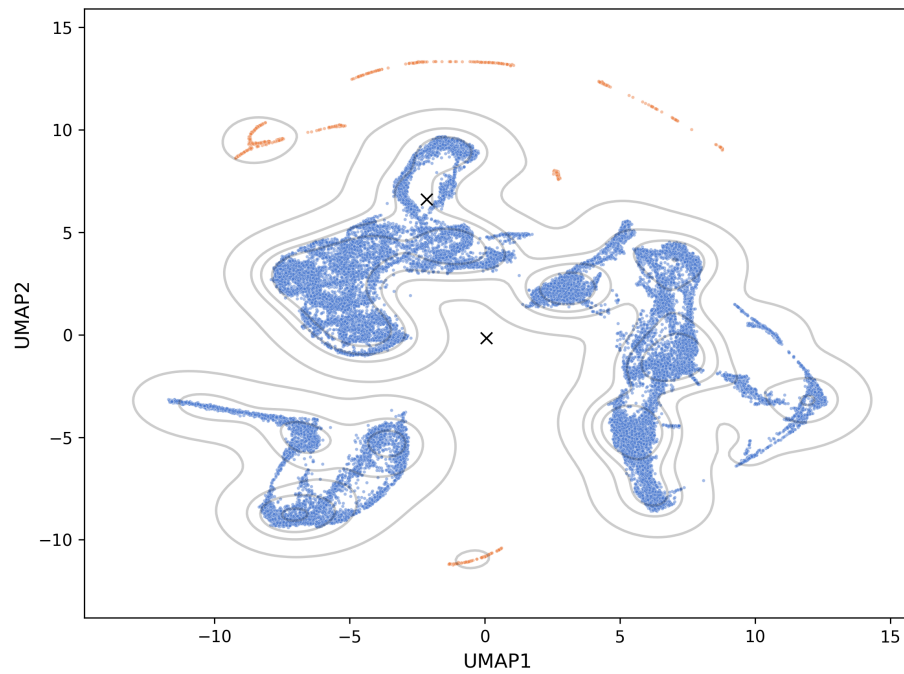


(c) VGG-19 embeddings.

Figure 4.1. 2D UMAP embeddings of bottom center image tiles across different pre-trained models.



(a) Cluster results of UMAP CLIP embeddings of upper left tiles using HDBSCAN.



(b) Cluster results of VGG-19 CLIP embeddings of upper left tiles using HDBSCAN.

Figure 4.2. Examples of strong and poor clustering convergence for embeddings.

Table 4.2. Clustering performance metrics across pre-trained model and embedding spaces

Pre-trained Model	Reduction	Method	Silhouette	Davies-Bouldin
CLIP	UMAP	HDBSCAN	0.441	0.847
		K-means	<b>0.747</b>	<b>0.800</b>
	PCA	HDBSCAN	0.218	1.591
		K-means	0.168	2.024
ResNet50	UMAP	HDBSCAN	0.239	1.172
		K-means	0.667	0.926
	PCA	HDBSCAN	0.279	1.771
		K-means	0.197	2.767
VGG-19	UMAP	HDBSCAN	0.047	1.123
		K-means	0.639	0.964
	PCA	HDBSCAN	0.265	1.478
		K-means	0.166	1.957

#### 4.1.3 Diversity Sampling Evaluation

Figure 4.3 and Table 4.3 present a comprehensive comparison of our Domain-Informed Diversity Sampling (DIDS) approach against established baselines: random sampling, TypiClust, and ProbCover. DIDS demonstrates superior performance across multiple diversity metrics, with particularly notable improvements in embedding coverage and displacement range representation.

The unnormalized diversity metrics (Figure 4.3) reveal that while ProbCover achieves comparable LPIPS distances to DIDS, our method shows marked improvements in embedding space coverage and displacement range representation. Specifically, DIDS achieves an average MST edge cosine distance of approximately 0.055 compared to 0.038 for ProbCover, indicating better coverage of the latent space. The displacement metric shows even more pronounced improvements, with DIDS maintaining an average value of 0.17 compared to 0.15 for ProbCover and 0.12 for TypiClust. The normalized metrics in Figure 4.4 show even more pronounced differences between approaches.



When examining normalized metrics (Table 4.3), the optimal DIDS configuration (CLIP-PCA embeddings with KMeans clustering at  $\gamma = 0.85$ ) achieves the highest overall diversity score of 0.907, substantially outperforming the best baseline scores from ProbCover (0.491), Random sampling (0.336), and TypiClust (0.299). This superior performance stems from DIDS's balanced optimization across all three diversity measures, with normalized scores of 0.955 for latent spread, 0.867 for LPIPS, and 0.923 for displacement range.

The consistent performance advantage of DIDS over baseline methods demonstrates the effectiveness of incorporating domain-specific displacement information alongside traditional embedding-based diversity metrics. This balanced approach ensures comprehensive coverage of both the visual feature space and the underlying physical deformation characteristics of the material system.

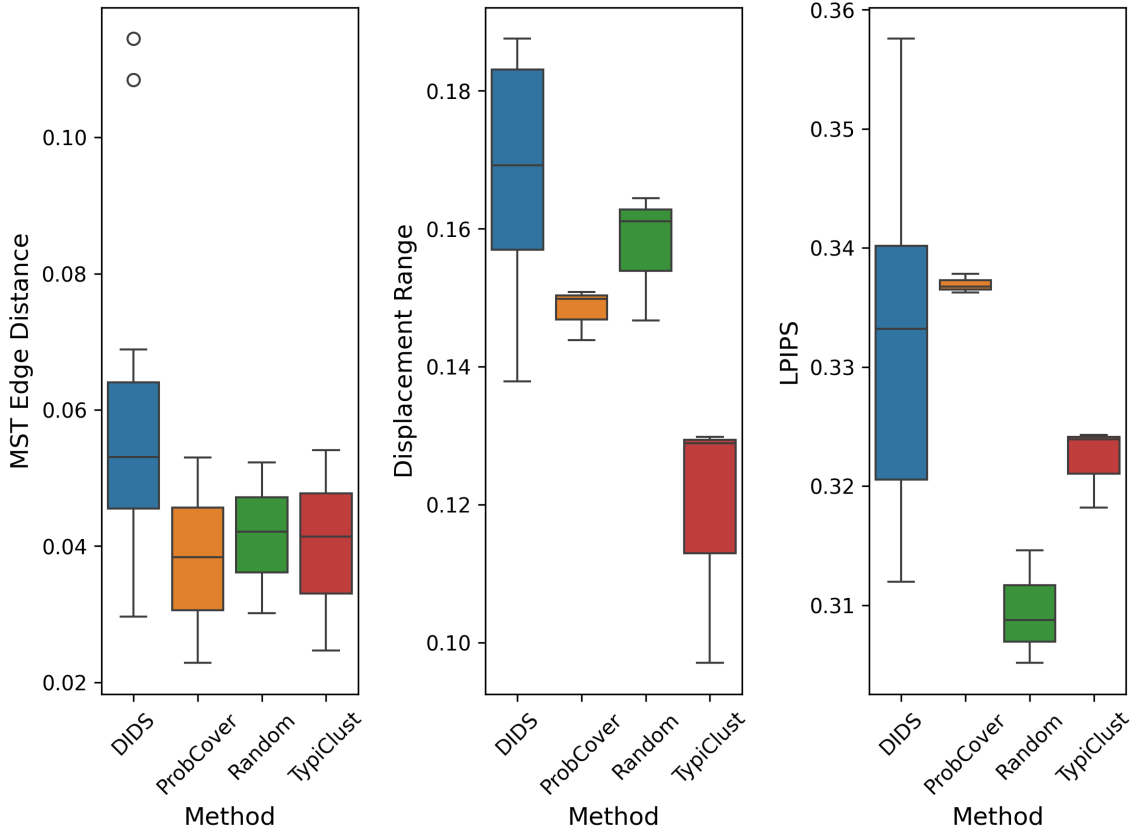


Figure 4.3. Distribution of unnormalized diversity metrics for different sampling methods.

Table 4.3. Normalized diversity metrics and overall diversity score across different sampling strategies

Encoder	Method	Latent Spread	LPIPS	Displacement	Diversity Score
CLIP	DIDS (PCA, KMeans, 0.85)	0.955	0.867	0.923	<b>0.907</b>
VGG19	DIDS (PCA, KMeans, 0.85)	0.914	0.820	0.853	0.852
VGG19	DIDS (PCA, KMeans, 0.5)	0.727	0.758	<b>1.000</b>	0.849
ResNet50	DIDS (UMAP, HDBSCAN, 0.85)	0.994	0.595	<b>1.000</b>	0.837
CLIP	DIDS (PCA, KMeans, 0.5)	0.732	0.626	<b>1.000</b>	0.797
VGG19	DIDS (PCA, KMeans, 1.0)	<b>1.000</b>	<b>1.000</b>	0.451	0.781
ResNet50	DIDS (PCA, KMeans, 0.5)	0.980	0.498	0.927	0.766
CLIP	DIDS (PCA, KMeans, 1.0)	<b>1.000</b>	0.675	0.737	0.765
ResNet50	DIDS (UMAP, HDBSCAN, 0.5)	0.811	0.503	<b>1.000</b>	0.764
ResNet50	DIDS (PCA, KMeans, 0.85)	0.675	0.542	0.856	0.694
VGG19	DIDS (UMAP, HDBSCAN, 0.5)	0.248	0.554	<b>1.000</b>	0.671
VGG19	DIDS (UMAP, HDBSCAN, 0.85)	0.263	0.559	0.951	0.656
ResNet50	DIDS (PCA, KMeans, 1.0)	0.790	0.708	0.538	0.656
VGG19	DIDS (UMAP, KMeans, 0.85)	0.198	0.534	0.951	0.634
VGG19	DIDS (UMAP, HDBSCAN, 1.0)	0.267	0.655	0.741	0.612
CLIP	DIDS (UMAP, HDBSCAN, 0.5)	0.241	0.344	<b>1.000</b>	0.586
ResNet50	DIDS (UMAP, KMeans, 0.85)	0.488	0.199	<b>1.000</b>	0.577
CLIP	DIDS (UMAP, HDBSCAN, 0.85)	0.290	0.297	<b>1.000</b>	0.577
VGG19	DIDS (UMAP, KMeans, 0.5)	0.146	0.365	0.976	0.565
ResNet50	DIDS (UMAP, HDBSCAN, 1.0)	0.597	0.348	0.664	0.524
CLIP	DIDS (UMAP, KMeans, 0.85)	0.211	0.172	<b>1.000</b>	0.511
VGG19	DIDS (UMAP, KMeans, 1.0)	0.203	0.510	0.654	0.506
CLIP	DIDS (UMAP, KMeans, 0.5)	0.184	0.134	<b>1.000</b>	0.491
ResNet50	DIDS (UMAP, KMeans, 1.0)	0.458	0.300	0.683	0.484
ResNet50	DIDS (UMAP, KMeans, 0.5)	0.213	0.000	<b>1.000</b>	0.443
CLIP	DIDS (UMAP, HDBSCAN, 1.0)	0.226	0.171	0.625	0.364
CLIP	DIDS (UMAP, KMeans, 1.0)	0.200	0.134	0.661	0.358
VGG19	ProbCover	<b>0.056</b>	0.605	<b>0.594</b>	<b>0.491</b>
CLIP	ProbCover	0.000	0.595	0.517	0.445
ResNet50	ProbCover	0.000	<b>0.625</b>	0.584	0.483
ResNet50	Random	<b>0.219</b>	0.005	0.707	<b>0.329</b>
VGG19	Random	0.000	0.073	<b>0.745</b>	0.327
CLIP	Random	0.215	<b>0.185</b>	0.548	0.336
VGG19	TypiClust	0.030	<b>0.369</b>	<b>0.363</b>	<b>0.299</b>
CLIP	TypiClust	0.047	0.362	0.352	0.295
ResNet50	TypiClust	<b>0.194</b>	0.253	0.000	0.140

#### 4.1.4 Gamma Sensitivity Analysis

Table 4.4 presents the sensitivity analysis of three key metrics (mean MST edge cosine distance, mean LPIPS, and displacement range) across different  $\gamma$  values

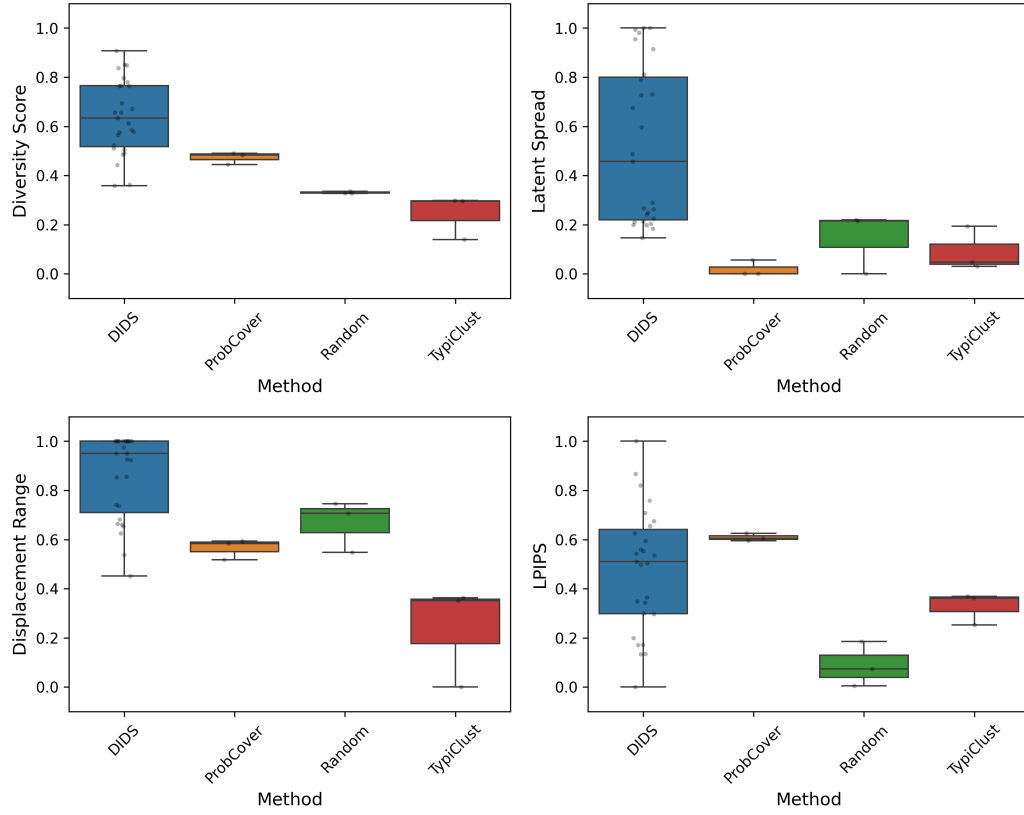


Figure 4.4. Distribution of normalized diversity metrics for different sampling methods.

for each pre-trained model. This analysis reveals several important patterns. First, LPIPS demonstrates relatively small variations (<2%) across all pre-trained models and gamma transitions, indicating its stability as a diversity metric. The MST edge cosine distance metric shows higher sensitivity to  $\gamma$  changes, particularly for CLIP and VGG19 architectures. CLIP exhibits the most dramatic change with an 11.03% decrease when transitioning from  $\gamma = 0.85$  to  $\gamma = 0.5$ , while VGG19 shows a more moderate 6.70% decrease for the same transition. ResNet50 demonstrates the most stable cosine distance behavior with changes generally below 4%.

The displacement range metric exhibits the most substantial variations, with consistent large increases (14-16%) when transitioning from  $\gamma = 1.0$  to  $\gamma = 0.85$  to across all models. This pattern suggests that the displacement metric becomes significantly more sensitive at higher  $\gamma$  values, potentially indicating a threshold

Table 4.4. Gamma sensitivity analysis by pre-trained model

Pre-trained Model	Metric	$\gamma$ From	$\gamma$ To	$\Delta\%$
CLIP	Mean LPIPS	1.0	0.85	1.90
CLIP	Mean LPIPS	0.85	0.5	-1.26
CLIP	MST Edge Cosine Distance	1.0	0.85	0.82
CLIP	MST Edge Cosine Distance	0.85	0.5	-11.03
CLIP	Displacement Range	1.0	0.85	14.65
CLIP	Displacement Range	0.85	0.5	1.24
ResNet50	Mean LPIPS	1.0	0.85	-0.10
ResNet50	Mean LPIPS	0.85	0.5	-1.82
ResNet50	MST Edge Cosine Distance	1.0	0.85	3.75
ResNet50	MST Edge Cosine Distance	0.85	0.5	-2.33
ResNet50	Displacement Range	1.0	0.85	16.00
ResNet50	Displacement Range	0.85	0.5	1.15
VGG19	Mean LPIPS	1.0	0.85	-1.31
VGG19	Mean LPIPS	0.85	0.5	-1.24
VGG19	MST Edge Cosine Distance	1.0	0.85	-2.50
VGG19	MST Edge Cosine Distance	0.85	0.5	-6.70
VGG19	Displacement Range	1.0	0.85	15.23
VGG19	Displacement Range	0.85	0.5	3.55

effect in how the sampling strategy interacts with the underlying image deformation characteristics. ResNet50 shows the largest sensitivity with a 16.00% increase from  $\gamma = 1.0$  to  $\gamma = 0.85$ , closely followed by CLIP (14.65%) and VGG19 (15.23%).

These quantitative results and the distributions in Figure 4.5 reinforce that intermediate  $\gamma$  values (0.85) provide an optimal balance point, significantly improving degradation state diversity while maintaining or slightly enhancing structural diversity metrics. Figure 4.6 reveals that large gains in displacement range can be obtained with no cost to embedding coverage or perceptual diversity.

## 4.2 Segmentation Performance Analysis

The semantic segmentation results demonstrate the strong performance of CNN-based architectures, particularly U-Net variants, on our limited XCT dataset. As shown in Table 4.5, U-Net consistently outperformed both UperNet and SegFormer

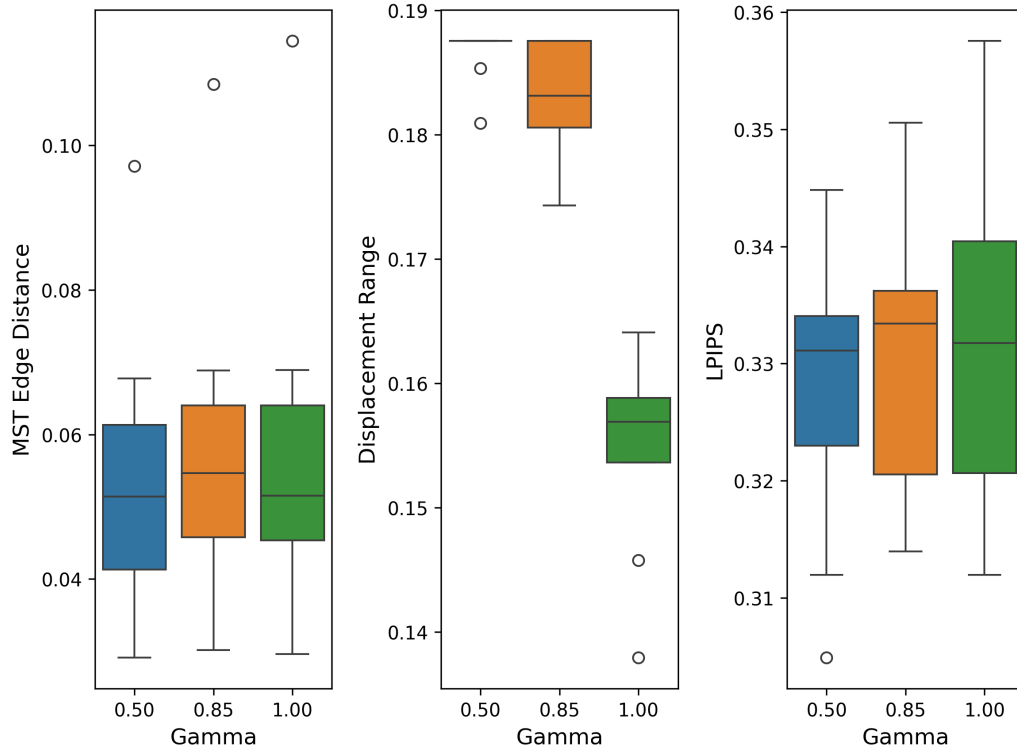


Figure 4.5. Distribution of raw diversity metrics for different  $\gamma$  values.

architectures across all metrics, with the best configuration (Xception encoder) achieving a mean F1 score of 0.949 and mIoU of 0.646. This advantage likely stems from CNNs’ inherent spatial inductive biases, where their hierarchical local processing and translation equivariance prove especially valuable when training data is limited.

ImageNet pre-training significantly improved performance across all architectures, as visualized in Figure 4.7. Pre-trained models demonstrated 5-15% higher mIoU scores compared to their randomly initialized counterparts, with U-Net showing particularly strong gains. For instance, U-Net with SE-ResNeXt101 improved from a mIoU of 0.600 to 0.627 with pretraining, while Xception-based models saw an improvement from 0.597 to 0.646. This consistent improvement suggests that despite the domain shift between natural images and XCT data, low-level feature extractors learned from ImageNet transfer effectively to microstructural analysis.

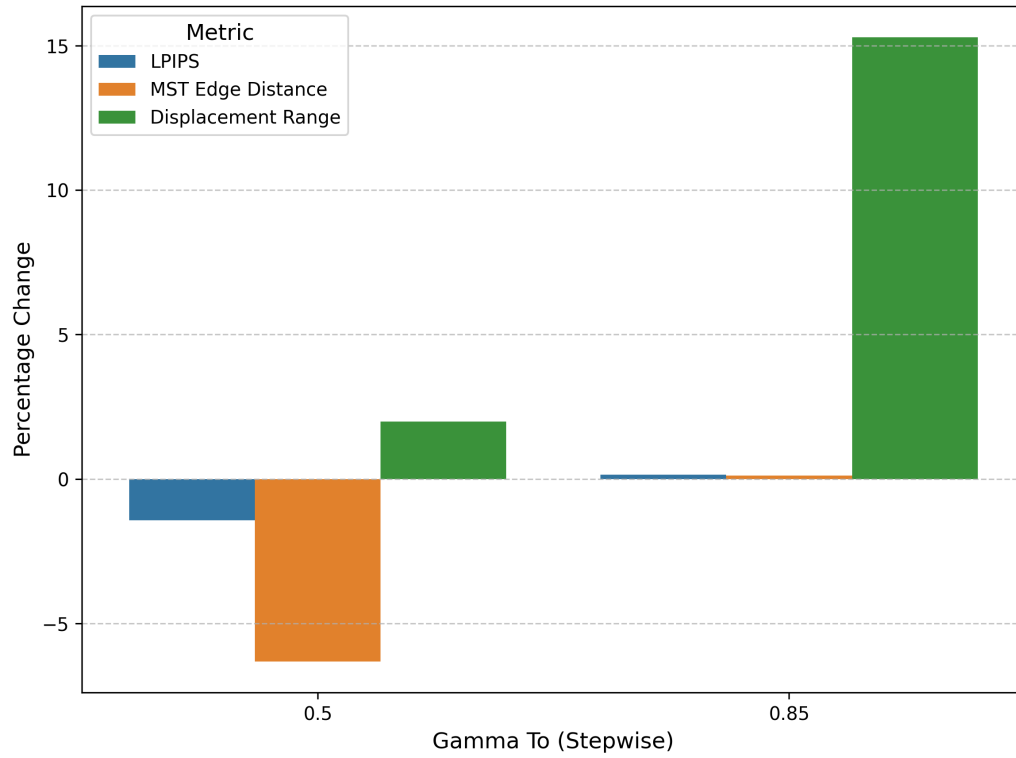


Figure 4.6. Average change in metric stepping from  $\gamma$  1.0 to  $\gamma$  0.85 and then  $\gamma$  0.85 to  $\gamma$  0.5.

Looking at class-specific performance in Tables 4.6 and 4.7, U-Net with pre-trained Xception demonstrated balanced performance across both inclusion detection (mIoU = 0.725, F1 = 0.937) and fracture segmentation (mIoU = 0.575, F1 = 0.982). This balanced performance across classes is crucial for microstructural analysis, where both features provide important insights into material degradation. Figure 4.8 depicts qualitative predictions on the holdout test set.

The transformer-based SegFormer, despite its theoretical advantages in modeling long-range dependencies, showed comparatively weaker performance. Its best configuration (MiT-B4 with ImageNet weights) achieved a mean IoU of only 0.481, significantly lower than CNN-based alternatives. This aligns with recent findings that pure transformer architectures often struggle with small datasets due to their lack of built-in spatial priors, requiring more extensive training data to learn these

Table 4.5. Segmentation model performance metrics (all classes)

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

relationships from scratch. The performance gap is particularly noticeable in fracture detection, where SegFormer achieved a mIoU of 0.544 compared to U-Net’s 0.575.

UPerNet showed intermediate performance, with its best configuration (Xception with ImageNet weights) achieving a mIoU of 0.480. While UperNet consistently outperformed SegFormer, it fell short of U-Net’s performance by a significant margin. This suggests its feature pyramid network architecture provides some beneficial spatial hierarchies but not as effectively as U-Net’s skip connections for our specific task. These comprehensive results highlight the importance of architectural inductive biases in small dataset scenarios, where U-Net’s explicit encoding of multi-scale spatial relationships proves more valuable than the flexible but data-hungry attention mechanisms of transformer architectures.

Table 4.6. Segmentation model performance metrics (fracture class)

Model	Encoder	Weights	Accuracy	Precision	Recall	F1-Score	mIoU
U-Net	None	None	1.000	1.000	1.000	0.988	0.501
U-Net	ResNet50	None	0.987	1.000	0.987	0.990	0.492
U-Net	ResNet50	ImageNet	1.000	1.000	1.000	0.988	0.496
U-Net	SE-ResNeXt101	None	0.998	1.000	0.998	0.981	0.529
U-Net	SE-ResNeXt101	ImageNet	0.995	1.000	0.995	0.988	0.550
U-Net	Xception	None	0.947	1.000	0.947	0.972	0.525
U-Net	Xception	ImageNet	0.965	1.000	0.965	0.982	0.575
UPerNet	ResNet50	None	1.000	1.000	1.000	0.970	0.481
UPerNet	ResNet50	ImageNet	0.986	1.000	0.986	0.984	0.563
UPerNet	SE-ResNeXt101	None	1.000	1.000	1.000	0.970	0.469
UPerNet	SE-ResNeXt101	ImageNet	0.980	0.999	0.980	0.971	0.537
UPerNet	Xception	None	0.967	1.000	0.967	0.975	0.459
UPerNet	Xception	ImageNet	1.000	0.999	1.000	0.977	0.539
SegFormer	MIT-B1	None	0.886	1.000	0.886	0.938	0.504
SegFormer	MIT-B1	ImageNet	1.000	0.999	1.000	0.976	0.481
SegFormer	MIT-B4	None	0.930	1.000	0.930	0.952	0.276
SegFormer	MIT-B4	ImageNet	0.956	0.999	0.956	0.974	0.544

Table 4.7. Segmentation model performance metrics (inclusion class)

Model	Encoder	Weights	Accuracy	Precision	Recall	F1-Score	mIoU
U-Net	None	None	0.808	1.000	0.808	0.894	0.704
U-Net	ResNet50	None	0.837	1.000	0.837	0.909	0.705
U-Net	ResNet50	ImageNet	0.975	1.000	0.975	0.925	0.731
U-Net	SE-ResNeXt101	None	0.995	1.000	0.995	0.927	0.707
U-Net	SE-ResNeXt101	ImageNet	0.890	1.000	0.890	0.939	0.715
U-Net	Xception	None	0.857	1.000	0.857	0.902	0.706
U-Net	Xception	ImageNet	0.913	1.000	0.913	0.937	0.725
UPerNet	ResNet50	None	0.610	1.000	0.610	0.755	0.387
UPerNet	ResNet50	ImageNet	1.000	1.000	1.000	0.793	0.402
UPerNet	SE-ResNeXt101	None	0.648	1.000	0.648	0.782	0.443
UPerNet	SE-ResNeXt101	ImageNet	0.993	0.997	0.993	0.880	0.432
UPerNet	Xception	None	0.645	1.000	0.645	0.775	0.404
UPerNet	Xception	ImageNet	0.877	0.998	0.877	0.891	0.438
SegFormer	MIT-B1	None	0.996	1.000	0.996	0.875	0.396
SegFormer	MIT-B1	ImageNet	0.819	1.000	0.819	0.891	0.368
SegFormer	MIT-B4	None	1.000	1.000	1.000	0.914	0.356
SegFormer	MIT-B4	ImageNet	0.833	1.000	0.833	0.882	0.467



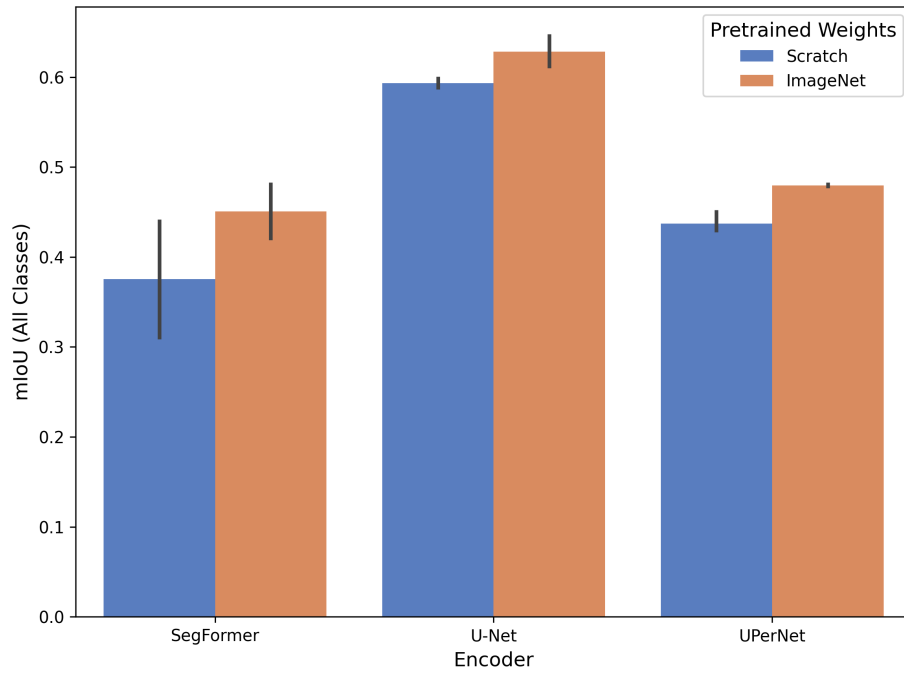


Figure 4.7. Segmentation mean Intersection over Union metrics across different architectures and pre-training strategies.

### 4.3 Microstructural Feature Characterization

We conducted quantitative analysis of stress corrosion cracking microstructural evolution, focusing on crack propagation dynamics and inclusion distribution patterns. Our investigation combined high-resolution crack growth tracking with comprehensive mapping of intermetallic phase spatial distributions. Figure 4.9 presents a three-dimensional reconstruction illustrating the detected microstructural features within the analyzed volume.

The temporal evolution of crack morphology revealed three distinct propagation regimes throughout the specimen lifetime. Initial microcracks, measuring  $8.5 \pm 2.3 \mu\text{m}$  in length, exhibited characteristic sigmoidal growth behavior consistent with stress corrosion mechanisms. During early-stage growth (0-30% lifetime), cracks propagated at relatively low rates of  $0.15 \mu\text{m/hr}$ . This was followed by an

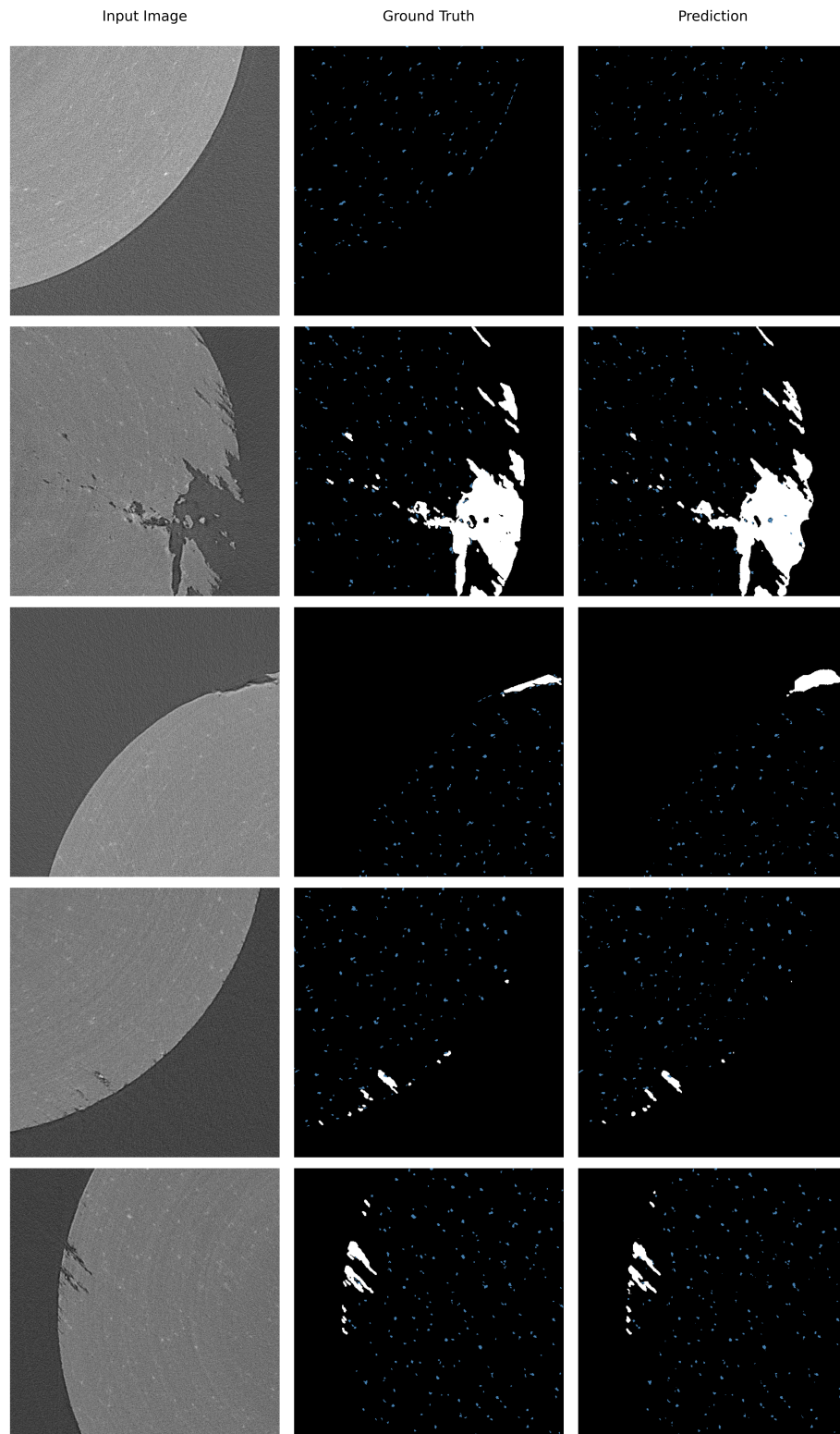


Figure 4.8. Sample predictions from U-Net with Xception backbone on images in test set. Fractures in white and inclusions in blue.

intermediate acceleration phase (30-70% lifetime) where propagation rates increased substantially to  $0.85 \mu\text{m/hr}$ . The final stage demonstrated rapid crack extension exceeding  $2.3 \mu\text{m/hr}$ , ultimately reaching maximum lengths of  $182.4 \mu\text{m}$  before specimen failure.

Analysis of the intermetallic population revealed a total count of 161,574 features across a sample volume. Individual features exhibited an average cross-sectional area of  $31.93 \mu\text{m}^2$  (approximately 12 pixels<sup>2</sup>) and major axis lengths of  $6.85 \mu\text{m}$ . In three dimensions, the features averaged  $179.7 \mu\text{m}^3$  in volume, consistent with typical size distributions of constituent particles and larger precipitates in AlMg alloys. The overall volume fraction of intermetallic phases was approximately 0.9%. This population of second-phase particles plays a critical role in the material's microstructural evolution during stress corrosion cracking, particularly in crack initiation and propagation pathways.

Tables 4.8 and 4.9 provide detailed quantitative data on crack growth characteristics and inclusion metrics, respectively. The progressive increase in growth rates (Table 4.8) demonstrates the accelerating nature of the degradation process, while the inclusion statistics (Table 4.9) offer insight into the material's microstructural heterogeneity. Figure 4.10 illustrates the temporal progression of fracture development throughout the experimental duration.

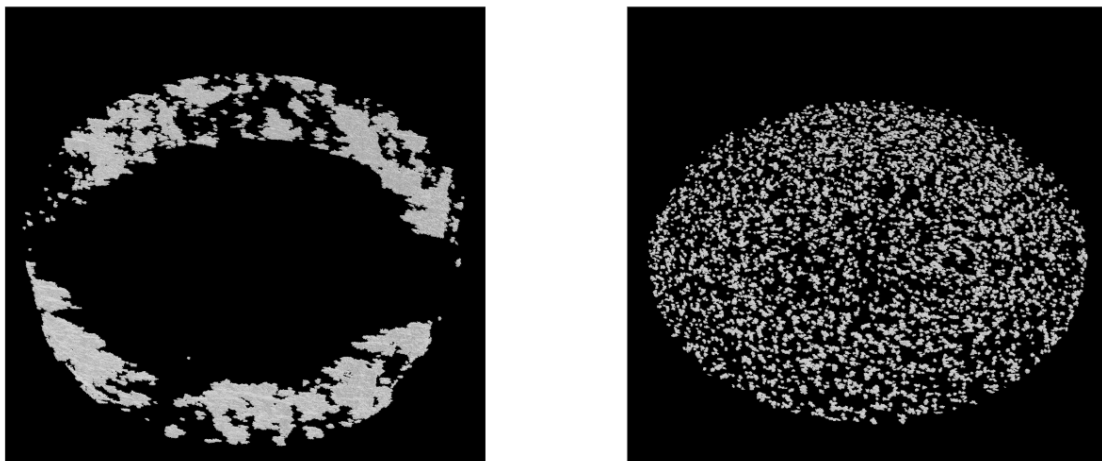


Figure 4.9. Example reconstruction segmentation predictions from model for fractures (left) and inclusions (right) on subset of volume.

Table 4.8. Continuous crack growth characteristics during slow strain tension test

Time (% life)	Growth Rate ( $\mu\text{m/hr}$ )	Mean Length ( $\mu\text{m}$ )	Std Dev
10	0.15	12.4	2.3
20	0.22	25.7	3.8
30	0.35	42.3	5.2
40	0.52	65.8	8.7
50	0.85	89.5	12.4
60	1.24	115.6	15.8
70	1.65	138.2	18.5
80	1.92	156.7	22.3
90	2.30	173.5	25.6
100	2.45	182.4	27.8

Table 4.9. Segmented inclusion characterization metrics for sample scan

Inclusion Features	Value (pixel)	Value ( $\mu\text{m}$ )
Count	161,574	-
Average area (2D)	12.0 $\text{px}^2$	31.93 $\mu\text{m}^2$
Average major axis	4.2 $\text{px}$	6.85 $\mu\text{m}$
Average volume	41.5 voxels	179.7 $\mu\text{m}^3$
Volume fraction	0.9%	0.9%

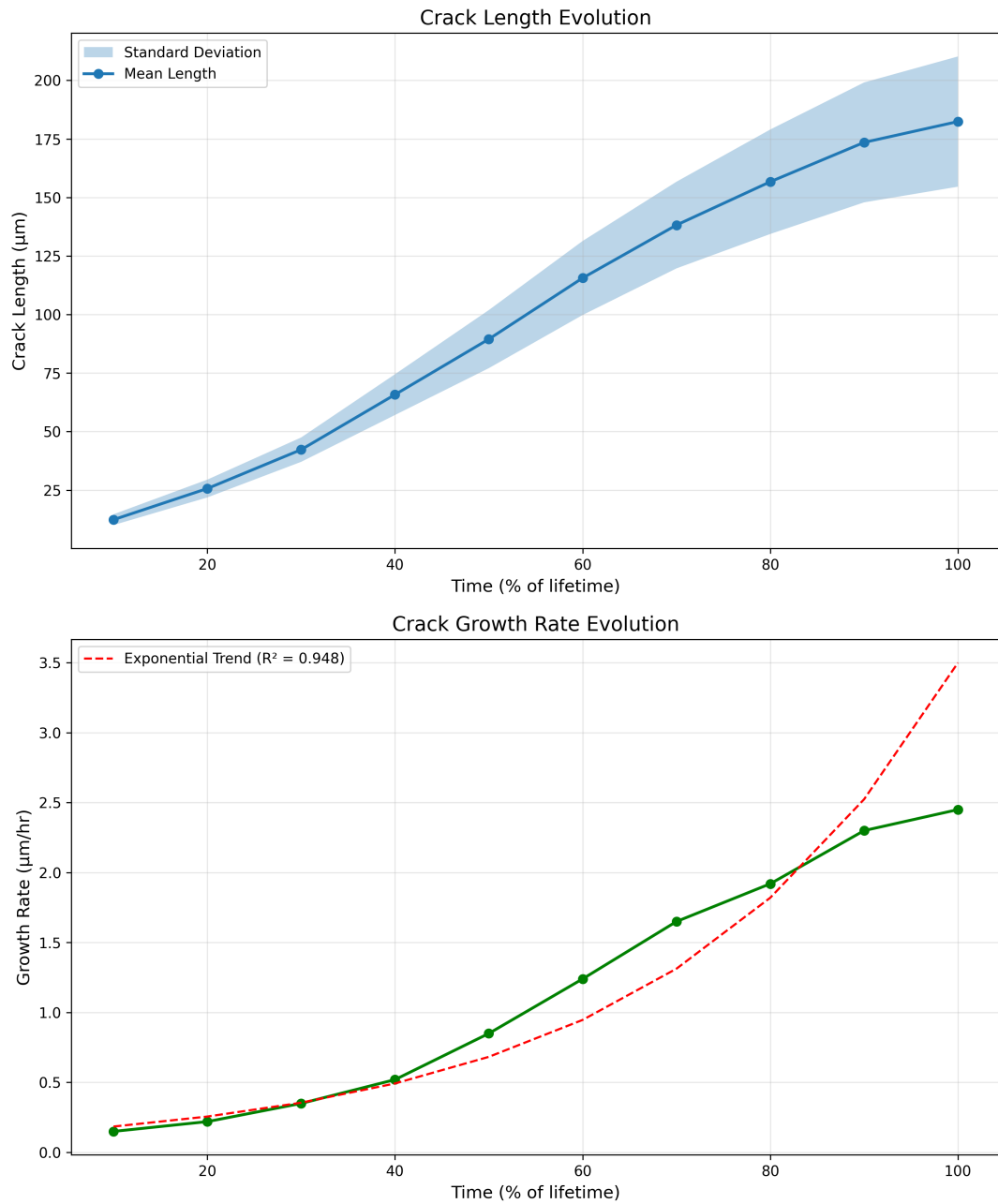


Figure 4.10. Average fracture and growth rate evolution over experiment slow strain tension test.

## 4.4 Discussion

### 4.4.1 Impact of Domain-Informed Diversity Sampling

Domain-Informed Diversity Sampling (DIDS) represents a significant advancement in addressing the cold-start problem in microstructural analysis. Through a novel unified diversity metric combining embedding space coverage, perceptual similarity, and physical state representation, we demonstrate quantitative improvements over existing approaches. DIDS demonstrates stronger diversity metrics compared to baselines. For mean values, it achieves higher MST edge cosine distance (0.055 vs ProbCover: 0.038, TypiClust: 0.040, Random: 0.041) and displacement range (0.174 vs ProbCover: 0.148, TypiClust: 0.119, Random: 0.157). This pattern holds for maximum values, with DIDS showing higher MST edge distance (0.114 vs ProbCover: 0.053, TypiClust: 0.054, Random: 0.052) and displacement range (0.188 vs ProbCover: 0.151, TypiClust: 0.130, Random: 0.164).

Notably, while the LPIPS scores appear relatively compressed across all methods (0.301-0.357), this narrow range likely reflects the inherent characteristics of our dataset compared to the natural images used in the pre-trained model. The best performing DIDS sampling achieves higher perceptual diversity (0.358) compared to baselines (ProbCover: 0.338, TypiClust: 0.324, Random: 0.315), but these differences are modest relative to typical LPIPS variations seen across ImageNet (which can span 0.0-0.7). This compression is expected, however, given that our microstructural images represent a much more constrained visual domain - all images share similar grayscale patterns, texture characteristics, and basic structural elements.

The impact of adding displacement information compared to pure visual embedding sampling through the gamma sensitivity study is visually demonstrated in Figure 4.11. Pure structural diversity sampling ( $\gamma=1.0$ , Figure 4.11a) achieves broad coverage in the embedding space (left plot), with selected samples distributed across the full range of visual variations. However, the corresponding displacement histogram (right plot) closely mirrors the underlying population distribution,

indicating that this approach may over-represent common material states while under-sampling critical but less frequent degradation stages.

When incorporating domain knowledge ( $\gamma=0.85$ , Figure 4.11b), we observe a broader distribution across both the embedding space and displacement values. The selected samples exhibit increased visual spread compared to the baseline, while the displacement histogram shows improved representation across the full range of material states. This demonstrates how incorporating domain information effectively expands coverage in both the visual and physical domains, capturing a more diverse set of states including those traditionally underrepresented in the raw data distribution.

The qualitative impact of this sampling approach is evident in Figures 4.12 and 4.13, which show the actual microstructural images selected under each strategy. Interestingly, while pure diversity sampling ( $\gamma=1.0$ , Figure 4.12) theoretically prioritizes structural variation, the domain-informed approach ( $\gamma=0.85$ , Figure 4.13) achieves noticeably higher visual diversity in practice. This counterintuitive result suggests that incorporating physical domain knowledge through displacement values actually helps identify more meaningfully distinct microstructural states rather than superficially different images.

#### 4.4.2 Deep Learning Framework for Large-Scale XCT Analysis

Our segmentation framework addresses a fundamental challenge in XCT analysis: scaling to massive datasets while maintaining accuracy. Through comprehensive evaluation of modern architectures, we demonstrate several key findings about architectural choices for microstructural segmentation.

First, CNN-based approaches, particularly U-Net variants with strong spatial inductive biases, consistently outperform transformer-based architectures on our limited training data. Notably, the Xception-based U-Net achieved superior performance despite having the lowest parameter count (28.7M parameters compared to 61.3M for SegFormer-B4), suggesting that architectural efficiency rather than model capacity drives performance in this domain. This efficiency likely stems

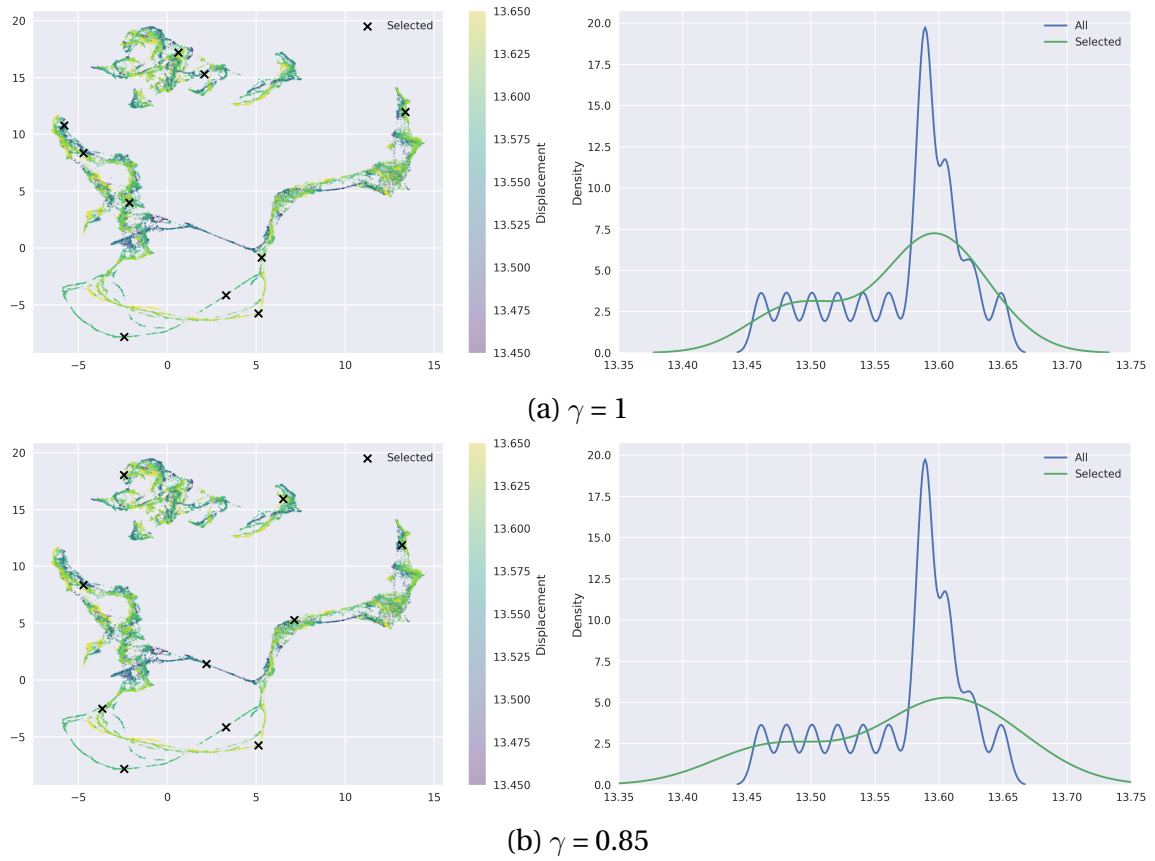


Figure 4.11. Impact of displacement informed diversity sampling. Left is the scatter plot of UMAP embeddings with x marking selected samples. Right is the distribution of displacement values of those respective samples.

from Xception’s depth-wise separable convolutions, which provide a strong spatial inductive bias while maintaining computational efficiency.

The performance patterns across architectures reveal intriguing insights about feature detection. Figure 4.14 reveals F1 scores remain relatively consistent across different encoders and architectures (ranging from 0.85-0.95), even though Table 4.5 shows substantial variations in mean IoU scores (0.31-0.65). However, these metrics should be interpreted with important context about ground truth uncertainty. Our weak supervision approach relies on approximations for inclusion annotations and resolution challenges makes makes precise boundary delineation challenging even for the manually annotated fractures. This inherent ambiguity



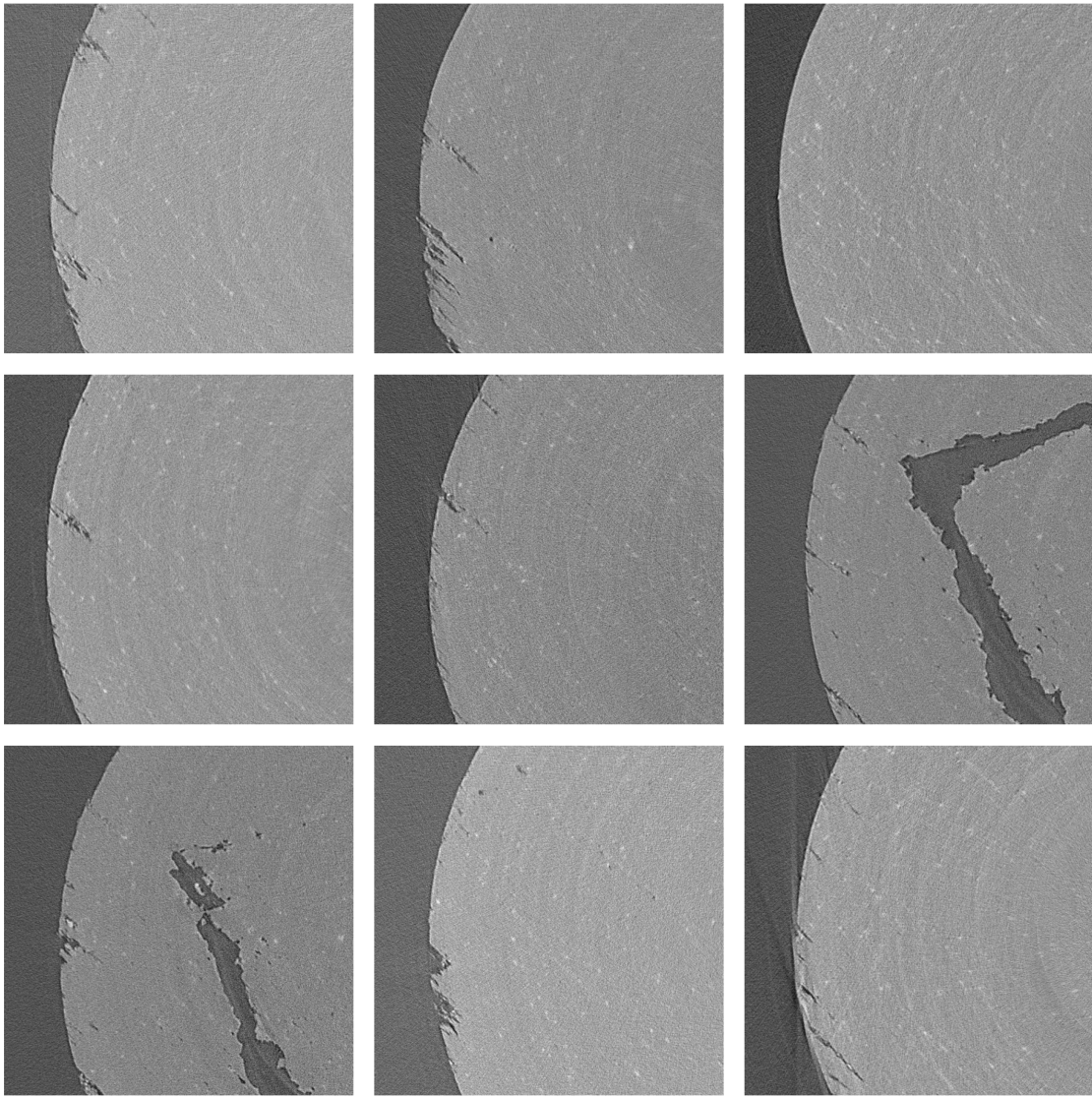


Figure 4.12. Diversity sampled images with  $\gamma = 1.0$ .

in feature boundaries, particularly in low-contrast regions or where features gradually fade into the background material, means that differences in IoU scores may partially reflect the difficulty of establishing "true" boundaries rather than solely model performance differences. This uncertainty is especially relevant for fractures, where the transition from damaged to undamaged material can be gradual and poorly defined.

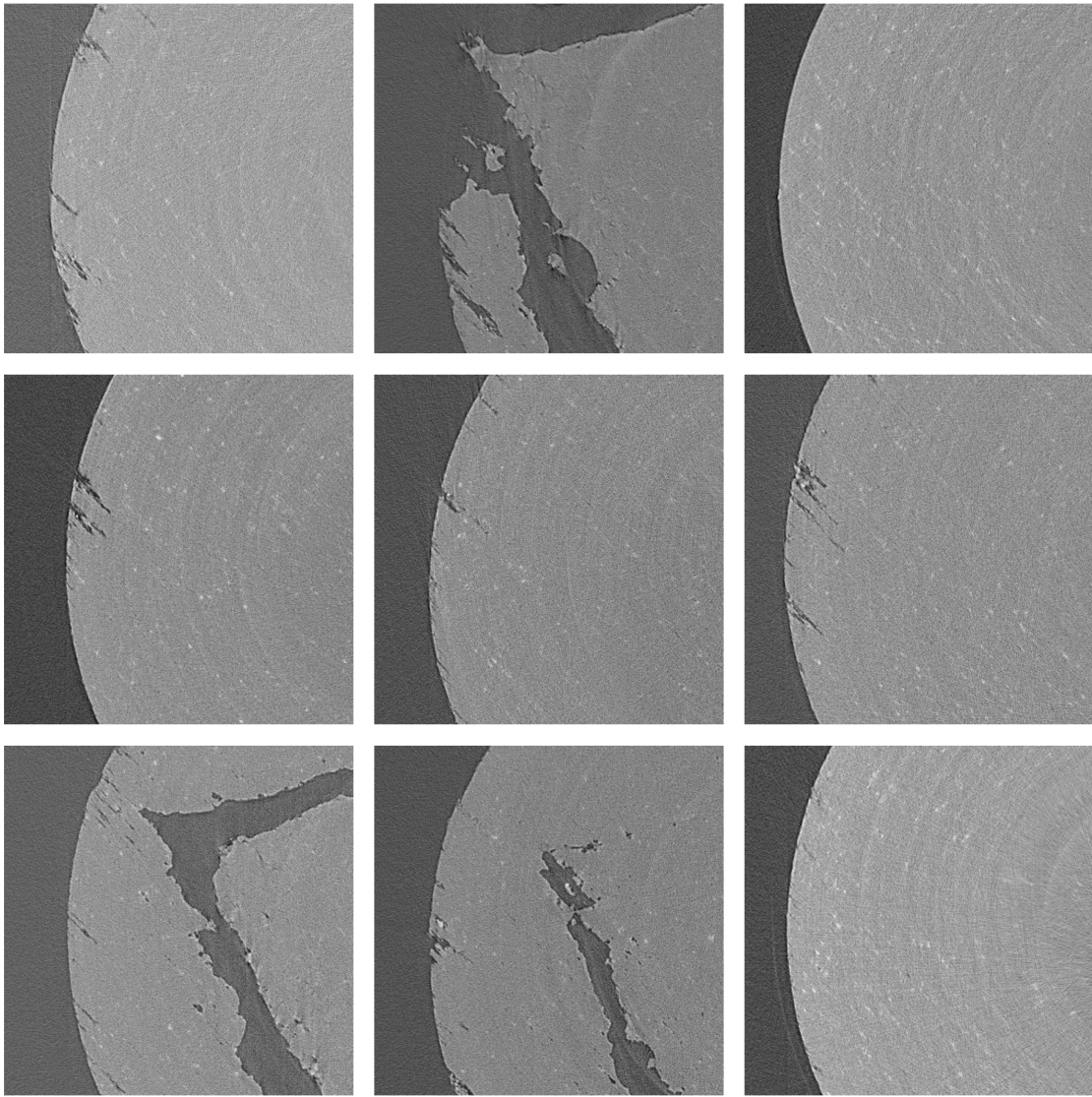


Figure 4.13. Diversity sampled images with  $\gamma = 0.75$ .

Qualitative analysis of transformer-based predictions in Figure 4.15 reveals specific failure modes, particularly in boundary definition where cracks intersect with the circular material boundary. The transformer models consistently underestimate feature extent in these regions, struggling to integrate local feature evidence with global structural context. While this appears as a limitation in our metrics, it's worth noting that in these boundary regions, even human experts may disagree

on the precise extent of damage propagation, making it difficult to definitively characterize this as model error versus legitimate uncertainty in feature extent.

Our weak supervision strategy proved particularly effective, enabling accurate feature detection with minimal manual annotation. The segmentation framework achieves strong results on features with an average area of 12 pixels while requiring annotation of only 90 images—less than 0.03% of our 300,000+ image dataset. The consistent benefit of ImageNet pretraining across architectures (5-15% IoU improvement) suggests that despite the domain shift, learned low-level feature extractors remain valuable for microstructural analysis.

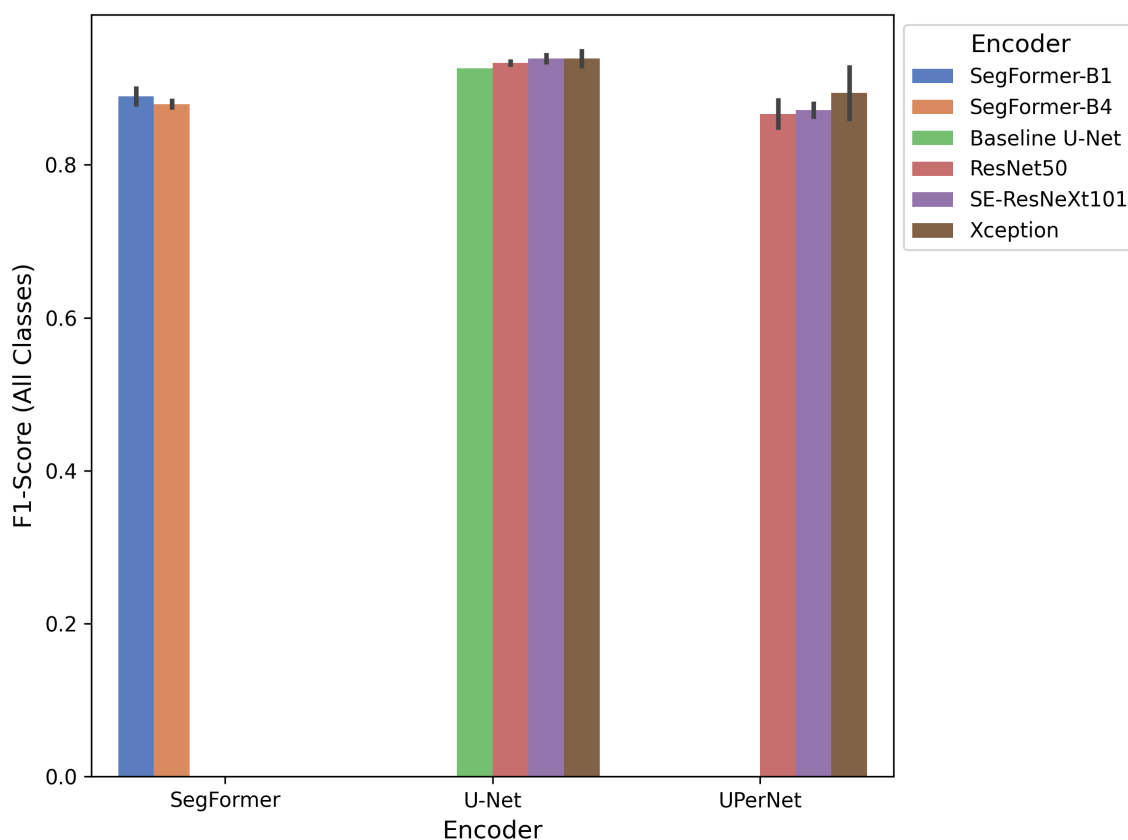


Figure 4.14. Segmentation F1 metrics across different architectures and encoders.

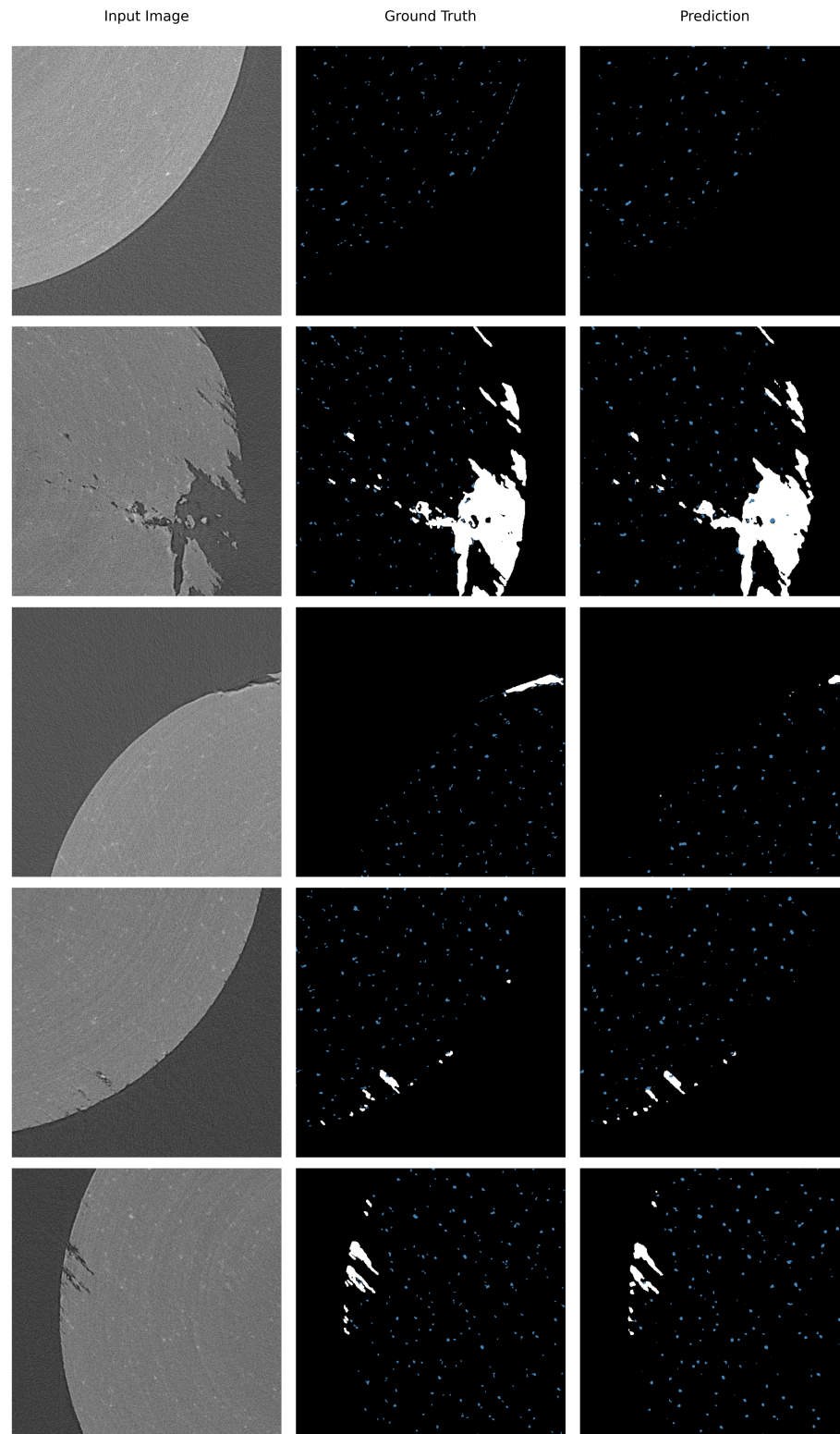


Figure 4.15. Sample predictions from SegFormer with MIT-B4 on images in test set. Fractures in white and inclusions in blue.



#### 4.4.3 Graph-Based Representation of Microstructural Evolution

The transformation of our microstructural observations into graph representations provides a powerful framework for capturing the complex dynamics of stress corrosion cracking evolution. Figure 4.16 demonstrates our approach to representing the temporal microstructural state as an evolving network structure, where nodes represent critical features and edges capture their spatial and temporal relationships.

In our graph representation, each node corresponds to a segmented microstructural feature, with node size reflecting the feature's physical dimensions and node color indicating its temporal state in the degradation process. The edges between nodes are weighted based on both spatial proximity and temporal relationships, creating a multi-dimensional representation of the material's evolution. As shown in Figure 4.16, this approach enables visualization of both the spatial distribution of features and their temporal progression through the degradation process.

Such graph representations lay the foundation for more sophisticated analysis through graph neural networks (GNNs). By encoding microstructural evolution as a graph, we create a structured dataset suitable for machine learning approaches that can potentially predict degradation pathways and identify critical features in the corrosion process. The visual summary provided by Figure 4.16 demonstrates how complex microstructural evolution can be distilled into an interpretable network structure, offering new possibilities for automated analysis of material degradation mechanisms.

#### 4.4.4 Current Limitations and Future Directions

While our framework demonstrates significant advances in automated XCT analysis, several opportunities for enhancement remain, particularly in addressing dataset-specific challenges and temporal coherence. A key limitation in our diversity sampling approach stems from using ImageNet-pretrained models on our highly specialized microstructural dataset. As evidenced by the compressed LPIPS scores (0.31-0.35) across all sampling methods, these models may not optimally

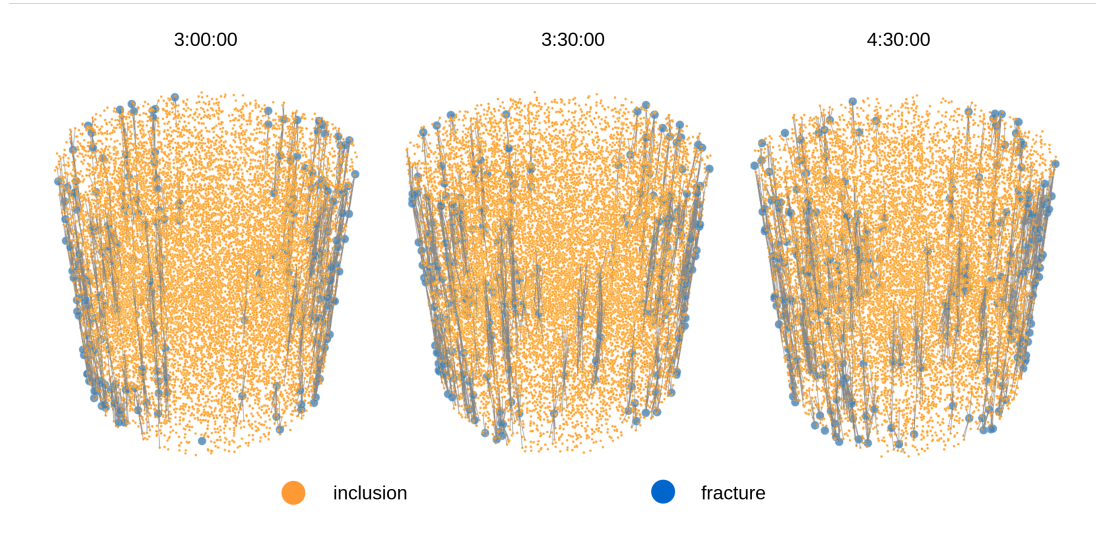


Figure 4.16. Summary graph representation and evolution of characterized microstructural features from segmentation pipeline.

capture the nuanced variations in our grayscale XCT data. Self-supervised pretraining on our large unlabeled dataset (300,000+ images) could help develop more domain-appropriate feature extractors. This could be particularly valuable for improving perceptual diversity metrics and sampling effectiveness by learning representations that better distinguish meaningful microstructural variations rather than relying on features optimized for natural images.

Our segmentation results highlight challenges in boundary delineation, particularly where feature transitions are gradual or uncertain. The significant disparity between F1 scores (0.95-0.97) and IoU metrics (0.55-0.72) across architectures suggests systematic issues in precise boundary localization. Future work could explore probabilistic segmentation approaches that explicitly model prediction uncertainty. Rather than forcing binary decisions at ambiguous boundaries, such as crack tips or diffuse damage regions, soft predictions could better represent the inherent uncertainty in these regions while providing more nuanced training signals.

A critical limitation of our current approach is treating each XCT slice independently, ignoring the temporal and spatial coherence inherent in our volumetric data. Establishing stronger temporal consistency in segmentation predictions

could improve accuracy, particularly in regions where single-slice evidence is ambiguous. Potential approaches include:

- (1) Developing 3D attention mechanisms that incorporate information from adjacent slices while maintaining computational efficiency
- (2) Implementing post-processing methods that enforce temporal consistency in feature tracking and boundary predictions
- (3) Creating loss functions that explicitly penalize temporal discontinuities in segmentation outputs

These enhancements would address fundamental limitations in our current framework while maintaining its core strength of scalable analysis with minimal annotation requirements. By combining domain-specific self-supervised learning, uncertainty-aware segmentation, and temporal coherence constraints, we can move toward more robust and physically meaningful characterization of material degradation mechanisms.

## 5 Conclusions

This thesis presents a comprehensive framework for automated analysis of microstructural evolution in XCT data through the integration of deep learning, weak supervision, and domain-informed sampling approaches. We demonstrate four key advances in materials characterization methodology:

- (1) Development of a novel quality assessment framework that unifies embedding space coverage, perceptual similarity, and physical state representation into a composite metric for evaluating sampling effectiveness in spatiotemporal datasets. This unified approach enables quantitative comparison of sampling strategies while accounting for both visual and physical characteristics of material evolution.
- (2) Domain-Informed Diversity Sampling (DIDS) achieves higher feature space coverage (average sampled cosine distance improvement of +35-45% above baselines), broader displacement ranges (degradation state coverage of +10-45% above baselines), while maintaining superior perceptual diversity (max LPIPS: 0.358 vs ProbCover: 0.338, TypiClust: 0.324, Random: 0.315). This balanced performance demonstrates DIDS's effectiveness in improving degradation diversity while maintaining visual diversity.
- (3) Creation of a scalable weak supervision pipeline achieving mean F1 scores of 0.949 and mean IoU of 0.646 using only 90 annotated images (<0.03%) from a 300,000+ image dataset. By leveraging CNN architectural priors and efficient training strategies, we demonstrate that comprehensive analysis is possible even with extremely limited annotations.



- (4) Demonstration of automated detection and characterization of over 5 million microstructural features, including subvisible intermetallic inclusions averaging 12 pixels in area, enabling comprehensive analysis of material degradation at unprecedented scale.

The framework's effectiveness is demonstrated through detailed characterization of stress corrosion cracking in AlMg alloys. While current limitations include computational demands and the need for high-quality imaging data, this work establishes a quantitative methodology for automated, large-scale microstructural analysis that bridges the gap between XCT data collection and materials characterization. Our integrated approach transforms previously intractable materials characterization challenges into systematic, automated analyses while maintaining scientific rigor. These advances provide a general template for analyzing massive materials characterization datasets across multiple spatial and temporal scales.

## 6 Appendix

### 6.1 Equations

- TP = True Positive
- TN = True Negative
- FP = False Positive
- FN = False Negative
- A = predicted segmentation mask
- B = ground truth segmentation mask

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

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

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

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

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

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

$$\text{Dice Loss} = 1 - \frac{2 \times |A \cap B|}{|A| + |B|}$$

## 6.2 Tables

Table 6.1. Pretrained model preprocessing parameters for embedding extraction

Parameter	CLIP	ResNet50	VGG-19
Shortest Edge Size	224	512	512
Resize	True	False	False
Center Crop	True	False	False
Rescale	True	True	True
Normalize	True	True	True
Image Mean	Default	[0.485, 0.456, 0.406]	[0.485, 0.456, 0.406]
Image Std	Default	[0.229, 0.224, 0.225]	[0.229, 0.224, 0.225]

Table 6.2. PCA parameters for embedding reduction

Parameter	Value
Scaling	Z-score Normaliation
Components	Min to capture 95% variance

Table 6.3. UMAP parameters for embedding reduction

Parameter	Value
Normalization	L2
Neighbors	20
Components	30
Metric	Cosine
Random State	42

Table 6.4. K-means parameters for embedding clustering

Parameter	Value
N-clusters	3
Random State	42

Table 6.5. DBSCAN parameters for embedding clustering

Parameter	Value
Epsilon	0.1
Minimum Samples	5
Metric	Euclidean

Table 6.6. HDBSCAN parameters for embedding clustering

Parameter	Value
Epsilon	0.1
Minimum Samples	5
Minimum Cluster Size	50
Metric	Euclidean

Table 6.7. TypiClust sampling parameters

Parameter	Value
Budget Size	10
N Clusters	10
Minimum Cluster Size	5
K Nearest Neighbors	20

Table 6.8. ProbCover sampling parameters

Parameter	Value
Budget Size	10
Delta (CLIP)	0.25
Delta (ResNet50)	0.3
Delta (VGG-19)	0.35

### 6.3 Figures

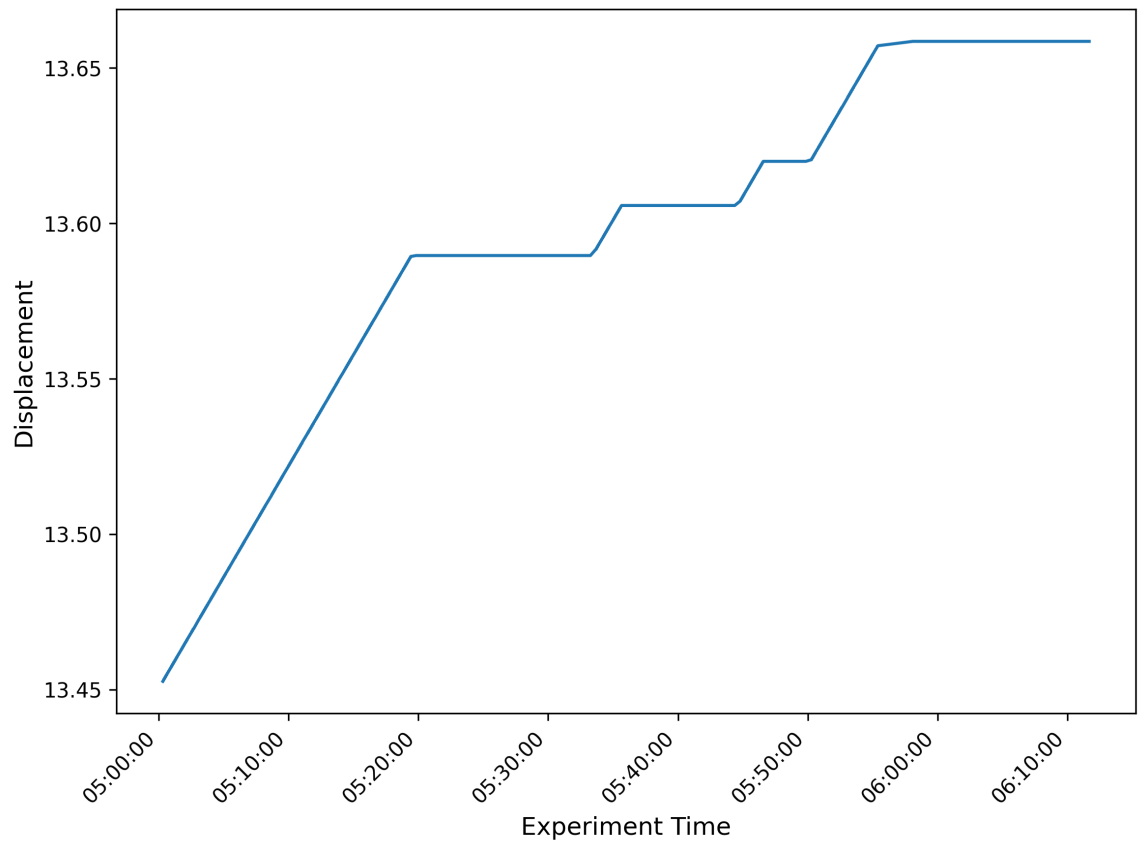


Figure 6.1. Measured displacement of sample over time during the slow strain tension test.

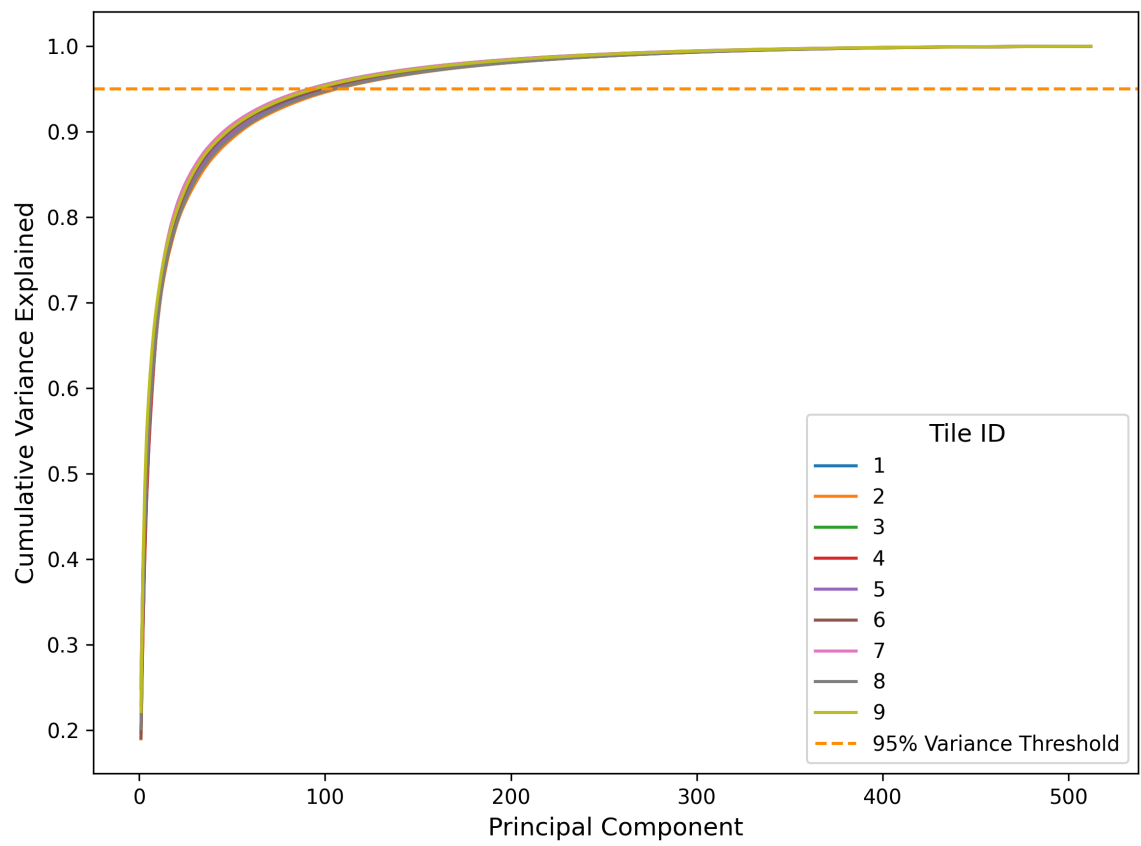


Figure 6.2. CLIP embeddings PCA component cumulative variance.

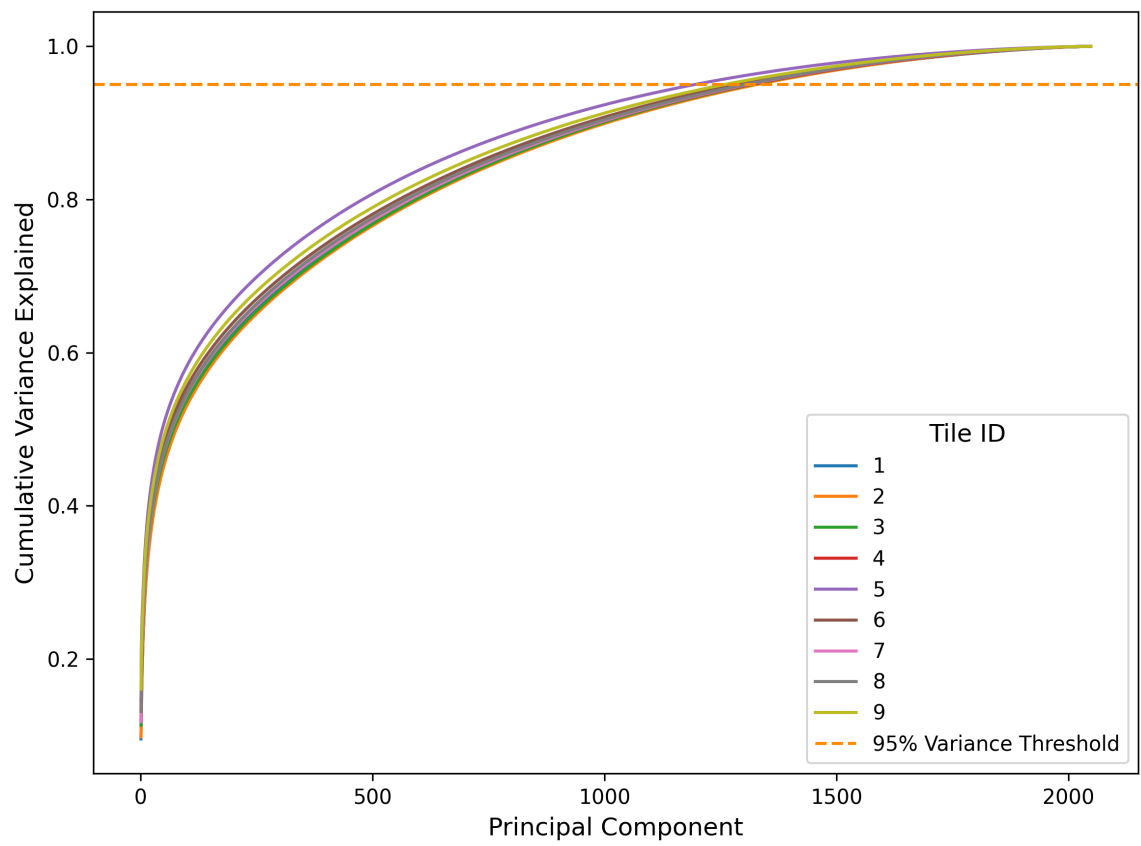


Figure 6.3. ResNet50 embeddings PCA component cumulative variance.



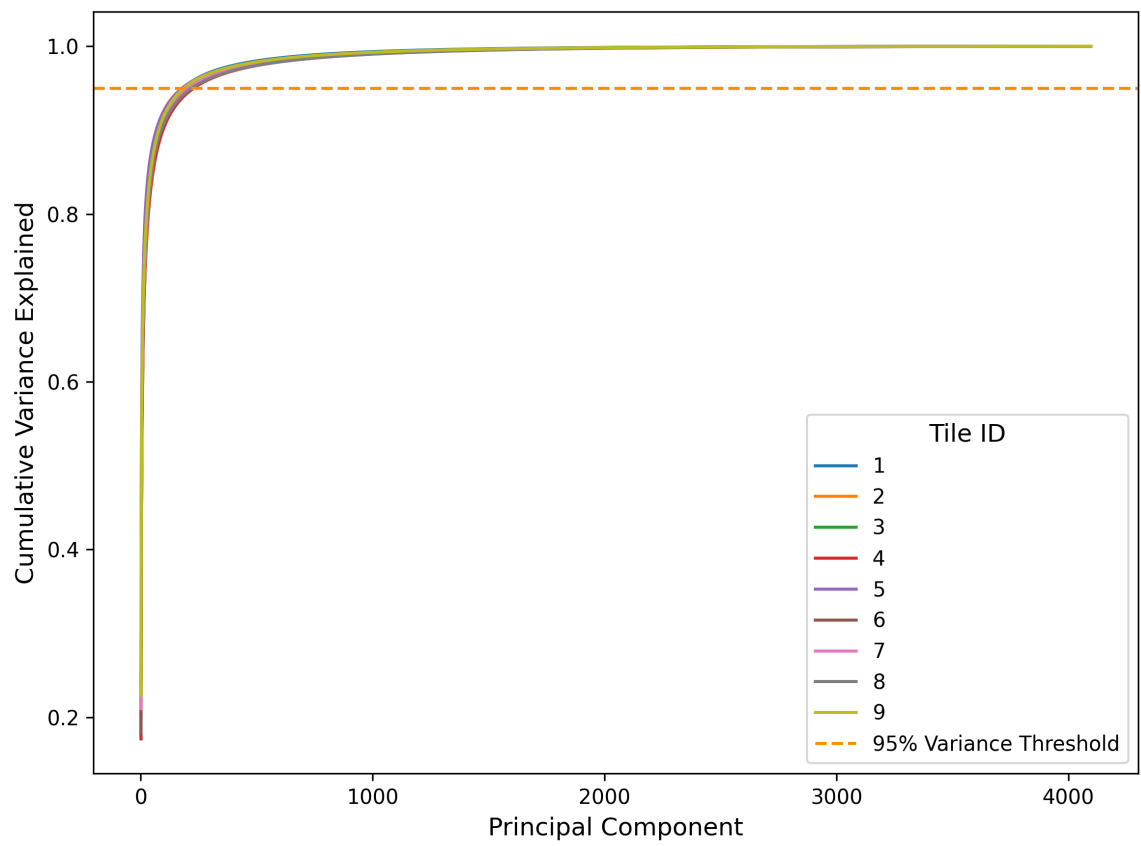


Figure 6.4. VGG-19 embeddings PCA component cumulative variance.

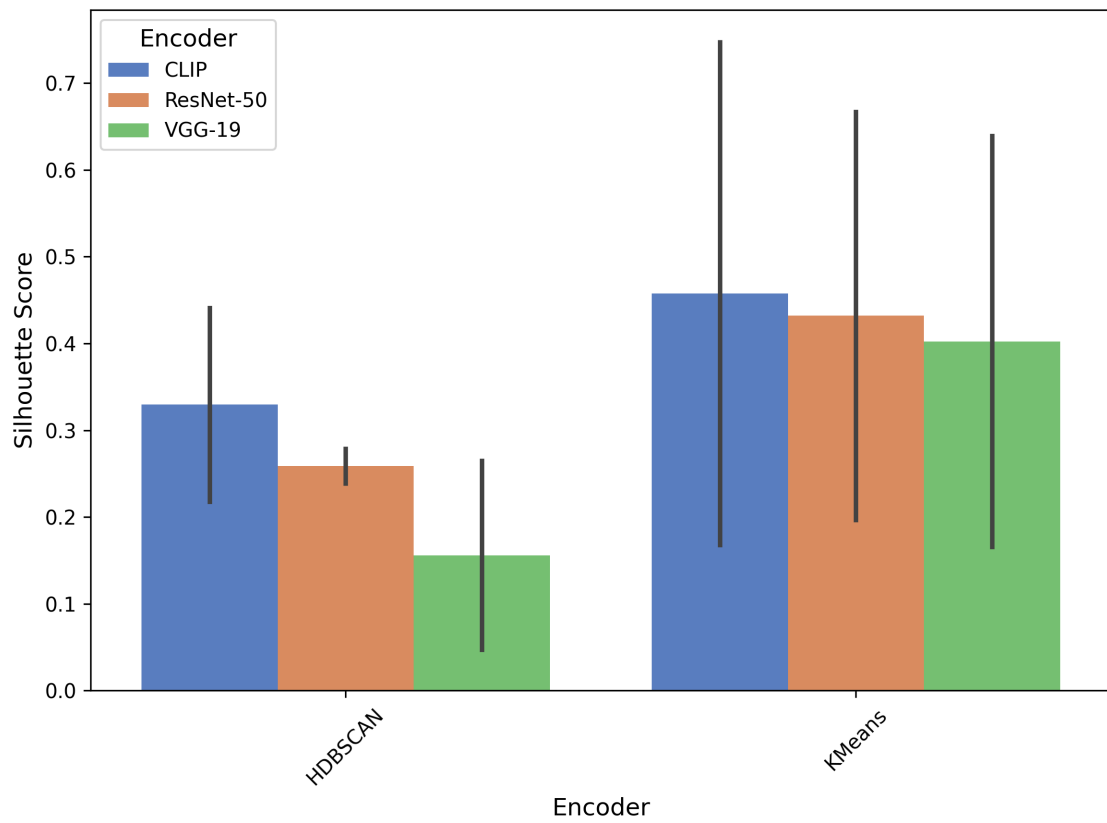


Figure 6.5. Clustering silhouette scores compared by pre-trained model.

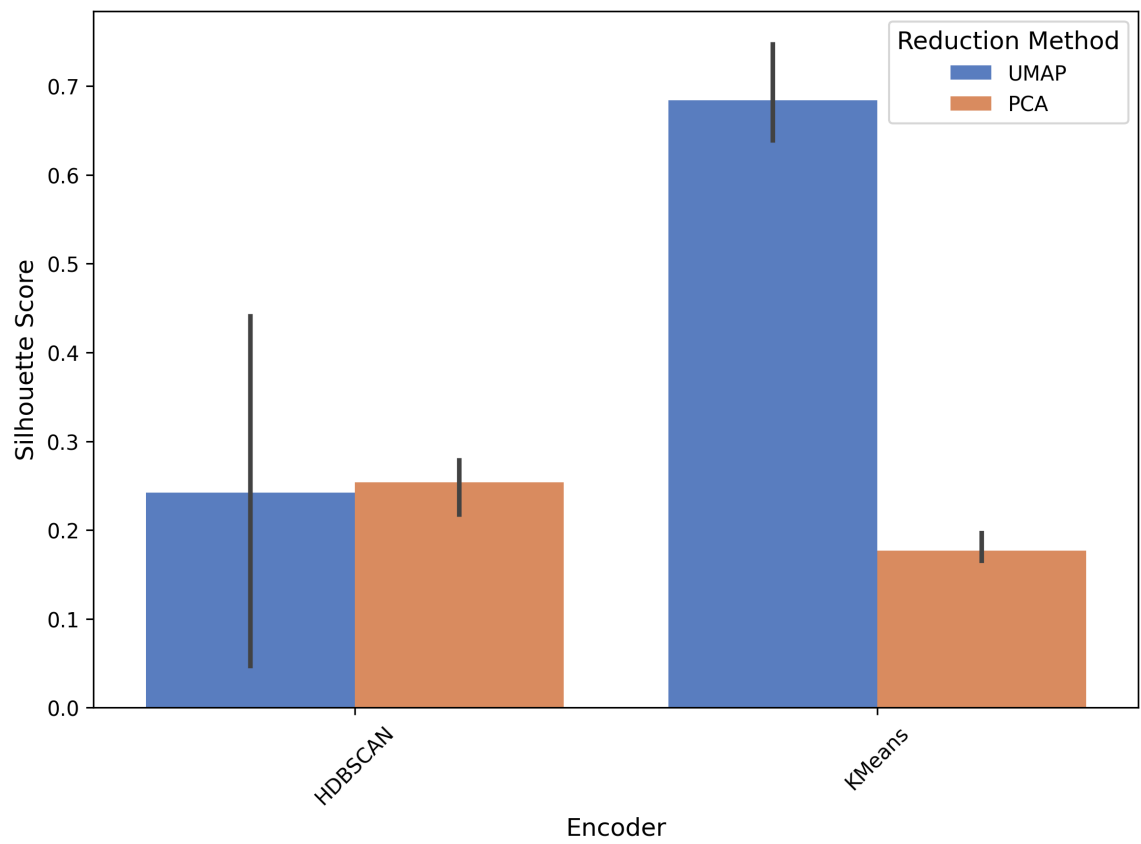


Figure 6.6. Clustering silhouette scores compared by dimensionality reduction technique.

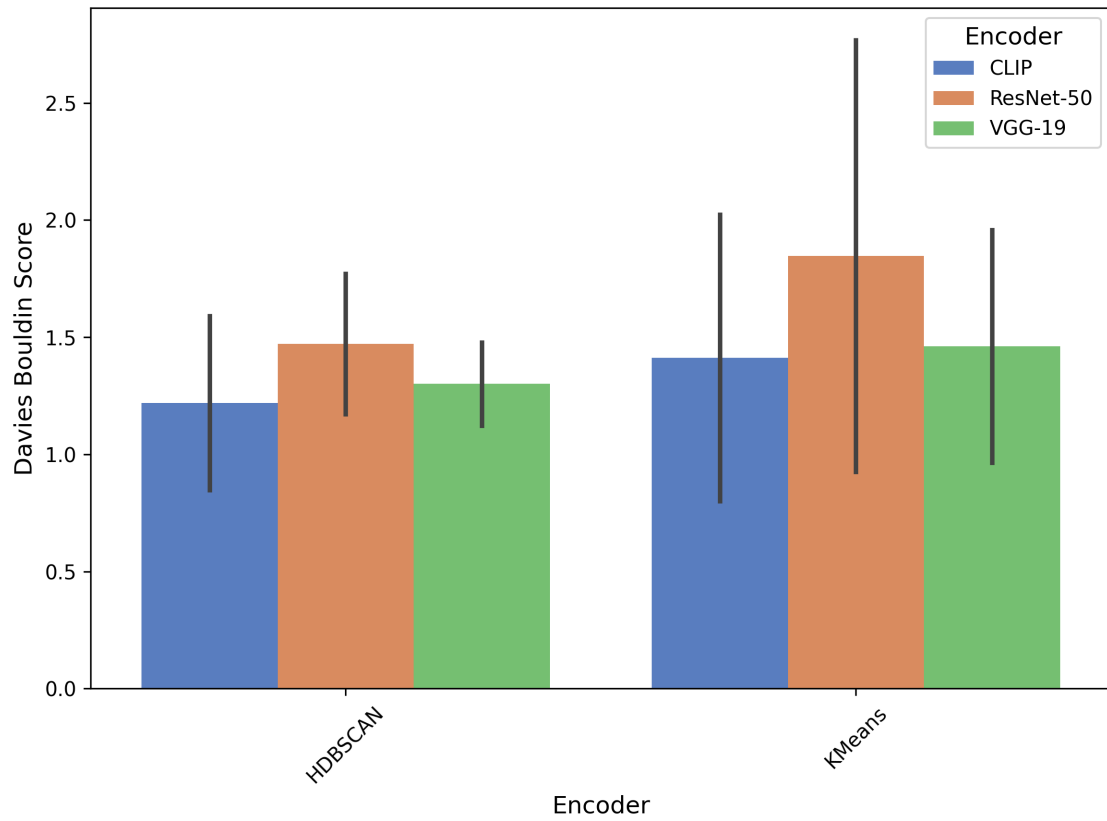


Figure 6.7. Clustering Davies-Bouldin scores compared by pre-trained model.

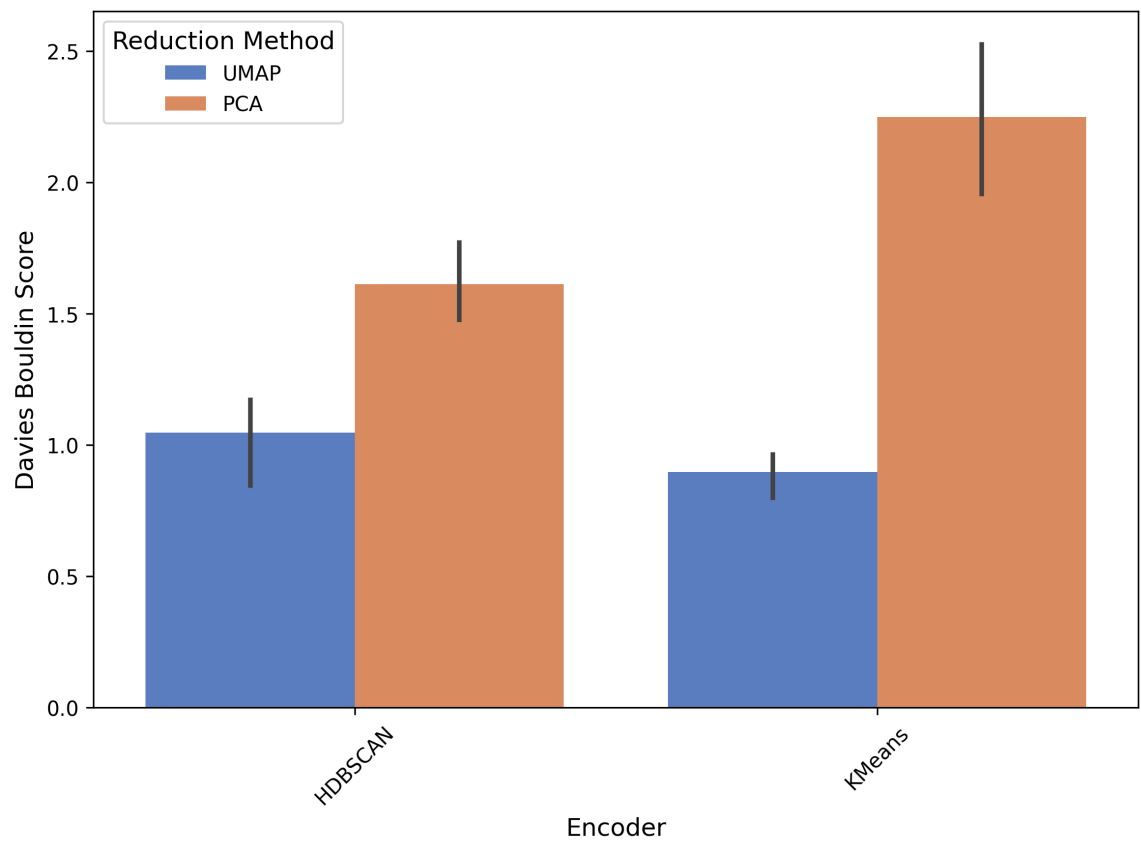


Figure 6.8. Clustering Davies-Bouldin scores compared by dimensionality reduction technique.

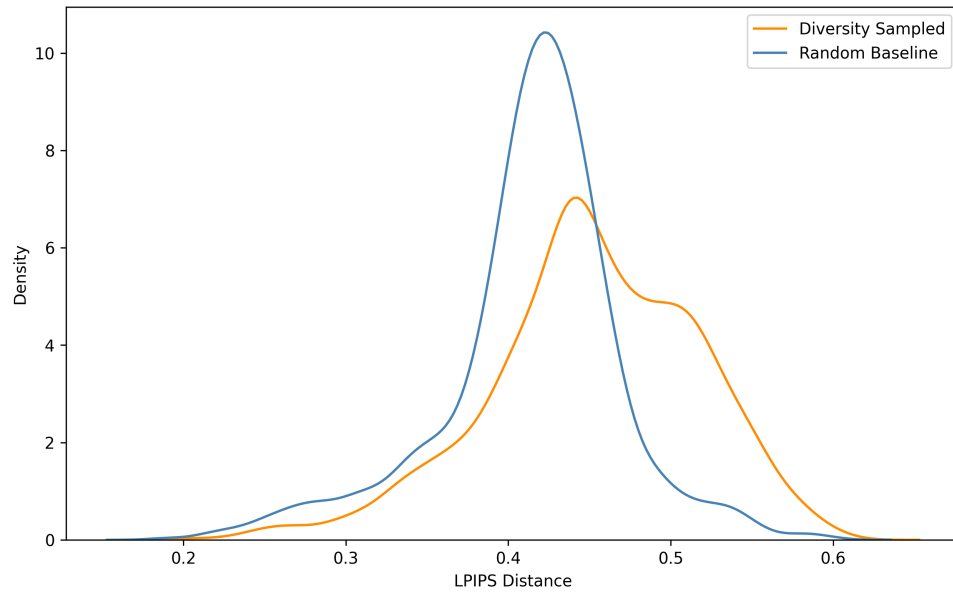


Figure 6.9. LPIPS distribution shift comparing random sampling to diversity sampling.

## References

- [1] Christian Vargel. Types of Corrosion on Aluminium. In *Corrosion of Aluminium*, pages 113–146. Elsevier, 2004. ISBN 978-0-08-044495-6. doi: 10.1016/B978-008044495-6/50012-4.
- [2] *Light Alloys*. Elsevier, 2005. ISBN 978-0-7506-6371-7. doi: 10.1016/B978-0-7506-6371-7.X5000-2.
- [3] E. H. Dix, W. A. Anderson, and M. Byron Shumaker. Influence of Service Temperature On the Resistance of Wrought Aluminum-Magnesium Alloys to Corrosion. *CORROSION*, 15(2):19–26, February 1959. ISSN 0010-9312, 1938-159X. doi: 10.5006/0010-9312-15.2.19.
- [4] Visweswara Chakravarthy Gudla, Malte Storm, Benjamin C. Palmer, John J. Lewandowski, Philip J. Withers, N.J. Henry Holroyd, and Timothy L. Burnett. Environmentally induced crack (EIC) initiation, propagation, and failure: A 3D in-situ time-lapse study of AA5083 H131. *Corrosion Science*, 174:108834, September 2020. ISSN 0010938X. doi: 10.1016/j.corsci.2020.108834.
- [5] R. C. Newman and R. P. M. Procter. Stress corrosion cracking: 1965–1990. *British Corrosion Journal*, 25(4):259–270, January 1990. ISSN 0007-0599. doi: 10.1179/000705990799156373.
- [6] Richard Schwarting and Greg Ebel. Manufacturing techniques and process challenges with cg-47 class ship aluminum superstructure modernization and repairs. 2011. URL <https://api.semanticscholar.org/CorpusID:109985385>.
- [7] Timothy L. Burnett, N.J. Henry Holroyd, Geoffrey M. Scamans, Xiaorong Zhou, George E. Thompson, and Philip J. Withers. The role of crack branching in stress corrosion cracking of aluminium alloys. *Corrosion Reviews*, 33(6):443–454, November 2015. ISSN 2191-0316, 0334-6005. doi: 10.1515/corrrev-2015-0050.

- [8] Sudhanshu S. Singh, Jason J. Williams, X. Xiao, F. De Carlo, and N. Chawla. In Situ Three Dimensional (3D) X-Ray Synchrotron Tomography of Corrosion Fatigue in Al7075 Alloy. In T. S. Srivatsan, M. Ashraf Imam, and R. Srinivasan, editors, *Fatigue of Materials II*, pages 17–25. Springer International Publishing, Cham, 2013. ISBN 978-3-319-48583-6 978-3-319-48105-0. doi: 10.1007/978-3-319-48105-0\_2.
- [9] Fabian Eckermann, Thomas Suter, Peter J. Uggowitzer, Andreas Afseth, Marco Stampanoni, Federica Marone, and Patrik Schmutz. In Situ Microtomographically Monitored and Electrochemically Controlled Corrosion Initiation and Propagation in AlMgSi Alloy AA6016. *Journal of The Electrochemical Society*, 156(1):C1, 2009. ISSN 00134651. doi: 10.1149/1.2996269.
- [10] Chunpeng Wang, Ullrich Steiner, and Alessandro Sepe. Synchrotron Big Data Science. *Small*, 14(46):1802291, November 2018. ISSN 1613-6810, 1613-6829. doi: 10.1002/smll.201802291.
- [11] Francisca G. Caballero, editor. *Encyclopedia of Materials: Metals and Alloys*. Elsevier, Amsterdam, 2022. ISBN 978-0-12-819733-2.
- [12] G.M. Scamans, N.J.H. Holroyd, and C.D.S. Tuck. The role of magnesium segregation in the intergranular stress corrosion cracking of aluminium alloys. *Corrosion Science*, 27(4):329–347, January 1987. ISSN 0010938X. doi: 10.1016/0010-938X(87)90076-X.
- [13] R. Zhang, S.P. Knight, R.L. Holtz, R. Goswami, C.H.J. Davies, and N. Birbilis. A Survey of Sensitization in 5xxx Series Aluminum Alloys. *Corrosion*, 72(2): 144–159, February 2016. ISSN 0010-9312, 1938-159X. doi: 10.5006/1787.
- [14] J. L. Searles, P. I. Gouma, and R. G. Buchheit. Stress corrosion cracking of sensitized AA5083 (Al-4.5Mg-1.0Mn). *Metallurgical and Materials Transactions A*, 32(11):2859–2867, November 2001. ISSN 1073-5623, 1543-1940. doi: 10.1007/s11661-001-1036-3.
- [15] Alison J. Davenport, Yudie Yuan, Rajan Ambat, Brian J. Connolly, Martin Strangwood, Andreas Afseth, and G.M. Scamans. Intergranular Corrosion and Stress Corrosion Cracking of Sensitised AA5182. In *Materials Science Forum*, volume 519–521, pages 641–646. Trans Tech Publications Ltd., Stafa, July



2006. ISBN 978-0-87849-408-8. doi: 10.4028/0-87849-408-1.641.
- [16] Ramasis Goswami and Ronald L. Holtz. Transmission Electron Microscopic Investigations of Grain Boundary Beta Phase Precipitation in Al 5083 Aged at 373 K (100 °C). *Metallurgical and Materials Transactions A*, 44(3):1279–1289, March 2013. ISSN 1073-5623, 1543-1940. doi: 10.1007/s11661-012-1166-9.
- [17] Young-Ki Yang and Todd Allen. Direct visualization of  $\beta$  phase causing intergranular forms of corrosion in Al–Mg alloys. *Materials Characterization*, 80: 76–85, June 2013. ISSN 10445803. doi: 10.1016/j.matchar.2013.03.014.
- [18] R. H. Jones, J. S. Vetrano, and C. F. Windisch. Stress Corrosion Cracking of Al–Mg and Mg–Al Alloys. *CORROSION*, 60(12):1144–1154, December 2004. ISSN 0010-9312, 1938-159X. doi: 10.5006/1.3299228.
- [19] Cortney Crane, Robert Kelly, and Richard Gangloff. Crack Chemistry Control of Intergranular SCC in Sensitized Al–Mg. *Corrosion*, page 1852, November 2015. ISSN 0010-9312, 1938-159X. doi: 10.5006/1852.
- [20] Cortney Crane and Rick Gangloff. Stress Corrosion Cracking of Al–Mg Alloy 5083 Sensitized at Low Temperature. *Corrosion*, page 150819163648003, August 2015. ISSN 0010-9312. doi: 10.5006/1766.
- [21] P.J. Steiner and J.T. Burns. Mechanistic Studies of Intergranular Stress Corrosion Cracking in Al–Mg Alloys under Atmospheric Exposure Conditions. *Corrosion*, 74(10):1117–1131, October 2018. ISSN 0010-9312, 1938-159X. doi: 10.5006/2853.
- [22] Roger W. Staehle. Quantitative Micro-Nano (QMN) Approach to SCC Mechanism and Prediction-Starting a Third Meeting. In Jeremy T. Busby, Gabriel Ilevbare, and Peter L. Andresen, editors, *Proceedings of the 15th International Conference on Environmental Degradation of Materials in Nuclear Power Systems — Water Reactors*, pages 1535–1629. Springer International Publishing, Cham, 2011. ISBN 978-3-319-48760-1. doi: 10.1007/978-3-319-48760-1\_95.
- [23] A.C. Thompson, J. Llacer, L. Campbell Finman, E.B. Hughes, J.N. Otis, S. Wilson, and H.D. Zeman. Computed tomography using synchrotron radiation. *Nuclear Instruments and Methods in Physics Research*, 222(1-2):319–323, May 1984. ISSN 01675087. doi: 10.1016/0167-5087(84)90550-7.

- [24] Tyler J. Stannard, Jason J. Williams, Sudhanshu S. Singh, Arun Sundar Sundaram Singaravelu, Xianghui Xiao, and Nikhilesh Chawla. 3D time-resolved observations of corrosion and corrosion-fatigue crack initiation and growth in peak-aged Al 7075 using synchrotron X-ray tomography. *Corrosion Science*, 138:340–352, July 2018. ISSN 0010938X. doi: 10.1016/j.corsci.2018.04.029.
- [25] Sudhanshu S. Singh, Jason J. Williams, Tyler J. Stannard, Xianghui Xiao, Francesco De Carlo, and Nikhilesh Chawla. Measurement of localized corrosion rates at inclusion particles in AA7075 by in situ three dimensional (3D) X-ray synchrotron tomography. *Corrosion Science*, 104:330–335, March 2016. ISSN 0010938X. doi: 10.1016/j.corsci.2015.12.027.
- [26] Claudia Draxl and Matthias Scheffler. NOMAD: The FAIR concept for big data-driven materials science. *MRS Bulletin*, 43(9):676–682, September 2018. ISSN 0883-7694, 1938-1425. doi: 10.1557/mrs.2018.208.
- [27] Karl Pearson. LIII. *On lines and planes of closest fit to systems of points in space.* *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, November 1901. ISSN 1941-5982, 1941-5990. doi: 10.1080/14786440109462720.
- [28] Laurens van der Maaten and Geoffrey Hinton. Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008. ISSN 1533-7928.
- [29] Leland McInnes, John Healy, and James Melville. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, September 2020.
- [30] S. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, March 1982. ISSN 1557-9654. doi: 10.1109/TIT.1982.1056489.
- [31] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD’96, pages 226–231, Portland, Oregon, August 1996. AAAI Press.

- [32] Leland McInnes, John Healy, and Steve Astels. Hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software*, 2(11):205, March 2017. ISSN 2475-9066. doi: 10.21105/joss.00205.
- [33] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, October 1986. ISSN 0028-0836, 1476-4687. doi: 10.1038/323533a0.
- [34] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, May 2015. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature14539.
- [35] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is All you Need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.
- [36] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation, May 2015.
- [37] Tete Xiao, Yingcheng Liu, Bolei Zhou, Yuning Jiang, and Jian Sun. Unified Perceptual Parsing for Scene Understanding, July 2018.
- [38] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M. Alvarez, and Ping Luo. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers, October 2021.
- [39] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition, April 2015.
- [40] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition, December 2015.
- [41] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, and Enhua Wu. Squeeze-and-Excitation Networks, May 2019.
- [42] François Chollet. Xception: Deep Learning with Depthwise Separable Convolutions, April 2017.
- [43] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning Transferable Visual Models

- From Natural Language Supervision, February 2021.
- [44] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric, April 2018.
  - [45] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.
  - [46] Guy Hacohen, Avihu Dekel, and Daphna Weinshall. Active Learning on a Budget: Opposite Strategies Suit High and Low Budgets, June 2022.
  - [47] Ofer Yehuda, Avihu Dekel, Guy Hacohen, and Daphna Weinshall. Active Learning Through a Covering Lens, December 2022.
  - [48] Paul Doucet, Benjamin Estermann, Till Aczel, and Roger Wattenhofer. Bridging Diversity and Uncertainty in Active learning with Self-Supervised Pre-Training, March 2024.
  - [49] Jieyu Zhang, Cheng-Yu Hsieh, Yue Yu, Chao Zhang, and Alexander Ratner. A Survey on Programmatic Weak Supervision, February 2022.
  - [50] Kristen J. Hernandez, Thomas G. Ciardi, Rachel Yamamoto, Mingjian Lu, Arafath Nihar, Jayvic Cristian Jimenez, Pawan K. Tripathi, Brian Giera, Jean-Baptiste Forien, John J. Lewandowski, Roger H. French, and Laura S. Bruckman. L-PBF High-Throughput Data Pipeline Approach for Multi-modal Integration. *Integrating Materials and Manufacturing Innovation*, 13(3): 758–772, September 2024. ISSN 2193-9764, 2193-9772. doi: 10.1007/s40192-024-00368-0.
  - [51] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, June 2009. doi: 10.1109/CVPR.2009.5206848.
  - [52] Joshua Stuckner, Bryan Harder, and Timothy M. Smith. Microstructure segmentation with deep learning encoders pre-trained on a large microscopy dataset. *npj Computational Materials*, 8(1):200, September 2022. ISSN 2057-3960. doi: 10.1038/s41524-022-00878-5.

- [53] Nizam Ud Din and Ji Yu. Training a deep learning model for single-cell segmentation without manual annotation. *Scientific Reports*, 11(1):23995, December 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-03299-4.
- [54] Prem Shrestha, Nicholas Kuang, and Ji Yu. Efficient end-to-end learning for cell segmentation with machine generated weak annotations. *Communications Biology*, 6(1):232, March 2023. ISSN 2399-3642. doi: 10.1038/s42003-023-04608-5.
- [55] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, August 2021. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-021-03819-2.
- [56] Thorsten Wagner, Felipe Merino, Markus Stabrin, Toshio Moriya, Claudia Antoni, Amir Apelbaum, Philine Hagel, Oleg Sitsel, Tobias Raisch, Daniel Prumbaum, Dennis Quentin, Daniel Roderer, Sebastian Tacke, Birte Siebolds, Evelyn Schubert, Tanvir R. Shaikh, Pascal Lill, Christos Gatsogiannis, and Stefan Raunser. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Communications Biology*, 2(1):218, June 2019. ISSN 2399-3642. doi: 10.1038/s42003-019-0437-z.
- [57] David Rolnick, Priya L. Donti, Lynn H. Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, Nikola Milojevic-Dupont, Natasha Jaques, Anna Waldman-Brown, Alexandra Luccioni, Tegan Maharaj, Evan D. Sherwin, S. Karthik Mukkavilli, Konrad P. Kording, Carla Gomes, Andrew Y. Ng, Demis Hassabis, John C. Platt, Felix Creutzig, Jennifer Chayes, and Yoshua Bengio. Tackling Climate Change with Machine Learning, November 2019.

- [58] Tekin Bicer, Doga Gursoy, Rajkumar Kettimuthu, Ian T. Foster, Bin Ren, Vincent De Andrede, and Francesco De Carlo. Real-Time Data Analysis and Autonomous Steering of Synchrotron Light Source Experiments. In *2017 IEEE 13th International Conference on E-Science (e-Science)*, pages 59–68, Auckland, October 2017. IEEE. ISBN 978-1-5386-2686-3. doi: 10.1109/eScience.2017.53.
- [59] Ankit Agrawal, Parijat D Deshpande, Ahmet Cecen, Gautham P Basavarsu, Alok N Choudhary, and Surya R Kalidindi. Exploration of data science techniques to predict fatigue strength of steel from composition and processing parameters. *Integrating Materials and Manufacturing Innovation*, 3(1):90–108, December 2014. ISSN 2193-9764, 2193-9772. doi: 10.1186/2193-9772-3-8.
- [60] Sanam Gorgannejad, Aiden A. Martin, Jenny W. Nicolino, Maria Strantz, Gabriel M. Guss, Saad Khairallah, Jean-Baptiste Forien, Vivek Thampy, Sen Liu, Peiyu Quan, Christopher J. Tassone, and Nicholas P. Calta. Localized key-hole pore prediction during laser powder bed fusion via multimodal process monitoring and X-ray radiography. *Additive Manufacturing*, 78:103810, September 2023. ISSN 22148604. doi: 10.1016/j.addma.2023.103810.
- [61] Anuj Karpatne, Gowtham Atluri, James H. Faghmous, Michael Steinbach, Arindam Banerjee, Auroop Ganguly, Shashi Shekhar, Nagiza Samatova, and Vipin Kumar. Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data. *IEEE Transactions on Knowledge and Data Engineering*, 29(10):2318–2331, October 2017. ISSN 1558-2191. doi: 10.1109/TKDE.2017.2720168.
- [62] Anuj Karpatne, Ramakrishnan Kannan, and Vipin Kumar, editors. *Knowledge Guided Machine Learning: Accelerating Discovery Using Scientific Knowledge and Data*. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series. CRC Press, Boca Raton, first edition edition, 2023. ISBN 978-1-003-14337-6.

- [63] Tobias Strohmann, Katrin Bugelnig, Eric Breitbarth, Fabian Wilde, Thomas Steffens, Holger Germann, and Guillermo Requena. Semantic segmentation of synchrotron tomography of multiphase Al-Si alloys using a convolutional neural network with a pixel-wise weighted loss function. *Scientific Reports*, 9(1):19611, December 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-56008-7.
- [64] Radmir Karamov, Christian Breite, Stepan V. Lomov, Ivan Sergeichev, and Yentl Swolfs. Super-Resolution Processing of Synchrotron CT Images for Automated Fibre Break Analysis of Unidirectional Composites. *Polymers*, 15(9): 2206, May 2023. doi: 10.3390/polym15092206.
- [65] Allard A. Hendriksen, Minna Bühner, Laura Leone, Marco Merlini, Nicola Vigano, Daniël M. Pelt, Federica Marone, Marco Di Michiel, and K. Joost Batenburg. Deep denoising for multi-dimensional synchrotron X-ray tomography without high-quality reference data. *Scientific Reports*, 11(1):11895, June 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-91084-8.
- [66] Seungjoo Shin, Min Woo Kim, Kyong Hwan Jin, Kwang Moo Yi, Yoshiki Kohmura, Tetsuya Ishikawa, Jung Ho Je, and Jaesik Park. Deep 3D reconstruction of synchrotron X-ray computed tomography for intact lungs. *Scientific Reports*, 13(1):1738, January 2023. ISSN 2045-2322. doi: 10.1038/s41598-023-27627-y.
- [67] G01 Committee. Test Method for Determining the Susceptibility to Intergranular Corrosion of 5XXX Series Aluminum Alloys by Mass Loss After Exposure to Nitric Acid (NAML Test).
- [68] Mohsen Seifi, Henry Holroyd, and John Lewandowski. Deformation Rate and Sensitization Effects on Environmentally Assisted Cracking of Al-Mg Naval alloys. *Corrosion*, page 1949, December 2015. ISSN 0010-9312, 1938-159X. doi: 10.5006/1949.
- [69] G. M. Ugiansky, L. P. Skolnick, and S. W. Stiefel. Directional Effects in the Stress Corrosion Cracking of an Aluminum Alloy. *CORROSION*, 25(2):77–86, February 1969. ISSN 0010-9312, 1938-159X. doi: 10.5006/0010-9312-25.2.77.

- [70] Mark Rivers. Synchrotron ‘pink beam’ tomography for the study of dynamic processes. *SPIE Newsroom*, August 2016. ISSN 18182259. doi: 10.1117/2.1201608.006674.
- [71] Dog̃a Gürsoy, Francesco De Carlo, Xianghui Xiao, and Chris Jacobsen. TomoPy: A framework for the analysis of synchrotron tomographic data. *Journal of Synchrotron Radiation*, 21(5):1188–1193, September 2014. ISSN 1600-5775. doi: 10.1107/S1600577514013939.
- [72] Felix Buchert, Nassir Navab, and Seong Tae Kim. Toward Label-Efficient Neural Network Training: Diversity-Based Sampling in Semi-Supervised Active Learning. *IEEE Access*, 11:5193–5205, 2023. ISSN 2169-3536. doi: 10.1109/ACCESS.2023.3236529.
- [73] Sebastian Raschka, Joshua Patterson, and Corey Nolet. Machine Learning in Python: Main developments and technology trends in data science, machine learning, and artificial intelligence. *arXiv preprint arXiv:2002.04803*, 2020.
- [74] Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, November 1987. ISSN 0377-0427. doi: 10.1016/0377-0427(87)90125-7.
- [75] David L. Davies and Donald W. Bouldin. A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2):224–227, April 1979. ISSN 1939-3539. doi: 10.1109/TPAMI.1979.4766909.
- [76] Joseph B. Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, 7(1):48–50, February 1956. ISSN 0002-9939, 1088-6826. doi: 10.1090/S0002-9939-1956-0078686-7.
- [77] Nicki Detlefsen, Jiri Borovec, Justus Schock, Ananya Jha, Teddy Koker, Luca Di Liello, Daniel Stancl, Changsheng Quan, Maxim Grechkin, and William Falcon. TorchMetrics - Measuring Reproducibility in PyTorch. *Journal of Open Source Software*, 7(70):4101, February 2022. ISSN 2475-9066. doi: 10.21105/joss.04101.



- [78] Maxim Tkachenko, Mikhail Malyuk, Andrey Holmanyuk, and Nikolai Liubimov. Label Studio: Data labeling software, 2020.
- [79] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library, December 2019.
- [80] Pavel Iakubovskii. Segmentation Models Pytorch, 2019.
- [81] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization, January 2017.
- [82] Ilya Loshchilov and Frank Hutter. SGDR: Stochastic Gradient Descent with Warm Restarts, May 2017.