

# Benchmarking Automated MIC Detection: The AI-Ready Space Biology SEM Dataset and Advanced Detection Methods

Tanzina Akter Tani<sup>1\*</sup>, Amber D Busboom<sup>1</sup>, Robert McLean<sup>1</sup>, Jelena Tešić<sup>1\*</sup>

<sup>1</sup>Texas State University, San Marcos, 78666, TX, U.S.A.

\*Corresponding author(s). E-mail(s): [tanzinatani123@gmail.com](mailto:tanzinatani123@gmail.com); [jtesic@txstate.edu](mailto:jtesic@txstate.edu);

Contributing authors: [abusboom@txstate.edu](mailto:abusboom@txstate.edu); [mclean@txstate.edu](mailto:mclean@txstate.edu);

## Abstract

Microbial-induced corrosion (MIC) severely threatens the structural integrity of metals in high-stakes environments, from aerospace to terrestrial infrastructure. Rapid, reliable detection is essential for informed mitigation and maintenance. We present a computer vision pipeline that sets a new benchmark in MIC region segmentation within scanning electron microscope (SEM) imagery. Central to this advancement is our expanded, expertly annotated dataset of 331 SEM images of MIC on stainless steel, the largest and first AI-ready segmentation dataset to include spaceflight samples to date. Leveraging this resource, we rigorously benchmark both classical and deep learning methods. Our classical Cluster with Location Features (CLF) method achieves a Dice score of 44.1%, nearly 10% higher than prior baselines, owing to improved annotations. We introduce two deep learning architectures: an enhanced SAM2 and a novel Prompt- and Heatmap-Guided FPN-based Lightweight Segmentation Model (Lightweight PH-FPNSeg), both engineered for challenging MIC contours. Both models deliver state-of-the-art results, achieving average Dice and IoU scores of 82% and 70%, respectively—substantially surpassing prior approaches. The study establishes a robust pipeline for precise, automated MIC detection, laying a foundation for advanced corrosion research, proactive asset management, and improved materials science across industry and research settings.

**Keywords:** Microbial-induced corrosion, MIC, NASA, Space, Traditional baseline, DeepLabv3+, Segment Anything Model, Prompt-based Model

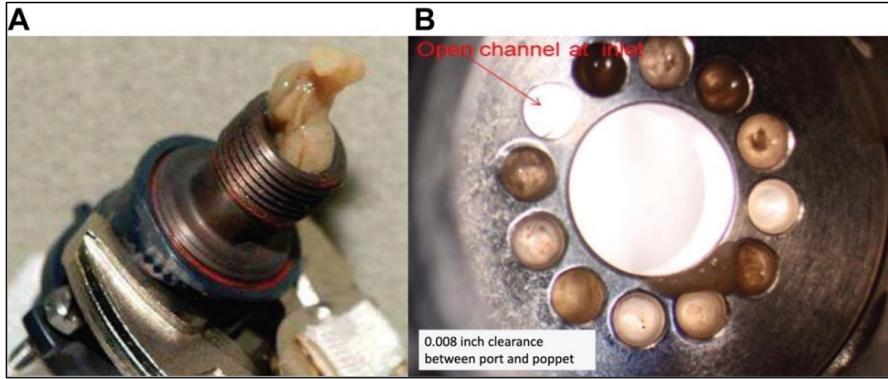
## 1 Introduction

Microbial damage refers to the harmful effects that microorganisms can inflict on a variety of surfaces, including metals, minerals, biological tissues, and medical equipment. Among the most significant mechanisms of such damage is the formation of biofilms. Microorganisms construct biofilms as three-dimensional communities, consisting of cells, water, and sticky substances known as extracellular polymeric substances (EPS) [1]. These complex structures allow microbes to survive longer and withstand environmental stresses, presenting substantial challenges in medical, industrial, and ecological contexts [2].

A significant consequence of biofilm formation is microbially induced corrosion (MIC). MIC typically manifests as localized pitting corrosion, where small, deep cavities form beneath biofilms on metal surfaces. Chemical changes within the biofilm drive this process—microorganisms release metabolic byproducts such as sulfides, organic acids, or manganese oxides, which reduce local pH and alter redox potential. Such changes disrupt the protective oxide layer of metals, accelerating corrosion at specific sites [3, 4]. The presence of MIC beneath biofilms produced by diverse microorganisms—including marine *Vibrio* and heat-resistant *Bacillus* and *Geobacillus*—underscores its wide-ranging implications for both industry and the environment [5, 6].

Furthermore, the challenges posed by biofilms extend well beyond metal corrosion. In water and wastewater systems, biofilm formation, fouling, and MIC persistently complicate disinfection and removal efforts, as biofilms protect microbes from external interventions [7]. This resistance leads to chronic contamination, increased maintenance costs, and elevated health risks due to biofilm-protected pathogens on storage tanks, medical devices, and instrumentation. In spaceflight environments, managing biofilms becomes even more critical. For example, the Waste Recycling System (WRS) aboard the International Space Station (ISS) collects and purifies waste liquids—including urine and condensed humidity—transforming them into potable water via distillation and disinfection [8]. In the microgravity conditions of space, limited fluid movement further encourages biofilm growth. Resulting biofilms in the WRS can cause corrosion, clog filters, and damage essential spacecraft systems, threatening mission reliability and astronaut health [9, 10]. Figure 1 shows examples of biofilms in bacterial fluids before disinfection [9].

Given these wide-ranging challenges, robust analysis and monitoring of biofilms and MIC are essential. Advanced image segmentation, especially when applied to microscopy images such as those obtained with scanning electron microscopy (SEM) or confocal laser scanning microscopy, has become a critical tool [12]. Effective segmentation enables precise identification of microbial cells, clusters, corrosion byproducts, and surface coverage, reducing quantification errors and providing deeper insight into biofilm heterogeneity. Researchers particularly rely on this approach for MIC investigations, where corrosion pits are subtle, highly localized, and morphologically irregular [13, 14]. Segmentation also provides quantitative metrics, including problems, density, and distribution. These metrics are central to corrosion risk assessment and mitigation planning across industries and spaceflight applications [15].



**Fig. 1** International Space Station (ISS) Water Recovery System (WRS) components that needed to be replaced due to microbial-induced corrosion (MIC): (A) water condensate line and (B) on a solenoid valve [11].

The growing complexity of imaging data has accelerated the adoption of automated, AI-based analysis. Today, AI plays a pivotal role in health, biology, and industry by automating analytical tasks and supporting more informed decision-making. However, reliable AI models require high-quality, well-curated, and contextually relevant data [16]. Preparation of such data depends on close collaboration between technical experts and domain specialists to avoid such problems as poor formatting, missing metadata, class imbalance, and mislabeled annotations, all of which can compromise model performance [17]. As a result, data preparation must be treated as a rigorous, iterative engineering process, continuously refined to meet scientific goals [18].

Researchers increasingly use automated image segmentation to detect MIC features and assess biofilm removal. The subtlety and variability of MIC features require expert annotation to generate high-quality AI training data. Inconsistent or inaccurate labeling can substantially impede model accuracy. Recognizing these challenges, our DataLab computer science team [19] and the McLean biology team [20] worked closely together to develop an MIC dataset that is both technically rigorous and biologically precise, designed explicitly for robust automated analysis.

Our study is a substantial extension of previous work, which systematically compared classical and deep learning methods for MIC segmentation on stainless steel samples collected from both microgravity (flight) and terrestrial (ground-control) conditions. However, the earlier study [21] was limited by a small dataset and insufficient methodological detail, reducing its impact. In this work, we address these limitations by introducing the largest—and, to our knowledge, the first—AI-ready segmentation dataset for MIC analysis that includes samples from spaceflight environments, comprising 331 expertly annotated SEM images. Importantly, we revisit the classical and deep learning segmentation approaches developed in our earlier work, re-evaluating and fine-tuning them using the new, expanded dataset. This process led to further improvements in segmentation accuracy and reliability, as the models benefited from enhanced data diversity and robust annotation. By systematically re-benchmarking these established methods, we provide a transparent and rigorous

baseline for subsequent advances. We further expand the methodological scope by integrating two advanced segmentation architectures: an enhanced Segment Anything Model 2 (SAM2) and the newly designed Prompt- and Heatmap-Guided FPN-based Lightweight Segmentation Model (Lightweight PH-FPNSeg). Both architectures are specifically optimized for the detection of MIC-affected regions in challenging SEM imagery, addressing the subtleties and irregularities inherent in MIC features. The integration of these modern approaches with improved baselines leads to significant improvements in segmentation accuracy, reproducibility, and quantitative rigor.

In summary, the present study establishes new benchmarks for automated MIC analysis in both spaceflight and terrestrial contexts by: *(1) providing the most comprehensive and specialized MIC segmentation dataset to date; (2) delivering a systematic, side-by-side comparison and optimization of classical and deep learning approaches on this dataset; and (3) further advancing automated image segmentation through the implementation of both an enhanced and a newly designed deep learning architecture.* Collectively, these contributions establish a robust, extensible platform for MIC segmentation that standardizes evaluation, enables scalable and reliable quantification, and informs the design of more effective mitigation strategies.

The supplementary code can be found in the GitHub repository as *SpaceMIC* [22].

## 2 Related Works

In this section, we review prior research on automated image segmentation for microbial-induced corrosion (MIC) detection, covering studies on biofilms and corrosion across diverse materials, imaging modalities, and computational approaches.

A super-resolution (DLSR) technique [23] has been used to enhance SEM biofilm image segmentation on material surfaces. It contributed to quantifying biofilm structure as the segmentation model, a fully convolutional network (FCN with a pretrained ResNet50 backbone), showed improvement. Another study [24] focuses on the biofilms produced by *Desulfovibrio alaskensis* G20—a sulfate-reducing bacterium—that were grown on mild steel. They collected a total of 77 SEM images for analysis. As for the automated system, a self-supervised learning with SimSiam and a ResNet encoder has been experimented with to detect MIC in biofilm images effectively. Microbial corrosion on AISI 304 stainless steel was also analyzed using SEM micrographs [25]. Corrosion tests exposed metal samples to artificial seawater and Stankey medium with salt marsh sediments to introduce bacteria. The OpenCV method measured biofilm number, size, and coverage, enabling detailed evaluation of material degradation and growth under different conditions. Another form of corrosion, caused by biofilm or rust on metal surfaces such as tanks, pipelines, bridges, and hulls, has been analyzed using an automated segmentation system [26]. For this, a comprehensive metal corrosion image dataset from the field and the internet, covering various corrosion types and severities, was collected. The HEU-Net deep learning model then identified and quantified corroded areas accurately. Another study used high-quality lab images to monitor and predict rust and corrosion on weathering steel [27]. The dataset includes images at various rust stages with extracted texture, color, and structural features. Segmentation methods like Random Forest, LightGBM, XGBoost, and

Gradient Boosting Decision Trees evaluated rust thickness loss and rust layer types. Corrosion on archaeological artifacts before restoration has also been studied to create an automated system [28]. It focuses on automatically segmenting rusted regions from stereo microscope images of iron objects, aiding faster and more objective assessment for restorers and archaeologists. Experiment with the U-Net segmentation model, the study highlighting its potential for non-destructive and efficient corrosion evaluation in cultural heritage conservation. Another study [29] created a new corrosion dataset by imaging SS304L stainless steel samples after controlled corrosion tests, selecting 84 high-quality SEM images from different regions of the structural alloy, and manually annotating corrosion areas at the pixel level with expert review. The dataset was combined with existing public datasets to train the YOLOv9-based segmentation model. Also, A different study used UAV images to detect corrosion on steel structures automatically [30]. The dataset includes 331 high-resolution UAV images of Tianjin steel surfaces, segmented into super-pixel regions and labeled for corrosion. The FPN and PANet models were employed to perform semantic segmentation. To segment and quantify bacterial cells in SEM images of biofilms, especially when cells overlap or touch, a study created a dataset with about 290 SEM images of *Desulfovibrio alaskensis* G20 biofilms, which were manually annotated pixel by pixel [1]. The approach uses a U-Net for segmentation and ellipse fitting to separate overlapping cells and extract properties like area and cell count. Similar work of bacterial cells in *Desulfovibrio alaskensis* G20 biofilms was used for automated segmentation and measurement [31]. 66 SEM images of biofilms on steel and copper at various growth stages were collected. The BiofilmScanner used Yolact for segmentation and moment invariants for geometric features, with preprocessing like CLAHE and augmentation. Another study used SEM images of microbial biofilms formed by *Oleidesulfovibrio alaskensis* (OA-G20) on graphene-nickel surfaces [32]. They collected the dataset by incubating graphene-nickel specimens with OA-G20 for 30 days and then capturing SEM images. As a deep learning segmentation model, TransUNet has achieved the highest result. High-resolution SEM images captured the biofilm growth of *Pseudomonas veronii*, *Acinetobacter baumannii*, and *Arthrobacter halodurans* on polymer supports in LB medium at various stages [33]. They manually annotated 100 images to identify individual bacterial cells and background regions. Also, VAEs, GANs, diffusion models, and CycleGAN are used to generate large, realistic synthetic datasets with masks, which facilitated the robust training of the segmentation Mask R-CNN. Another study related to MIC mechanisms in carbon steel was conducted by exposing coupons to microbial consortia and controls, and analyzing them after 30 days via X-ray micro-CT, SEM/EDS, FTIR, Raman, XRD, and 16S rRNA sequencing [34]. Machine learning methods, including 3D imaging with watershed segmentation and thresholding, were used for pore analysis to assess corrosion, phases, and porosity in both biotic and abiotic conditions.

Overall, most previous studies on biofilm formation and MIC have focused on terrestrial industrial environments, such as pipelines and water systems, using SEM imagery from laboratory or field samples. Recognizing the specific challenges of microbial contamination in spacecraft, our earlier research [21] filled this gap by analyzing stainless steel coupons exposed to mixed bacterial cultures in spaceflight missions

(SpaceX-21) and their ground controls. The whole process resulted in an extensive collection of SEM images, with 154 images carefully annotated for accurate segmentation. We improved the field by applying both traditional and deep learning segmentation models, including our CLF baseline, DeepLabv3+, and the Segment Anything Model (SAM). Building on this work, our current research aims to expand and refine the dataset and automated analysis of MIC segmentation, paving the way for broader, more robust applications in corrosion monitoring and biofilm studies for both space and terrestrial environments.

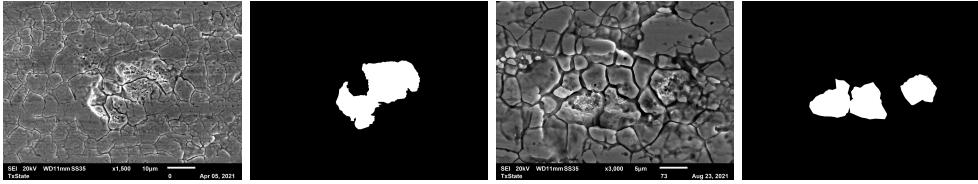
### 3 Dataset acquisition

The NASA Bacterial Adhesion and Corrosion (BAC) team conducted two recent studies aboard the SpaceX-21 (2020) and SpaceX-29 (2023) missions to examine biofilm formation and its potential role in microbially induced corrosion (MIC) in the spaceflight environment [35]. They investigated the ability of *Escherichia coli* and *Pseudomonas aeruginosa* to form biofilms and cause corrosion on stainless steel coupons under both microgravity and full-gravity conditions.

The McLean team [20] carried out Scanning Electron Microscopy (SEM) analysis on 316 stainless steel coupons from both the SpaceX-21 flight and ground control, incubating the coupons under identical conditions. They placed these coupons in the 12-well BioCell flight hardware [36]. Culture preparations for both missions were completed at Kennedy Space Center (Florida) immediately before launch and inoculated by the ISS crew during orbit. Asynchronous ground controls were performed at ASU a few hours after the spaceflight inoculations, allowing accurate replication of the experiment based on astronaut communications. The experimental components were delivered to the International Space Station (ISS) aboard SpaceX-21. At pre-scheduled intervals, selected wells were inoculated with a mixed bacterial culture by the ISS crew. The same setup and procedures were used for the ground controls.

When the flight and ground-control samples returned for SEM analysis, the McLean team [20] first processed them for confocal microscopy, keeping the BioCells intact. Then, they extracted the stainless steel coupons from the BioCells and chemically washed them to remove biofilms in preparation for SEM imaging. These coupons were mounted onto SEM stubs and imaged under vacuum, starting with low magnification ( $50\times$ ) for general scanning, followed by higher magnifications (typically  $1,500\times$ – $3,000\times$ ) to visualize bacterial structures and corrosion features. During SEM imaging, the electron beam excites atoms in the sample, which emit X-ray energy characteristic of specific elements. The initial SpaceX-21 flight experiment generated thousands of SEM images, captured using a JEOL SEM system operated under both low and high vacuum settings [20].

The complete dataset consists of 9,415 TIFF-format SEM images, each accompanied by metadata. Image characteristics vary based on parameters such as voltage, magnification, resolution, and imaging mode. To define a consistent baseline for analysis, we selected the most overlapping subset, reducing the dataset to 4,573 images. To establish ground truth for corrosion versus background segmentation, a domain expert used Label Studio to thoroughly annotate 154 images at the pixel level in the



**Fig. 2** Two sample images and their corresponding ground truth from our dataset.

prior study [21]. However, those previous annotations had limitations, such as missing some MIC regions and including chemically corroded areas where microorganisms were not present. To address the issue, the McLean team performed a new round of expert annotation, ensuring that only regions involving microorganism corrosion are labeled. The team meticulously re-annotated the data, producing 331 high-quality ground truth masks focused solely on corrosion caused by microbes. Among the annotated samples, 246 originate from microgravity (flight) conditions, and 85 come from terrestrial (ground-control) environments. The corresponding images and masks have resolutions of  $1280 \times 960$  and  $2560 \times 1920$  pixels, respectively. Representative samples of these images, along with their corresponding annotations, are shown in Figure 2.

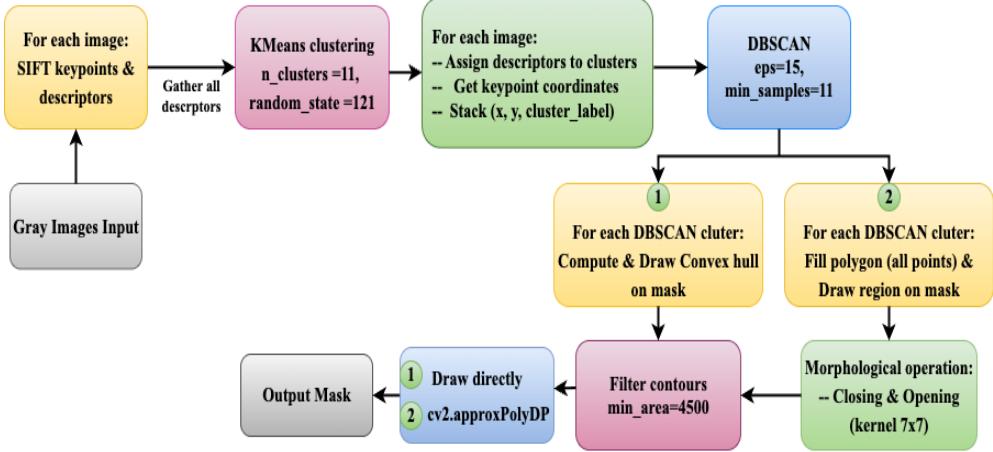
## 4 Methodology

The following section presents a comprehensive overview of both classical and deep learning methods employed in this study.

### 4.1 Classical Non-DL Technique

In our previous study [21], the Cluster with Local Features (CLF) algorithm was introduced only briefly. In this work, we apply CLF to the expanded dataset of 331 samples and provide a detailed exposition of the method, including a step-by-step flow chart highlighting its two best-performing variants.

In the proposed traditional CLF algorithm, we converted the input images to grayscale and processed them using the *SIFT (Scale-Invariant Feature Transform)* algorithm [37] to extract keypoints and descriptors for each image. SIFT was selected because it is robust to changes in scale and rotation, allowing for the reliable detection of objects regardless of their size or orientation. It helped to find matching local features [38] that are consistent across different views. After applying SIFT, all extracted descriptors were pooled together and clustered using the *K-Means algorithm* [39]. In the algorithm, we used 11 clusters with a random seed of 123. It helped to group similar local patterns or textures into a fixed set of descriptor types. Next, we *labeled descriptors* according to their K-Means cluster, and stacked each keypoint (x, y) coordinates with its cluster label for each image. The integration of spatial and appearance information enabled more meaningful clustering in the following stage. To further refine the segmentation, *DBSCAN* [40] clustering was applied to these combined vectors within each image, with eps set to 15 and a minimum of 11 samples. While K-Means groups features based on visual similarity, DBSCAN clusters those that are spatially contiguous, resulting in meaningful, connected regions.



**Fig. 3** Flow diagram of the proposed non-deep learning baseline algorithm: Cluster with Local Features (CLF).

After DBSCAN, we split the pipeline into two alternative approaches noted as 1 and 2 in the flow diagram, Fig. 3. Both approaches were motivated by different goals in mask construction. In the No. 1 approach, for each cluster identified by DBSCAN, the *convex hull* [41] around the cluster's keypoints was computed and drawn directly on a mask image. The process created mask regions that closely contain the detected feature clusters. In approach No.2, each DBSCAN cluster, the area defined by its keypoints, was filled as a *polygon* on the mask image without requiring convexity. To enhance mask quality, we used *morphological closing and opening* operations with a  $7 \times 7$  kernel [42] to smooth regions, fill small holes, and connect nearby components. After that, for both approaches, we extracted contours from the resulting masks and filtered them to keep only those larger than *4,500 pixels*, which reduced noise and focused on significant regions. In the convex hull approach, we drew the filtered *contours directly*, producing sharp, object-centered masks. In the polygon method, we used *cv2.approxPolyDP* to further simplify contours before being drawn, resulting in smoother and more natural object boundaries. We parameterized the proposed CLF algorithm through a trial-and-error process. As with most classical (non-deep learning) methods, which rely on human-crafted features rather than data-driven model learning, the effectiveness of this approach depends on both the structural characteristics of the images and the chosen parameter values. Consequently, we may need to adjust the parameters for a better fit with different image datasets or tasks.

## 4.2 Deep Learning Techniques

### 4.2.1 DeepLabV3+ Model

We implemented the DeepLabv3+ [43] model in the expanded dataset to evaluate segmentation under greater data diversity and scale. DeepLabv3+ uses an

encoder-decoder structure with a pretrained Xception or ResNet backbone for feature extraction. The encoder employs an Atrous Spatial Pyramid Pooling (ASPP) module to capture multi-scale context, while the decoder refines boundaries by integrating low- and high-level features. In our implementation, we added *dropout layers* and replaced standard convolutions with *modular convolutional blocks* for improved regularization and modularity. The experimented model provided a rigorous baseline for assessing performance gains in the extended dataset.

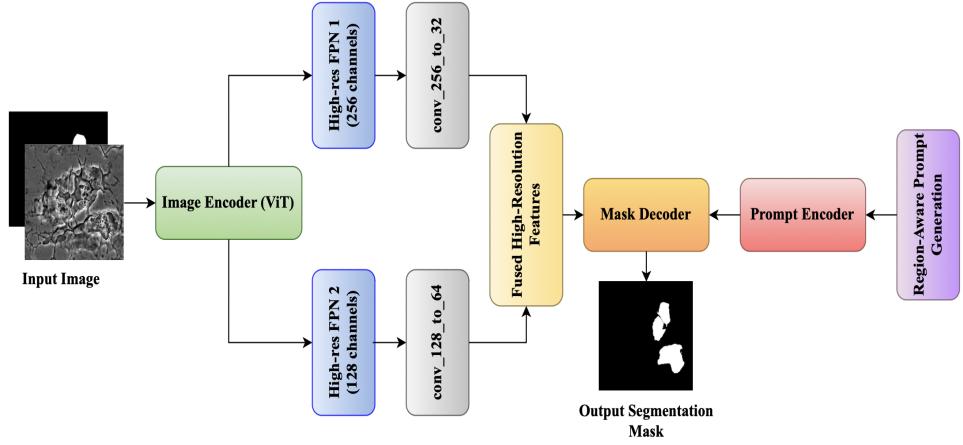
#### 4.2.2 SAM Model

The Segment Anything Model (SAM) [44], introduced in 2023, is a foundation model for prompt-driven, generalizable image segmentation. It accepts guidance such as points, boxes, or masks, enabling zero-shot adaptation to diverse tasks and unseen data. In this study, we used the sam-vit-base variant, which employs a ViT-based encoder pre-trained with masked autoencoders via HuggingFace Transformers. To preserve generalization, we froze the image and prompt encoders, training only the mask decoder for task-specific segmentation. Following SAM’s protocol, box prompts were derived from the smallest enclosing rectangles of ground-truth masks. Coordinates were pre-processed as model inputs with automated code, and small perturbations were optionally introduced during training to enhance robustness. These prompts provide spatial context, ensuring efficient, context-aware segmentation, particularly in cases of uncertainty or overlap.

The Segment Anything Model 2 (SAM2) [45], introduced in 2024, improves upon the first version by providing superior generalization and efficiency in image and video segmentation. Notable enhancements comprise a more robust hierarchical Vision Transformer (ViT) backbone for multi-scale feature extraction, improved prompt encoding strategies, compatibility with multi-type and multi-point prompts, and a lightweight mask decoder optimized for both speed and accuracy. SAM2 extends promptable segmentation into the spatio-temporal domain, demonstrating reliable performance in both static images and video frames.

#### 4.2.3 Enhanced SAM2 Model

We adapted SAM2 for our MIC segmentation, utilizing its promptable architecture that accepts flexible user input through points, boxes, and masks. The model incorporates a ViT-based image encoder for the extraction of multi-scale features, a prompt encoder for embedding user inputs, and a mask decoder that integrates these signals to produce segmentation masks [45][46]. The model diagram is given in Fig. 4. SAM2 model inherently supports multi-point prompting, allowing for the input of various positive and negative user points [46]. In our method, rather than relying on manual or randomly selected points, we introduced a systematic, *region-aware strategy* that automatically generates these prompts: we placed foreground points at the centroids of all connected components in the ground-truth mask and background points from non-object regions. The region-aware prompting strategy provides spatially rich and unambiguous guidance, essential when segmenting images with fragmented, overlapping, or irregularly shaped objects. Also, to enhance segmentation precision, our



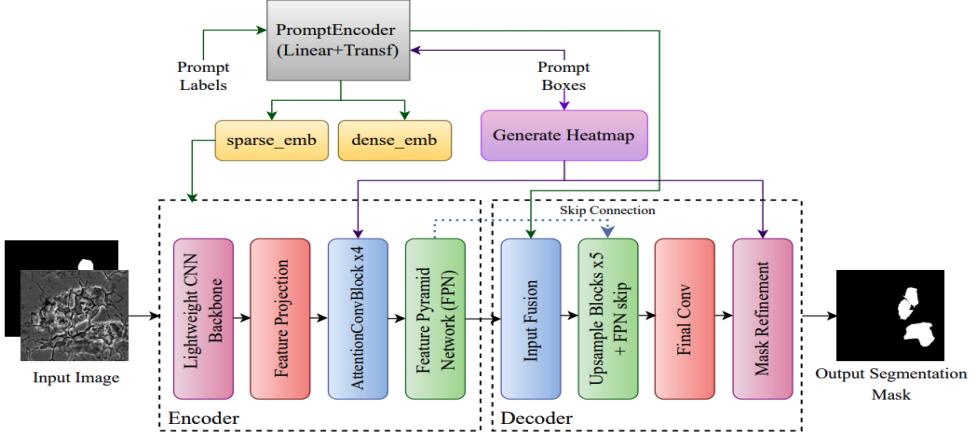
**Fig. 4** Enhanced SAM2 model with region-aware prompt and custom convolutional layers

method added extra  $1 \times 1$  convolutional layers to adapt and fuse high-resolution features from the encoder, specifically using shallow maps from *backbone\_fpn[0]* and *backbone\_fpn[1]*, which retain fine details for boundary detection. To improve compatibility, the process reduces extracted feature maps from 256 to 32 channels and from 128 to 64 channels. Merging these features with the main image features in the mask decoder enabled the model to utilize detailed spatial and semantic information, enhancing edge detection and sensitivity to subtle variations, especially useful when segmenting complex structures, such as in corrosion. The architecture consists of independent, trainable components: the image encoder, prompt encoder, and mask decoder, constructed in a modular fashion. These adaptations leverage the strengths of SAM2 while addressing the specific challenges associated with MIC segmentation, thereby facilitating accurate and detailed identification of complex corrosion structures.

#### 4.2.4 Proposed Lightweight PH-FPNSeg Model

Our proposed segmentation model is a prompt-guided encoder-decoder architecture designed to integrate user input with image context for precise mask generation. It introduces four key modules: a Prompt Encoder for structured user inputs, a Gaussian heatmap generator for spatial priors, a hierarchical image Encoder with stage-wise attention-convolution and feature pyramid refinement, and a multi-stage Decoder that reconstructs detailed segmentation masks. The overall architecture is illustrated in Fig. 5.

The **PromptEncoder** transformed user-provided bounding boxes and class labels into feature representations. Box coordinates and labels were embedded with linear layers, combined, and passed through a lightweight transformer encoder, allowing multiple bounding-box prompts from the same image to interact. The transformer output was pooled into a sparse embedding, a 256-dimensional vector that encodes the overall semantics of the prompts and provides global conditioning. This sparse embedding was used in the encoder to modulate multi-scale features and in the decoder to inject



**Fig. 5** Proposed Lightweight PH-FPNSeg Model comprises four modules: a PromptEncoder, a Heatmap generator, a hierarchical image Encoder and a multi-stage Decoder.

global context. The sparse embedding was further transformed into a *dense embedding*, resized to  $H/32 \times W/32$  to match the encoder’s deepest feature scale, combined with positional encoding for spatial awareness, and finally refined with a  $1 \times 1$  convolution, batch normalization, and ReLU. Unlike the sparse embedding, this dense embedding was only applied in the decoder to provide localized prompt-aware guidance for mask prediction.

The **Gaussian Heatmap** was generated from the centers of user-provided bounding boxes to form a smooth spatial map highlighting regions of interest. In the encoder, it modulated feature maps and was added to the attention logits, directing feature extraction toward regions indicated by the prompts. In the decoder, it was fused with mask uncertainty maps during refinement, allowing the network to fine-tune ambiguous boundaries and produce sharper, more accurate segmentation masks.

The **Image Encoder** integrates four main components: a lightweight convolutional backbone, attention-convolutional refinement blocks, prompt modulation, and a two-path feature pyramid network (FPN). Together, these transform the input into multi-scale, prompt-aware feature representations. The *backbone* is a four-stage CNN that reduces spatial resolution with strides of 4, 8, 16, and 32, while expanding the number of channels to 96, 192, 384, and 768, generating hierarchical features that capture both local detail and global semantics. Each stage *projects* its output to 256 channels via pointwise convolution and adds learned positional encodings resized to match spatial resolution. *Attention-convolutional blocks* then refine these representations, compressing queries and keys to 32 dimensions for efficiency. Gaussian heatmaps gate features before attention, bias the attention logits, and are reapplied after feed-forward refinement. Each block employs a dual-branch feed-forward module, where the low-rank and full-rank branches utilize multilinear projections with base and spline components, followed by depthwise convolution, batch normalization, and ReLU. Their outputs are concatenated, fused with a lightweight multilayer perceptron, projected back to 256 channels, further refined with another depthwise convolution, and stabilized by a scaled residual connection. Prompt modulation applies sparse

embeddings projected to 256 dimensions multiplicatively to the feature maps, enabling context-specific adaptation. Finally, the *two-path FPN* fuses multi-scale features: the top-down pathway propagates semantic context from deep to shallow maps, while the bottom-up pathway reinjects fine detail, with convolution–batch normalization-ReLU layers refining at each stage. The encoder outputs three intermediate feature maps as skip connections, a deepest fused map as the unified vision representation, and positional encodings to guide the decoder.

The **Decoder** began with an *input fusion phase* that merged the encoder’s deepest feature (*vision\_features*) with dense and sparse embeddings from the Prompt Encoder, enriching the input with spatial detail and prompt-specific context. The fused representation was progressively upsampled through five *convolutional blocks*, each doubling spatial resolution while reducing channel dimensionality. At each stage, *skip connections* from intermediate *FPN outputs* restored fine spatial details. A  $1 \times 1$  *convolution* generated the initial segmentation logits, which were further refined by a *lightweight head* that computed edge uncertainty from mask gradients and combined it with the Gaussian heatmap to produce a correction offset, improving accuracy along object boundaries.

Overall, the architecture effectively integrates prompt-aware attention, multi-scale feature refinement, and hierarchical fusion to generate rich, context-aware representations, enabling accurate segmentation performance.

## 5 Experimental Setup

### 5.1 Data Processing

**Cropping:** As illustrated in Fig. 2, the original images contain text at the lower margin, which is irrelevant for segmentation. To standardize the input, we first resized all images to  $1024 \times 1024$  pixels. Then the last 96 rows and columns were cropped from each image and its corresponding mask, resulting in a final dimension of  $928 \times 928$  pixels.

**Resizing:** We used the original image size of  $928 \times 928$  for the proposed baseline CLF algorithm. To lower computational complexity and achieve effective model training, images and ground truth masks used in all deep learning algorithms were scaled to  $512 \times 512$  pixels.

**Dataset Split:** We used all 331 images in the classical method to generate predicted masks and evaluated them against the ground truth. For the deep learning, the dataset was split into training and test sets with an 85:15 ratio. We applied a 3-fold cross-validation on the training set to mitigate overfitting and provide more reliable performance estimates. The held-out test set was used only for the final evaluation.

**Data Augmentation:** All four segmentation models—DeepLabv3+, SAM, SAM2, and Lightweight PH-FPNSeg—used geometric augmentations including random flips, rotations, and affine or shift/scale transformations to boost spatial diversity in training data and prevent overfitting. SAM2 and Lightweight PH-FPNSeg models utilized color and intensity augmentations, such as adjustments to brightness, contrast, gamma, hue, saturation, value, and noise, to enhance model robustness to appearance variations. Lightweight PH-FPNSeg model also included random blur effects to simulate

variations in sharpness, whilst DeepLabv3+ employed elastic transform and random cropping to capture intricate spatial distortions. All pipelines included normalization and tensor conversion. The architecture of each model defined the type and strength of the augmentations used. For example, SAM2 and Lightweight PH-FPNSeg employed more comprehensive color and noise modifications to accommodate greater appearance variability, whereas DeepLabv3+'s atrous convolutions benefited from spatial augmentations that highlighted positional and structural variation. These augmentations expose each model to a diverse range of realistic image variations, thereby enhancing generalization and reducing the risk of overfitting.

**Hyperparameters:** All models were trained for 250 epochs using early stopping with a patience of 25 epochs. For most of the models, early stopping triggered before reaching 200 epochs. However, given the continued performance improvements observed in Lightweight PH-FPNSeg, an additional 50 epochs were allocated to balance accuracy with computational efficiency. We saved only the checkpoint with the best validation dice score for each model to ensure we evaluated generalization at its peak. All models were trained with 8 batch size, except SAM2, which required a batch size of one owing to memory and architectural constraints. Optimization across all models was performed using the AdamW optimizer, chosen for its effective weight decay regularization and robustness in handling sparse gradients. To address class imbalance and enhance segmentation performance in MIC detection, we employed a composite loss function during the training of all models. The loss function combines binary cross-entropy with focal loss to focus learning on challenging pixels, alongside Dice loss to improve spatial overlap between predictions and ground truth. Furthermore, L2 regularization is applied to mitigate overfitting and promote generalization. For Lightweight PH-FPNSeg and SAM2, additional optimization strategies were used, including linear learning rate warm-up, cosine annealing, gradient clipping, mixed-precision training, and Xavier uniform weight initialization. We selected these techniques based on preliminary findings of optimization sensitivity in these models; their inclusion contributed to stabilizing the training process and preventing issues such as gradient explosion.

**Computational Resources:** We performed all computations on the LEAP2 HPC Cluster at Texas State University, which offers large-memory nodes with up to 1.5 TB RAM. All training and experiments ran on a single node equipped with an NVIDIA A100 80GB PCIe GPU, 256 GB RAM, 400 GB SSD storage, and a single processor. The configuration ensured access to high-performance computation and allowed for efficient model training and evaluation.

**Evaluation Metrics:** The metrics used in the experiments are accuracy (ACC), area under the curve (AUC), specificity (SP), recall (RC), Dice coefficient (Dice), and intersection over union (IoU).

## 6 Results Analysis

In this section, we present the quantitative and qualitative results across all experiments and discuss their relevance.

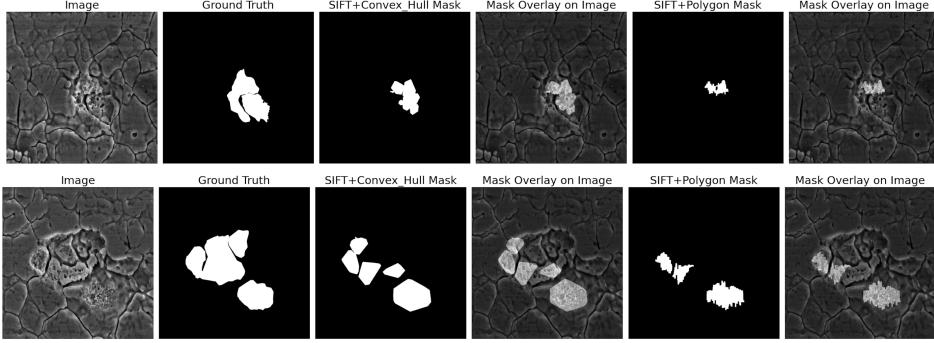
Table 1 and Fig. 6 display the quantitative results and visual outputs for the classical baseline CLF algorithm. The SIFT+Convex Hull variant of CLF successfully detected large, contiguous MIC regions and achieved the highest scores among traditional methods on the extended dataset. The variant achieved a Dice score of 44.07% and IoU of 31.40%, showing moderate overlap between predicted masks and ground truth. While it effectively captured the main corrosion areas, finer details were often missed. The accuracy reached 89.33%, showing that most pixels were correctly classified, and the specificity of 94.67% reflects strong avoidance of false positives in background regions. A recall of 58.13% indicates detection of just over half of the corrosion pixels, with the remainder missed. The AUC of 74.56% demonstrates moderate discriminative ability between corroded and non-corroded areas.

The SIFT+Polygon variant offers greater adaptability for tracing irregular object contours and occasionally captures smaller or more fragmented complex MIC regions. It achieved a Dice score of 37.20% and an IoU of 25.88%, indicating less overlap and more fragmented segmentations. Also, this variant obtained an accuracy of 90.66%, indicating almost same misclassification rate as the convex hull variant, and a specificity of 97.31%, confirming strong avoidance of false positives in background regions. The recall reached 37.91%, showing that less than half of the true corrosion pixels were detected, and the AUC of 67.61% reflected a weaker ability to distinguish corrosion from non-corrosion. Overall, both classical variants of the CLF method offered valuable initial segmentation by identifying primary corrosion regions; SIFT+Convex Hull excelled in overall coverage, whereas SIFT+Polygon gave greater adaptability for shape modification. However, neither fully captured all the subtle and irregular features of MIC. Their results offer a useful baseline but also underscore the challenges traditional algorithms face compared to more advanced deep learning models on complex data.

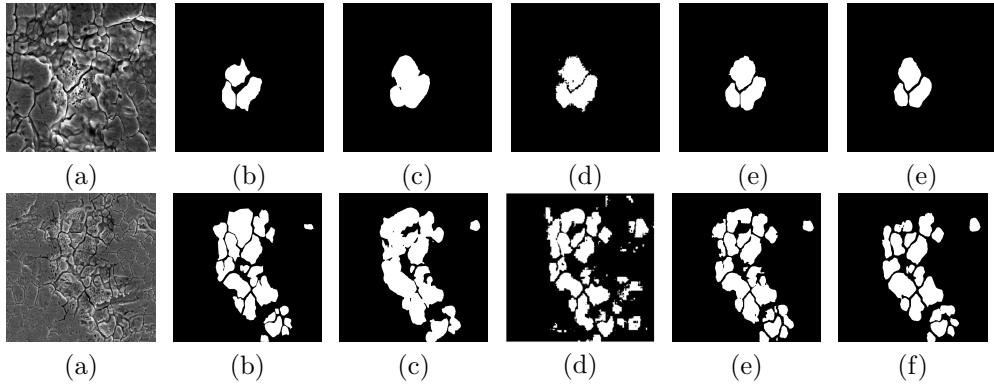
**Table 1** Performance Comparison of Implemented Approaches on the Proposed MIC Dataset

Methods	ACC%	AUC%	SP%	RC%	Dice%	IoU%
SIFT+Convex Hull	89.33	74.56	94.67	54.45	44.07	31.40
SIFT+Polygon	90.66	67.61	97.31	37.91	37.20	25.88
DeepLabv3+	96.05	97.90	97.51	83.85	75.76	62.97
SAM	95.12	97.58	97.22	75.22	73.67	59.68
<b>SAM2</b>	96.78	<b>98.69</b>	<b>98.43</b>	83.46	81.85	70.31
<b>Lightweight PH-FPNSeg</b>	<b>96.83</b>	98.40	97.63	<b>85.77</b>	<b>82.95</b>	<b>71.42</b>

Building on this, the deep learning methods achieved significant improvements in accuracy. They produced visually more consistent results, as reflected in both the quantitative metrics shown in Table 1 and the visual comparisons presented in Figure 7. To begin with, the DeepLabv3+ model achieved a Dice score of 75.76% and an IoU of 62.97%, indicating strong overlap between predicted masks and the ground truth. The model effectively covered most corrosion regions but missed some small structures and fine boundary details, reducing precision for those areas. The high



**Fig. 6** Non-deep learning proposed Cluster with Local Feature (CLF) algorithm visual output with SIFT+Convex Hull and SIFT+Polygon variant



**Fig. 7** Visual output comparison between experimented deep learning segmentation methods (a) MIC Image, (b) Ground Truth, (c) DeepLabv3+, (d) SAM, (e) SAM2, (f) Lightweight PH-FPNSeg

accuracy (96.05%) and specificity (97.51%) indicate that the model can effectively classify most background and foreground pixels. The result also showed that about 83.85% (recall) of actual corrosion pixels are successfully identified. In column (c) of Fig.7, it is evident that the DeepLabv3+ masks can locate the main corrosion areas and generally align with the ground truth shapes. However, they tend to be smoothed and sometimes overlap at boundaries where the ground truth separates neighboring regions.

The second model, SAM achieved a Dice score of 73.67% and an IoU of 59.68%, which is slightly lower than DeepLabv3+. However, these results still represent strong coverage of corrosion areas. It achieved high accuracy (95.12%) and specificity (97.22%), but the recall (75.22%) showed that it missed many MIC pixels, especially in fragmented or fine-grained regions. The output mask, as shown in column (d), means that SAM can identify the main corrosion areas; however, it still misses small or thin regions and does not always capture fine details. There are also some false-positive outputs where background areas are mistakenly labeled as corrosion, resulting in some noisy regions in the mask.

The enhanced SAM2 version demonstrated a clear improvement, achieving the highest Dice score (81.85%) and IoU (70.31%) among all previous models. These results reflected the model's ability to achieve substantial coverage and strong overlap with true corrosion regions, including complex and fragmented MIC structures. It also achieved high accuracy (96.78%) and specificity (98.43%), correctly classifying nearly all image pixels. Also, the model achieved a recall of 83.46%, detecting almost all corrosion pixels, and an AUC of 98.69%, confirming SAM2's strong ability to distinguish corrosion from background. Visual inspection of Fig. 7 (column (e)) shows that the predicted masks closely follow the smallest branches and clusters of corrosion, with boundaries that align well with expert annotations.

Additionally, the proposed new model, Lightweight PH-FPNSeg, performed at a similarly high level as SAM2, with a Dice score of 82.95% and an IoU of 71.42%. The accuracy (96.83%) and specificity (97.63%) are both high, but it stands out for achieving the highest recall (85.77%) of all models, showing Lightweight PH-FPNSeg rarely misses actual MIC compared to other models. It can find both small and fragmented regions effectively. In column (f) of Fig. 7, it is visually evident that the model accurately reconstructs both prominent and subtle MIC areas, displaying sharp, detailed boundaries that closely match the ground truth.

Overall, DeepLabv3+ and SAM serve as strong deep learning baselines, effectively segmenting the main MIC regions, but they miss finer details and provide partial coverage in complex areas. In contrast, SAM2 and Lightweight PH-FPNSeg's advanced architectures and prompt strategies offer higher accuracy and generate segmentation masks that closely match expert annotations in terms of completeness and detail.

## 6.1 Comparison with Previous Study

Table 2 compares the results of the current experiments with those reported in the previous work [21]. For clarity, the SIFT-based method with morphological operations of the prior study is referred to here as SIFT+Polygon, while the version without morphological operations is labeled as SIFT+Convex Hull. Both methods retain their original algorithms without changes. In the previous work [21], the SIFT+Polygon method achieved a Dice score of 34.89% and an IoU of 24.13%. In contrast, when applied to the new dataset, the same method yielded improved results, with a Dice score of 37.20% and an IoU of 25.88%. On the other hand, the SIFT+Convex Hull approach achieved the highest performance on the extended dataset, showing nearly a 10% increase in Dice score and a 6% improvement in IoU compared to the previous best results of SIFT with morphological operations.

Similarly, the DeepLabv3+ model reported in [21] achieved a Dice score of 69.53% and an IoU of 53.92%. In this study, these metrics improved to 75.76% and 62.97%, respectively. The SAM model also demonstrated enhanced performance, with Dice and IoU scores increasing from 71.36% and 58.92% in the previous work to 73.67% and 59.68% in the current experiments. Most notably, the SAM2 model and newly proposed Lightweight PH-FPNSeg model outperformed all prior approaches, achieving Dice scores exceeding 82% and IoU values above 70%. Together, these results demonstrate a significant advancement in microbial-influenced corrosion (MIC) segmentation on our extended dataset.

**Table 2** Performance comparisons with the previous study on the proposed MIC dataset

Methods	Dice %	IoU %
SIFT+Convex Hull	44.07	31.40
SIFT(without morph.ops) [21]	26.20	18.98
SIFT+Polygon	37.20	25.88
SIFT [21]	34.89	24.13
DeepLabv3+	75.76	62.97
Deeplabv3+ [21]	69.53	53.92
SAM	73.67	59.68
SAM [21]	71.36	58.92
<b>SAM2 (new)</b>	<b>81.85</b>	<b>70.31</b>
<b>Lightweight PH-FPNSeg (new)</b>	<b>82.95</b>	<b>71.42</b>

## 6.2 Ablation Study

We conducted an ablation study on our proposed Lightweight PH-FPNSeg model to evaluate the contribution of its individual components to MIC segmentation. The study began with a lightweight CNN encoder-decoder baseline, which achieved a Dice score of 63.42% and an IoU of 49.59%. We initially tested ConvNeXt as a backbone, motivated by its hybrid convolution-transformer design. However, its pre-trained representations failed to transfer effectively to MIC images, producing results nearly identical to the simpler CNN. We therefore adopted the lightweight CNN backbone, which retained the ConvNeXt-inspired staged structure while reducing the parameter count to just 16.47M. Next, we introduced four AttentionConv blocks, which added  $\sim 10$ M parameters and improved performance by +2.71% Dice and +2.80% IoU through attention-guided refinement and better handling of irregular corrosion textures.

**Table 3** Ablation study of the proposed Lightweight PH-FPNSeg model, demonstrating the contribution of each component on overall performance.

Configuration	Dice%	$\Delta$ Dice	IoU%	$\Delta$ IoU	Param(M)
Baseline(CNN backbone+Decoder)	63.42	-	49.50	-	16.47M
+ AttentionConv Blocks	66.13	+2.71	52.30	+2.80	26.54M
+ FPN	70.80	+4.67	57.02	+4.72	31.02M
+ HeatMap	76.63	+5.83	63.29	+6.27	31.02M
+ Prompt (Full Model)	82.95	+6.32	71.42	+8.13	32.41M

Adding a Feature Pyramid Network (FPN) further enhanced multi-scale feature fusion. It increased the parameter count by  $\sim 4.5$ M and provided a gain of +4.67% Dice and +4.72% IoU, emphasizing the role of multi-scale context in detecting corrosion regions of varying size and morphology. Integrating a Heatmap module directed the model to corrosion-relevant regions and refined boundary details, leading to an additional +5.83% Dice and +6.27% IoU. Finally, the Prompt Encoder enabled

prompt-aware adaptation, improving segmentation by +6.32% Dice and +8.13% IoU. With all components combined, the final Lightweight PH-FPNSeg model achieved 82.95% Dice and 71.42% IoU, while maintaining high efficiency with only 32.41M parameters. These results confirm that each component contributes incrementally, and together they achieve state-of-the-art accuracy with minimal computational cost, making the model both effective and practical for MIC segmentation.

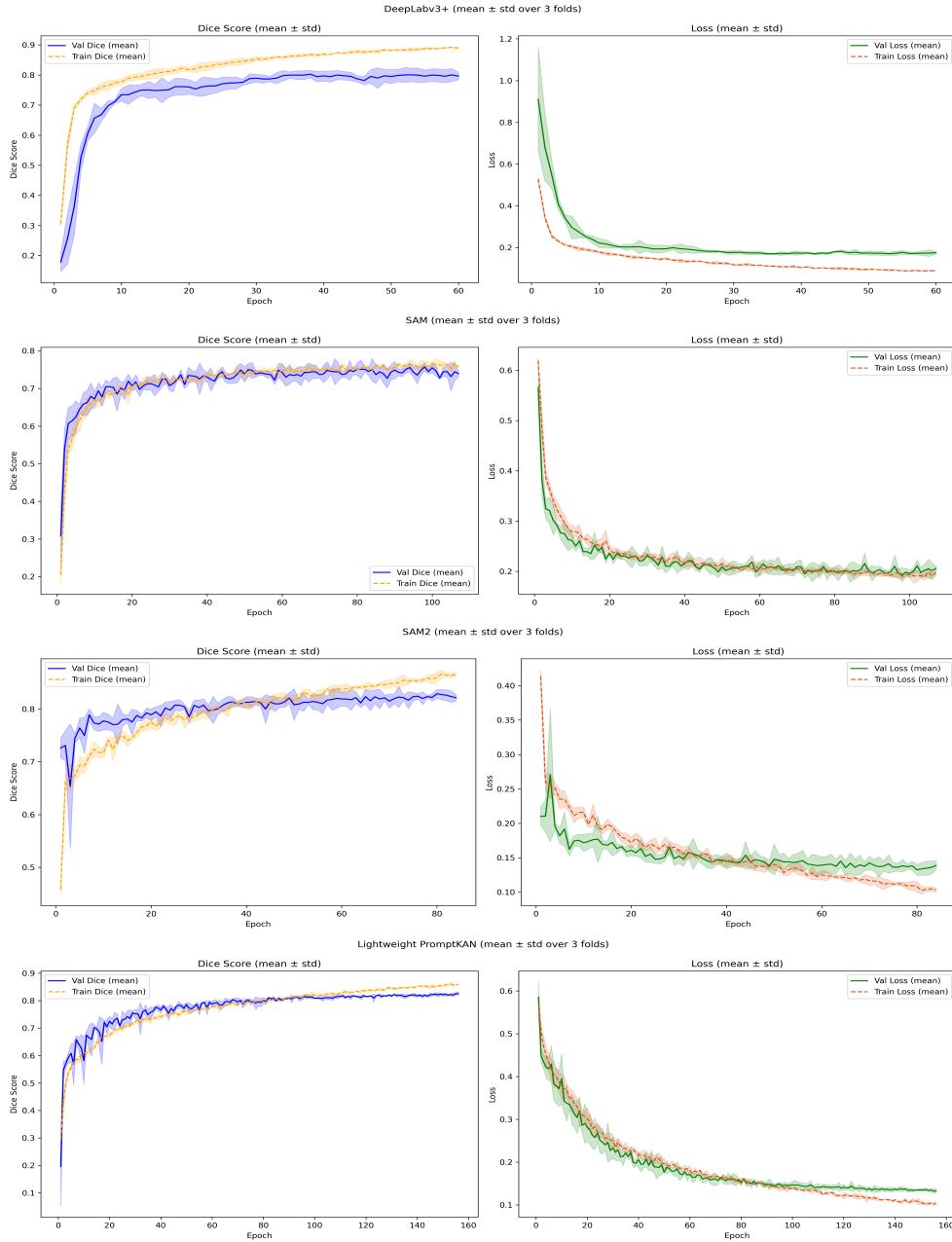
### 6.3 Accuracy vs. Computational Trade-offs

Among the deep learning models, DeepLabv3+ (17.9M parameters) trained the fastest, completing three folds in approximately 34 minutes by utilizing early stopping within a 250-epoch budget. Its relatively simple convolutional architecture and smaller parameter count contributed to this efficiency. However, DeepLabv3+ achieved a Dice score of 75.76%, which was lower than that of more advanced models, indicating limitations in capturing the complex spatial patterns characteristic of MIC segmentation. SAM with a ViT-B backbone, comprising 91M parameters, showed only marginal performance gains, attaining a Dice score of 73.67% after nearly 4 hours and 56 minutes of training, suggesting that increased architectural complexity did not necessarily translate to higher accuracy. SAM2, with a hierarchical Vision Transformer (ViT) backbone comprising 224.4 million parameters, achieved a Dice score of 81.85% after approximately 1 hour and 44 minutes of training. While it delivered improved accuracy with moderate training time, its huge parameter size increases memory usage and computational resource requirements. Lightweight PH-FPNSeg, with 32.41 million parameters, delivered the highest Dice score of 82.95%, though it required the longest training time of nearly 5 hours and 16 minutes. Despite the increased computational cost, this model offered the most favorable trade-off between accuracy and parameter efficiency, benefiting from its hierarchical structure and prompt-aware feature modulation. Overall, models incorporating prompt mechanisms and deeper architectures enhanced segmentation performance but incurred higher training times and computational demands.

### 6.4 Model Training Analysis

Figure 8 illustrates the training and validation Dice scores and losses for all evaluated models, offering a comparative analysis of their performance during the training process. For all deep learning models, results are averaged over three folds and truncated at the least number of epochs completed by any fold. It ensures that each point on the curve reflects all folds and allows for fair and unbiased comparisons.

In the DeepLabv3+ plot, training and validation Dice increased rapidly during the early epochs. Validation stabilized around 0.78–0.80 by epoch 20–30, while training climbed above 0.86. The widening gap reflects overfitting, even though validation stayed steady. Moderate fold-to-fold variability is evident in the Dice curves, whereas the loss curves dropped sharply and then stabilized at low values with minor variation. In contrast, SAM plot showed a fast increase and early plateau, with training and validation Dice converging near 0.75 and remaining closely aligned. Losses stabilized around 0.18–0.20, with consistently low variability across folds, reflecting stable



**Fig. 8** Learning curves (mean  $\pm$  std over 3 folds) for all models. Each plot shows training and validation Dice and loss, averaged across folds and truncated at the minimum epoch reached by all folds.

and reliable training. In the SAM2 plot, both training and validation Dice scores rise

quickly, with validation stabilizing just above 0.81 and training catching up gradually after epoch 30. The curves stay closely aligned, indicating minimal overfitting. Losses for both sets continued to decrease and converged near 00.12 to 0.15. Medium fold-to-fold variability was observed in both Dice and loss, indicating dependable but less uniform generalization across splits. In the Lightweight PH-FPNSeg model plot, Dice scores initially show higher variability, especially for validation during the first approximately 20 epochs. Both training and validation Dice eventually stabilize above 0.82, with only a small gap between them. The loss curves for training and validation follow a steady downward trend and align closely in later epochs, while fold-to-fold variability reduces as training progresses. In summary, despite differences in convergence dynamics and generalization, all models exhibited stable validation Dice and low, convergent losses.

## 7 Conclusion

Microbial-Induced Corrosion (MIC) poses a serious threat to metal infrastructure, particularly in extreme environments such as space. Our interdisciplinary collaboration, which combines advanced computer science analysis methods with expert microbiology, has significantly improved research quality, helping to ensure safer future space missions beyond low Earth orbit. Building on previous studies, we present an expanded dataset of 331 annotated SEM images of MIC on stainless steel, sourced from both microgravity (flight) and terrestrial (ground-control) conditions—making this the largest and first AI-ready MIC segmentation dataset currently available. We rigorously benchmarked both classical and deep learning methods on this dataset. The non-deep learning Cluster with Local Features (CLF) algorithm, utilizing SIFT features with a Convex Hull method, achieved the best baseline performance (Dice 44.07%, IoU 31.40%). In addition to established DeepLabv3+ and SAM models, we introduce two advanced deep learning segmentation architectures: an enhanced SAM2 and a novel Lightweight PH-FPNSeg model, specifically optimized for MIC detection. Both advanced segmentation models achieved state-of-the-art results, with average Dice scores of 82% and IoU scores of 70%. Our work lays the groundwork for accurate, automated MIC detection from SEM images. It provides a valuable resource for corrosion segmentation research across both spaceflight and terrestrial contexts, enabling more effective analysis and monitoring of microbial influenced corrosion in long-duration space missions.

**Acknowledgments.** This work has been partially supported by the NSF # 2224449 grant, the NASA # 80NSSC22K1361 grant, and the Texas State University Research Enhancement Program grant.

## References

- [1] Abeyrathna, D., Life, T., Rauniyar, S., Ragi, S., Sani, R., Chundi, P.: Segmentation of bacterial cells in biofilms using an overlapped ellipse fitting technique. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 3548–3554 (2021). IEEE

- [2] Flemming, H.-C., Wingender, J., Szewzyk, U., Steinberg, P., Rice, S.A., Kjelleberg, S.: Biofilms: an emergent form of bacterial life. *Nature Reviews Microbiology* **14**(9), 563–575 (2016)
- [3] Shi, X., Avci, R., Geiser, M., Lewandowski, Z.: Comparative study in chemistry of microbially and electrochemically induced pitting of 316l stainless steel. *Corrosion Science* **45**(11), 2577–2595 (2003)
- [4] Cheng, S., Tian, J., Chen, S., Lei, Y., Chang, X., Liu, T., Yin, Y.: Microbially influenced corrosion of stainless steel by marine bacterium vibrio natriegens:(i) corrosion behavior. *Materials Science and Engineering: C* **29**(3), 751–755 (2009)
- [5] Laleh, M., Hughes, A.E., Xu, W., Cizek, P., Tan, M.Y.: Unanticipated drastic decline in pitting corrosion resistance of additively manufactured 316l stainless steel after high-temperature post-processing. *Corrosion Science* **165**, 108412 (2020)
- [6] Gupta, S., Anand, S.: Induction of pitting corrosion on stainless steel (grades 304 and 316) used in dairy industry by biofilms of common sporeformers. *International Journal of Dairy Technology* **71**(2), 519–531 (2018)
- [7] Costerton, J.W., Lewandowski, Z., Caldwell, D.E., Korber, D.R., Lappin-Scott, H.M., et al.: Microbial biofilms (1995)
- [8] Obenhuber, D., Huff, T., Rodgers, E.: Microbial biofilm studies of the environmental control and life support system water recovery test for space station freedom. *SAE transactions*, 888–892 (1991)
- [9] Zea, L., McLean, R.J., Rook, T.A., Angle, G., Carter, D.L., Delegard, A., Denvir, A., Gerlach, R., Gorti, S., McIlwaine, D., et al.: Potential biofilm control strategies for extended spaceflight missions. *Biofilm* **2**, 100026 (2020)
- [10] Diaz, A.M., Li, W., Calle, L.M., Callahan, M.R.: Investigation of biofilm formation and control for spacecraft-an early literature review. In: International Conference on Environmental Systems (2019)
- [11] Carter, D.L., Brown, C.: Impact of biofilms on the design and operation of iss life support systems. In: Annual Meeting of the American Society for Gravitational and Space Research (ASGSR 2017) (2017)
- [12] Abeyrathna, D., Ashaduzzaman, M., Malshe, M., Kalimuthu, J., Gadhamshetty, V., Chundi, P., Subramaniam, M.: An ai-based approach for detecting cells and microbial byproducts in low volume scanning electron microscope images of biofilms. *Frontiers in Microbiology* **13**, 996400 (2022)
- [13] Rojas, D., Rueda, L., Ngom, A., Urrutia, H., Carcamo, G.: Biofilm image segmentation using optimal multi-level thresholding. In: 2009 IEEE International

- Conference on Bioinformatics and Biomedicine, pp. 185–190 (2009). IEEE
- [14] Gaalen, K., Gremse, F., Benn, F., McHugh, P., Kopp, A., Vaughan, T.: Automated ex-situ detection of pitting corrosion and its effect on the mechanical integrity of rare earth magnesium alloy-WE43, *Bioact. Mater.* **8** (2022) 545–558
  - [15] Vyas, N., Sammons, R., Addison, O., Dehghani, H., Walmsley, A.: A quantitative method to measure biofilm removal efficiency from complex biomaterial surfaces using sem and image analysis. *Scientific reports* **6**(1), 32694 (2016)
  - [16] Clark, T., Caufield, H., Parker, J.A., Al Manir, S., Amorim, E., Eddy, J., Gim, N., Gow, B., Goar, W., Haendel, M., et al.: Ai-readiness for biomedical data: bridge2ai recommendations. *bioRxiv* (2024)
  - [17] Hiniduma, K., Byna, S., Bez, J.L., Madduri, R.: Ai data readiness inspector (aidrin) for quantitative assessment of data readiness for ai. In: Proceedings of the 36th International Conference on Scientific and Statistical Database Management, pp. 1–12 (2024)
  - [18] Zha, D., Bhat, Z.P., Lai, K.-H., Yang, F., Jiang, Z., Zhong, S., Hu, X.: Data-centric artificial intelligence: A survey. *ACM Computing Surveys* **57**(5), 1–42 (2025)
  - [19] Tešić, J.: Data Lab. <https://datalab12.github.io>. Accessed: 2025-07-26 (2024)
  - [20] McLean, R.J.C.: Microbiology of Biofilms and Polymicrobial Communities. <https://mcleanlab.wp.txstate.edu/dr-mclean>. Accessed: 2025-07-24 (2024)
  - [21] Akter Tani, T., Scouten, A., Ortiz, E., McLean, R.J., Tešić, J.: Automated corrosion identification in metal imagery: Traditional vs. deep learning. In: International Symposium on Visual Computing, pp. 267–280 (2024). Springer
  - [22] Akter Tani, T.: Automated Space MIC Detection. <https://github.com/DataLab12/SpaceMIC>. Accessed: September 5, 2025 (2025)
  - [23] Ashaduzzman, M., Bommanapally, V., Subramaniam, M., Chundi, P., Kalimuthu, J., Talluri, S., Gadhamshetty, R.: Using deep learning super-resolution for improved segmentation of sem biofilm images. In: 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 3587–3593 (2022). IEEE
  - [24] Bommanapally, V., Ashaduzzman, M., Malshe, M., Chundi, P., Subramaniam, M.: Self-supervised learning approach to detect corrosion products in biofilm images. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 3555–3561 (2021). IEEE
  - [25] Iannucci, L., Lombardo, L., Parvis, M., Cristiani, P., Basséguy, R., Angelini, E., Grassini, S.: An imaging system for microbial corrosion analysis. In: 2019

IEEE International Instrumentation and Measurement Technology Conference (I2MTC), pp. 1–6 (2019). IEEE

- [26] Zhu, T., Zhu, S., Zheng, T., Ding, H., Song, W., Li, C.: Heu-net: hybrid attention residual block-based network with external skip connections for metal corrosion semantic segmentation. *The Visual Computer* **40**(2), 1273–1287 (2024)
- [27] Wang, B., Liu, L., Cheng, X., Wu, W., Liu, C., Zhang, D., Li, X.: Advanced multi-image segmentation-based machine learning modeling strategy for corrosion prediction and rust layer performance evaluation of weathering steel. *Corrosion Science* **237**, 112334 (2024)
- [28] Stoean, R., Bacanin, N., Ionescu, L., Boicea, M., Gărău, A.-M., Ghîtescu, C.-C.: Semantic segmentation for corrosion detection in archaeological artefacts before restoration. In: 2021 23rd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), pp. 246–251 (2021). IEEE
- [29] Zhao, L., Locke, J., Xu, F., Yao, T., Guo, X.: Accurate segmentation of localized corrosion in structural alloys via deep learning. *npj Materials Degradation* **9**(1), 80 (2025)
- [30] Han, Q., Zhao, N., Xu, J.: Recognition and location of steel structure surface corrosion based on unmanned aerial vehicle images. *Journal of Civil Structural Health Monitoring* **11**(5), 1375–1392 (2021)
- [31] Rahman, M.H., Azam, M.A., Hossen, M.A., Ragi, S., Gadhamshetty, V.: Biofilmscanner: A computational intelligence approach to obtain bacterial cell morphological attributes from biofilm image. arXiv preprint arXiv:2302.09629 (2023)
- [32] Gurung, B.D.S., Khanal, A., Hartman, T.W., Do, T., Chataut, S., Lushbough, C., Gadhamshetty, V., Gnimpieba, E.Z.: Transformer in microbial image analysis: a comparative exploration of transunet, unet, and doubleunet for sem image segmentation. In: 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 4500–4502 (2023). IEEE
- [33] Holicheva, A.A., Kozlov, K.S., Boiko, D.A., Kamanin, M.S., Provotorova, D.V., Kolomoets, N.I., Ananikov, V.P.: Deep generative modeling of annotated bacterial biofilm images. *npj Biofilms and Microbiomes* **11**(1), 16 (2025)
- [34] Albahri, M.B., Barifcani, A., Iglauder, S., Lebedev, M., O’Neil, C., Salgar-Chaparro, S.J., Machuca, L.L.: Investigating the mechanism of microbiologically influenced corrosion of carbon steel using x-ray micro-computed tomography. *Journal of Materials Science* **56**(23), 13337–13371 (2021)
- [35] NASA Ames Research Center: Bacterial Adhesion and Corrosion (SpaceX-21). Accessed: 2025-07-10 (2021). <https://www.nasa.gov/ames/space-biosciences/>

- [36] University of Colorado Boulder: BioServe Space Technologies: BioCell. <https://www.colorado.edu/center/bioserve/spaceflight-hardware/biocell>. Accessed: 2025-08-03 (2024)
- [37] Burger, W., Burge, M.J.: Scale-invariant feature transform (sift). In: Digital Image Processing: An Algorithmic Introduction, pp. 709–763. Springer, ??? (2022)
- [38] Smelyakov, K., Tovchyrechko, D., Ruban, I., Chupryna, A., Ponomarenko, O.: Local feature detectors performance analysis on digital image. In: 2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T), pp. 644–648 (2019). IEEE
- [39] Likas, A., Vlassis, N., Verbeek, J.J.: The global k-means clustering algorithm. Pattern recognition **36**(2), 451–461 (2003)
- [40] Deng, D.: Dbscan clustering algorithm based on density. In: 2020 7th International Forum on Electrical Engineering and Automation (IFEEA), pp. 949–953 (2020). IEEE
- [41] Jayaram, M., Fleyeh, H.: Convex hulls in image processing: a scoping review. American Journal of Intelligent Systems **6**(2), 48–58 (2016)
- [42] Bradski, G., Kaehler, A.: Learning OpenCV: Computer Vision with the OpenCV Library. O'Reilly Media, Inc., Sebastopol, CA (2008)
- [43] Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 801–818 (2018)
- [44] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., et al.: Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4015–4026 (2023)
- [45] Ravi, N., Gabeur, V., Hu, Y.-T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., et al.: Sam 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714 (2024)
- [46] Jiang, Y.: Prompt engineering in segment anything model: Methodologies, applications, and emerging challenges. arXiv preprint arXiv:2507.09562 (2025)