DeepSC-Edge: Scientific Corrosion Segmentation with Edge-Guided and Class-Balanced Losses

Biao Yin[1], Nicholas Josselyn[1], Thomas Considine[2], John Kelley[2], Berend Rinderspacher[2], Robert Jensen[2], James Snyder[2], Ziming Zhang[1,3], Elke Rundensteiner[1,2]

[1]Department of Data Science, Worcester Polytechnic Institute, Worcester, MA
[2]Department of Computer Science, Worcester Polytechnic Institute, Worcester, MA
[3]Department of Electrical and Computer Engineering, Worcester Polytechnic Institute, Worcester, MA
[4]Army Research Directorate, DEVCOM Army Research Laboratory, Aberdeen Proving Ground, MD, USA

***Abstract*—Corrosion is a prevalent issue in numerous industrial fields, causing expenses nearing $3 trillion or 4% of the GDP annually with safety threats and environmental pollution. To timely qualify and validate new corrosion-inhibiting materials on a large scale, accurate and efficient corrosion assessment is crucial. Yet it is hindered by a lack of automatic tools for expert-level corrosion segmentation of material science experimental images. Developing such tools is challenging due to limited domain-valid data, image artifacts visually similar to corrosion, various corrosion morphology, strong class imbalance, and millimeter-precision corrosion boundaries. To help the community address these challenges, we curate the first expert-level segmentation annotations for a real-world image dataset [1] for scientific corrosion segmentation. In addition, we design a deep learning based model, called DeepSC-Edge that achieves guidance of ground-truth edge learning by adopting a novel loss that avoids over-fitting to edges. It also is enriched by integrating a class-balanced loss that improves segmentation with small area but crucial edges of interest for scientific corrosion assessment. Our dataset and methods pave the way to advanced deep-learning models for corrosion assessment and generation – promoting new research to connect computer vision and material science discovery. Once the appropriate approvals have been cleared, we expect to release the code and data at: https://arl.wpi.edu/**

I. Introduction

Background. Corrosion is defined as the gradual degradation of a metal over time due to chemical interactions with its environment. It results in major safety risks worldwide and negatively impacts nature environment, societal health, national infrastructure, manufacturing, and transportation. The associated economic burden is substantial, with global losses estimated to be around 4% of the gross domestic product (GDP), equivalent to approximately $2.5 trillion [2], [3]. Consequently, the study of corrosion is an active research field in material science that aims to innovate environmentally friendly materials capable of corrosion resistance. This involves conducting scientific corrosion assessment by various industries, government agencies, and countries [4]–[9].

Motivation. Traditional methods for assessing corrosion often involve manual segmentation, which is both labor-intensive and subject to human error. This lack of precise, quick, and safe assessment methods impedes not only the understanding of corrosion as well as material discovery in general. While the need for efficient and accurate corrosion assessment has been acknowledged [10], the transition to automated methods using Machine Learning (ML) has been slow. One of the significant roadblocks is the absence of high-quality scientific data for corrosion segmentation. This type of data is costly and complex to acquire, often requiring the expertise of seasoned professionals to identify subtle and varied corrosion patterns, especially those with artifacts like water stains with similar colors or textures to corrosion.

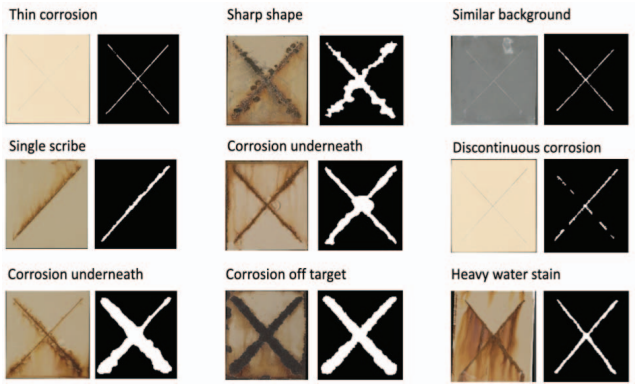

Figure 1: Sample corrosion images with the segmentation ground truth we provide with corrosion experts.

Shown in Figure 1, some corrosion forms, such as corrosion underneath, i.e. under the coatings, necessitate invasive and potentially hazardous methods for accurate assessment where the expert needs to scrape the corresponding coating out. There is also the challenge of millimeter-level observations, particularly crucial when defining the boundaries of corroded areas. Given these complexities, there is an urgent need for an automated corrosion assessment tool that can accurately deal with the nuances of various types of corrosion, thereby aiding both academic research and industrial applications.

Data and Ground Truth. Our original 600 corrosion panel images are fully taken from an open domain dataset [1] derived from standard corrosion science experiments according to ASTM standards [7]. Ground truth, expertly curated, binary segmentation is produced for each of the 600 images in this

1946-0759/23/$31.00 ©2023 IEEE
DOI 10.1109/ICMLA58977.2023.00251

work. In consultation with domain experts, we develop criteria to segment corrosion of interest while excluding superfluous background pixels.

Challenges. From examples shown in Figure 1, challenges that are both domain-specific and which extend to the general computer vision community [11], include: i) a variety of textures and shapes in limited images, ii) tiny-scale and detailed areas to segment, iii) similarities in color with coating or image artifacts such as water staining, iv) areas of corrosion not visibly apparent on the images but identifiable to domain experts, and v) class imbalance between corrosion and background pixels which arises naturally due to the relatively small size and irregular distribution of interest, such as any item above. These challenges lead us to the research questions below:

- To what extent can a deep learning architecture be trained to learn the expert-curated segmentation including the challenging corrosion field of interest?
- On limited data, how to better guide the model by exploiting strategies tied to the above observations, such as edge detection or class imbalance methods?

Our Proposed Methodology. We first train and compare two popular widely-used deep learning architectures, UNet [12] and MedTransformer [11], that had been shown to be successful for tasks to segment tiny but important pixels. Although UNet outperforms MedTransformer on our small domain-representative dataset, we found that UNet cannot fully learn edge information of the small-scaled corrosion crucial for corrosion assessment. This then motivates us to guide UNet with ground-truth edge maps generated by deterministic computer vision techniques such as the Canny edge operator [1]. We thus develop an edge-guided baseline model, ET-UNet. This model adopts the backbone from UNet [12], while now augmented with edge-guidance capabilities [13].

While ET-UNet is designed to guide segmentation boundaries, it lacks the ability to guide the boundary via edge map features in decoder layers. Typically, this Decoder edge guidance can cause overfitting to edges but not the segmentation within.

To address the issues, we propose three alternate strategies on ET-UNet: 1) constructing targeted edge guidance on Decoder, 2) innovating losses to regularize the caused overfitting issue, 3) employing a class-balanced method, Focal Tversky Loss (FTL) to further distinguish challenging corrosion of interest in the proposed edge guidance model. Our proposed models are listed in Table I.

Findings. We demonstrate that our proposed strategies effectively recover critical edge details while learning the segmentation – providing a 1% dice and 1.43% IOU increase over 86.72% dice or 77.27% IOU score from the UNet baseline on our image data set. After sorting and weighting the test set by its prediction rank performance using baseline UNet (weighting baseline poorer performing images higher), our proposed methods outperform UNet by 1.66% dice and 2.41% IOU. This demonstrates our ability to handle more challenging, difficult-to-segment images. The predictions on certain difficult test images are improved by about 8% dice or IOU increase.

Contributions. In summary, our key contributions in this work are as follows:

- We design a loss method enabling ground-truth edge guidance via decoder output to enhance segmentation performance while preventing the guidance overfits to edges (Equation 6.), compared to traditional encoder guidance
- We employ a class-imbalanced method to our novel edge guidance loss – effectively learning segmentation with the challenging edges of corrosion
- DeepSC-Edge over traditional methods is affirmed by a Rank metric measuring their performance on corrosion images with challenging edges.
- We provide the first expert-curated segmentation image dataset, opening rich research opportunities for scientific material discovery, deep learning, and computer vision.

II. Related Work

Deep learning & data sets in material science.. For corrosion science, machine learning solutions have been applied to automate engineering tasks such as defect detection [14]–[17] and corroded pipe detection [18]. However, there is no well-performing ML work for scientific corrosion segmentation [19]. Due to the difficulty and expertise required to annotate corrosion, to the best of our knowledge, our work provides the first expert-level segmentation dataset for scientific corrosion assessment and corresponding high-performing deep learning methods for standardized corrosion segmentation.

In the broader sense, deep learning has been widely used for image segmentation tasks. UNet is a popular deep learning model for various image segmentation tasks [12], [20]–[22] - first introduced for biomedical image segmentation. It primarily introduced the utilization of a skip-connection structure to propagate information from the encoder to the decoder. Further, there exist works that aim to refine predicted segmentation by integrating additional boundary and edge information [23]–[26]. Losses to address class imbalance during training in tasks like object detection by applying a modulating term to the cross entropy loss to focus learning on hard misclassified examples has been shown to be effective. [27]. In this work, we also incorporate edge information into UNet. However, we demonstrate methods involving ground-truth edges and a class-balanced loss to benefit scientific corrosion segmentation, especially on difficult-to-segment images that pose both domain-related and general-purpose segmentation challenges.

III. Data Set Description

Experimental Workflows for Coating and Rating Panels for Corrosion Levels. Corrosion panel testing is used globally in industrial and government labs to conduct standardized corrosion tests on surfaces with protective coatings [1]. In corrosion science, experimentalists develop and validate new

[1] https://kornia-tutorials.readthedocs.io/en/latest/_nbs/filtering_edges.html#canny-edges

Model	Description (Equation)	Method
UNet	UNet baseline for image segmentation (1)	Baseline 1
ET-UNet	Edge guidance architecture based on UNet (2)	Baseline 2
Guided ET-UNet	ET-UNet (2) + additional edge guidance (5)	Proposed Strategy 1
S-Guided ET-UNet	ET-UNet (2) + additional edge guidance in a soft manner (6)	Proposed Strategy 1, 2
UNet + FTL	UNet (1) + class balanced loss (7)	Baseline 1 + Proposed Strategy 3
ET-UNet + FTL	ET-UNet (2) + class balanced loss (7)	Baseline 2 + Proposed Strategy 3
Guided ET-UNet + FTL	ET-UNet (2) + additional edge guidance (5) + class balanced loss (7)	Proposed Strategy 1, 3
S-Guided ET-UNet + FTL	ET-UNet (2) + additional edge guidance in a soft manner (6) + class balanced loss (7)	Proposed Strategy 1, 2, 3

Table I: Model descriptions: UNet-based models that reflect our methods combining baselines and proposed strategies. Strategy 1 or 2 is proposed upon ET-UNet. DeepSC-Edge A is Guided ET-UNet. DeepSC-Edge B is S-Guided ET-UNet.

anti-corrosive materials following standard material science procedures [28]. Based on the corrosion area of interest, they rate the panels from 0 (heavy corrosion) to 10 (no corrosion) according to the defined ASTM standards. As the rating scale is at a millimeter level, it is necessary to segment corrosion out from the background in a detailed tiny scale where the experts exploit magnifiers to verify it point by point. These tests are widely conducted to evaluate the performance of the newly-invented coatings in preventing corrosion.

Experimental Data Collection. Our dataset is derived from a related, open dataset [1] includes 600 images of corroded panels that have been expertly rated but not yet segmented. This data set represents diverse material types in coating stack layers (5 substrates, 2 profiles, 21 pre-treatments, 4 primers, and 5 topcoats) and interaction of distinctive features (water spots, under-coating rust, tiny shapes, and edges largely of interest, etc.) not found in popular image datasets like ImageNet. Their images are evenly balanced across 5 critically defined rating categories of 5 to 9, and are resampled and rating-stratified to 10-fold cross-validation sets with a held-out test set containing 60 images. An example image of each rating class can be seen in Figure 1. In this work, we provide expertly curated, binary segmentations for each of these 600 images – representing the ground truth for scientific corrosion segmentation. They were obtained using the OpenCV GrabCut algorithm [29] to separate areas of scribe corrosion from the surrounding background.

Scientific Corrosion Segmentation Image Preparation. To ensure accuracy and completeness, we consulted with corrosion domain experts to develop criteria for segmenting scribe corrosion areas. This allowed us to include all necessary areas of corrosion in the segmentation while excluding irrelevant background pixels, such as dark water stains. Using GrabCut, we define a bounding box including all scribe corrosion and allow the algorithm to determine preliminarily the location of the foreground (corrosion) and background. Then, we manually refine the areas of corrosion and background using GrabCut until we achieve an accurate segmentation for the given panel [29]. The segmentation output from GrabCut is then transformed into a grayscale image and binarized such that all scribe corrosion areas are set to 255 and all background pixels are set to zero. We then apply a median filter [30] and morphological operators to the binary images to remove any remaining salt and pepper noise that wasn't removed initially by GrabCut. The binary segmentations were reviewed by corrosion domain experts and further refined collaboratively thereafter. Examples can be seen in Figure 1.

IV. Methodology

Our proposed solution, DeepSC-Edge, consists of two sub-models, DeepSC-Edge A (Guided ET-UNet in Table I) and B (S-Guided ET-UNet in Table I).

Inspired by the expert annotation process, it is noticeable that corrosion boundary plays a vital role in learning detailed expert-level knowledge about corrosion on a panel. In UNet baseline, early encoder layers are designed to learn low-level features [12] such as textures and colors of corrosion pixels. To guide these layers in learning the edges of expert ground truth segmentation, we built an UNet edge guidance baseline, ET-UNet, according to a widely used edge guidance network ET-Net [13]. However, corrosion boundaries may not always be visually apparent and low-level features might not represent the complete segmentation easily. Even worse, parts of boundaries might be similar to certain background pixels or to corrosion not of interest. Additionally, the expert rules to determine these boundaries are sample-dependent and varied considering complex factors from actual environments, raw materials, or even errors during the manufacturing and experimental process.

In order to ensure that ET-UNet is learning low-level features within the edge boundaries, we propose a strategy that considers additional edge guidance (Strategy 1 in Table I) on ET-UNet as Guided ET-UNet. Further, to avoid overfitting to the boundary rather than the segmentation, we propose to soft-guide by gradually injecting it as segmentation loss decreases during training (Strategy 2 in Table I) on Guided ET-UNet as S-Guided ET-UNet.

For edges or areas of corrosion that are the minority compared to other areas, we propose to involve the class-balanced loss, FTL, to exploit models for better recognition. We describe all of the models with methods in Table I.

Baselines. UNet is a convolutional neural network architecture that uses an encoder-decoder structure to perform segmentation by predicting pixel-wise labels. The encoder extracts low-level features from the input image, which are then used by the decoder to predict the segmentation shape.

We aim to investigate how edge guidance could be involved in this classic deep learning architecture – approaching a better automation of scientific corrosion segmentation. Thus, as an alternative baseline, we also develop an edge-guided UNet

model inspired by the edge guidance modules in ET-Net [13], called ET-UNet. Figure 2 illustrates how ET-UNet involves ground-truth edge guidance to the encoder-decoder architecture in UNet. In the model, the Edge Guidance Module (EGM) is designed to learn low-level features related to segmentation edges, and Weighted Aggregation Module (WAM) to capture multi-scale features to form the segmentation. **Strategies.** We propose the following strategies to improve baselines inspired by our observations and understanding of scientific corrosion segmentation: i) We involve additional guidance – by verifying edges of the segmentation prediction in ET-UNet with edges of ground truth. ii) In order to avoid the issue of over-fitting to the edges, we propose to add this guidance in a soft manner – increasing it as the segmentation loss decreases gradually. iii) Further, we utilize Focal Tversky Loss (FTL) [31] instead of Dice loss as a regularization to overcome the class imbalance challenge in the domain.

Strategy 1 is going to enhance ET-UNet by ensuring its prediction edge to approach that of the ground truth edge. Strategy 2 is proposed to avoid over-fitting to the edges while not the segmentation itself when applying Strategy 1. Strategy 3 aims at emphasizing learning the challenging corrosion of interest with minority but critical edges.

Loss functions to exploit deep learning models. Our strategies are added to our DeepSC-Edge models listed in Table I. We define the loss functions of these deep learning models below. Combining Binary Cross Entropy (BCE) and Dice loss ($Dice\_Loss$) [12], UNet loss function is calculated by ground truth segmentation Y and the segmentation prediction $\hat{Y}$:

$$\mathcal{L}_{UNet} = BCE(Y,\ \hat{Y}) + Dice\_Loss(Y,\ \hat{Y}) \tag{1}$$

Other than UNet, ET-UNet also learns edges of ground truth segmentation in early encoder layers so that its loss function can be described as:

$$\mathcal{L}_{ET-UNet} = \mathcal{L}_{UNet} + \lambda \cdot \mathcal{L}_{edge} \tag{2}$$

$$where\ \mathcal{L}_{edge} = BCE(Yedge, \hat{Y}edge) + Dice\_Loss(Yedge, \hat{Y}edge), \tag{3}$$

In Equation (3), $Yedge$ is the edge map of ground truth segmentation, $\hat{Y}edge$ is the predicted edge map from its Edge Guidance Module illustrated in Figure 2, and λ controls the strength of this edge guidance. We then involve the additional edge guidance shown in Table I:

$$\mathcal{L}_{edge^*} = BCE(Yedge, \hat{Y}edge^*) + Dice\_Loss(Yedge, \hat{Y}edge^*). \tag{4}$$

where $\hat{Y}edge^*$ denotes the edge map of segmentation prediction output from Weighted Aggregation Module of ET-UNet – instead of $\hat{Y}edge$ that is output from the Edge Guidance Module as shown in Figure 2. We control this additional guidance using γ. So that:

$$\mathcal{L}_{Guided\ ET-UNet} = \mathcal{L}_{ET-UNet} + \gamma \cdot \mathcal{L}_{edge^*} \tag{5}$$

$$\mathcal{L}_{S-Guided\ ET-UNet} = \mathcal{L}_{ET-UNet} + \frac{1}{\max\left(\mathcal{L}_{seg} - \theta,\ 0\right) + 1} \cdot \mathcal{L}_{edge^*} \tag{6}$$

By assigning θ appropriately, we involve the additional guidance, $\mathcal{L}_{edge^*}$, in the following soft (S) manner: the prediction edge, $\hat{Y}edge^*$, will provide its guidance in inverse proportion to $\mathcal{L}_{seg}$, and the full guidance to the prediction segmentation only if $\mathcal{L}_{seg}$ is smaller than or equal to θ. For Strategy 3, class balanced loss FTL [31] combines Tversky Index [31] and Focal loss [27]:

$$FTL = (1 - TI)^{1/\beta} \tag{7}$$

$$where\ TI = \frac{TP + \varepsilon}{TP + \alpha FN + (1-\alpha)FP + \varepsilon} \tag{8}$$

We embed FTL loss into our models (+FTL) shown in Table I instead of their $Dice\_Loss$. In Equation (8), we calculate True Positive (TP), False Negative (FN) and False Positive (FP) pixels according to ground truth and prediction in $Dice\_Loss$ used in the model. If α is larger than 0.5, the loss penalizes FN so the model will emphasize learning these challenging corrosion areas more. β further down-weights the contribution of corrosion that is easy to be predicted and focuses more on challenging corrosion during training. In this way, the model would learn to preserve sharp, minority, or challenging edges rather than avoid them, and as result, this would make the prediction of edges smooth.

V. Experimental Study and Analysis

Experimental Setup. Each model for a particular set of hyperparameters was trained in a range 1 to 2 hours using an A100 GPU. Thereafter, each image is inferred in less than a second. This efficiency was critical for our targeted applications to integrate the AI models in a corrosion data collection and assessment tool as an iOS APP for our domain collaborators to collect and work with their data.

We trained and evaluated all models using our dataset consisting of 600 corrosion image pairs: the original images and their corresponding segmentation we provide in this work.

To grid-search hyperparameters, we select learning rates from a set {1e-1, 1e-2, 1e-3, 1e-4, 1e-5} and batch size from a set {8, 16}. We found that all models work the best with learning rate 1e-3 and batch size 8. The best λ in ET-UNet (2) is 0.5 tuned from a set {0.001, 0.1, 0.5, 1.0, 1.5}. The best γ in Guided ET-UNet (5) based on the best ET-UNet is 1.0 tuned from a set {0.001, 0.1, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0}. The best θ in S-Guided ET-UNet (6) based on the best ET-UNet is 3.0 tuned from a set {1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0}.

For Proposed Strategy 3 on all the above best models, we found the best penalization weight α in FTL is 0.88 for ET-UNet + FTL and S-Guided ET-UNet + FTL models, and 0.90 for other models related – both showing a high false negative rate penalization is beneficial in the dataset.

We also present an ablation study – showing our results are robust to hyperparameter choice in the proposed strategies. We

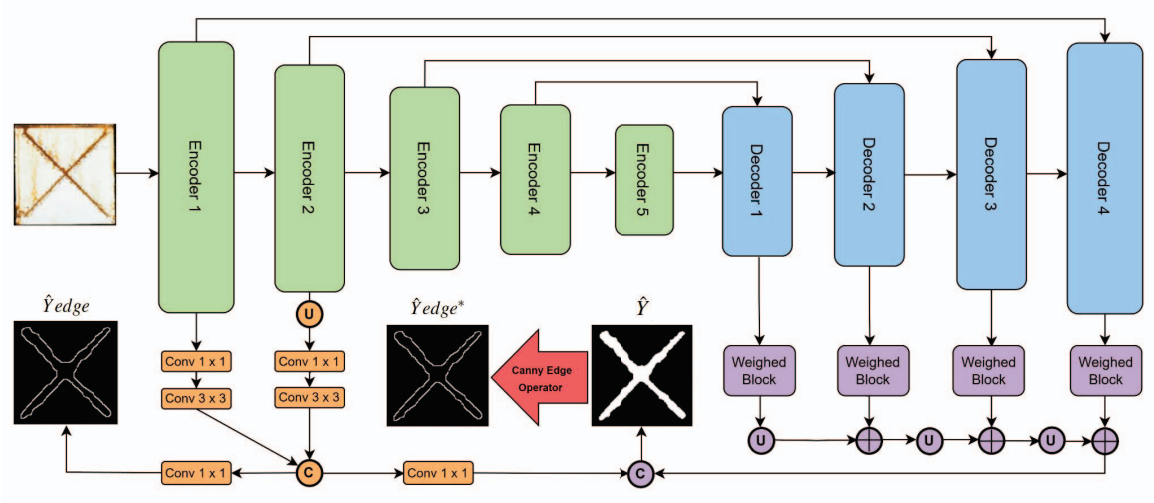

Figure 2: Guided ET-UNet: UNet encoder layers are in green, while UNet decoder layers are highlighted in blue. Edge Guidance Module [13] is in orange. Weighted Aggregation Module [13] is highlighted in purple. 'Conv', 'U', 'C', and '+' signify the convolutional layer, upsampling, concatenation, and addition layers, respectively. Red Arrow illustrates the deterministic canny edge operator. Our Strategy 1 develops a loss (Equation 5) to train this new decoder edge guidance with traditional encoder edge guidance loss (Equation 2) upon UNet loss (Equation 1) for segmentation. Our Strategy 2 proposes a novel loss (Equation 6) to regularize the proposed Guided ET-UNet loss so not to overfit to edges via weighting it with the segmentation loss (Equation 1). The decoder edge guidance in Guided ET-UNet is expected to be gradually increased if it does encourage the overall corrosion segmentation. However, this guidance will not be effective once the segmentation has been fully learned, as the weight will be 0 if the segmentation loss is smaller than a hyperparameter $theta$. In our Strategy 3 (similar to Strategies 1 and 2), we integrate a class imbalanced method (Equation 7) into the segmentation and related edge guidance methods for telling corrosion with challenging but critical boundaries.

evaluate the model performance using Dice and IOU scores as broadly used in machine segmentation tasks [12].

In order to better observe the evaluation, we also defined their variants other than scores – rank, increase, and weighted increase (W-Inc as shown in Table II). The Dice or IOU rank takes the average rank performance of a model after calculating its rank compared with other models on every test image. The Dice or IOU weighted increase (W-Inc) takes the weighted average score improvement of a model rather than the UNet baseline after weighting each test image with its prediction rank using the UNet baseline. In this way, we evaluate how a model is or is not able to predict challenging corrosion of interest better. If W-Inc is larger than 0, it indicates that the model outperforms UNet in predicting corrosion segmentation, especially on challenging corrosion.

Comparative Study of Corrosion Segmentation Performance. Shown in Table II, our results indicate that all of our strategies improve baseline models. S-Guided ET-UNet (+FTL) model performs the best based on its Dice, IOU, and corresponding ranks and increases, demonstrating its superior performance in solving the corrosion segmentation task compared to other strategies and baselines, especially on challenging corrosion. Further, incorporating any of our proposed strategies elevates ET-UNet performance over UNet – showing its effectiveness in edge-guidance. This is a notable

Model	Dice				IOU			
	Score ↑	Rank ↓	Increase ↑	W-Inc ↑	Score ↑	Rank ↓	Increase ↑	W-Inc ↑
UNet	0.8672 ± 0.0179	5.5167	0.0000	0.0000	0.7727 ± 0.0263	5.5167	0.0000	0.0000
ET-UNet	0.8670 ± 0.0181	5.6667	-0.0002	0.0014	0.7721 ± 0.0268	5.7000	-0.0006	0.0014
Guided ET-UNet	0.8687 ± 0.0176	5.3167	0.0015	0.0028	0.7749 ± 0.0260	5.2500	0.0022	0.0040
S-Guided ET-UNet	0.8696 ± 0.0199	4.6333	0.0024	0.0045	0.7763 ± 0.0288	4.6167	0.0036	0.0062
UNet + FTL	0.8738 ± 0.0152	4.0167	0.0066	0.0149	0.7823 ± 0.0226	4.0500	0.0096	0.0133
ET-UNet + FTL	0.8763 ± 0.0151	3.6833	0.0091	0.0094	0.7855 ± 0.0227	3.6833	0.0128	0.0214
Guided ET-UNet + FTL	0.8768 ± 0.0168	3.8500	0.0096	0.0155	0.7864 ± 0.0254	3.8667	0.0137	0.0232
S-Guided ET-UNet + FTL	**0.8772 ± 0.0140**	**3.3167**	**0.0100**	**0.0166**	**0.7870 ± 0.0212**	**3.3167**	**0.0143**	**0.0241**

Table II: Test performance using 10-fold cross-validation: each cell shows the average test performance of a model under a metric. The best-performing cell in each metric column is highlighted in BOLD. The MedTransformer performance (Dice Score: 0.85; IOU: 0.75) cannot beat UNet on our domain small dataset so it was not considered as a basic architecture to build our proposed strategies for scientific corrosion segmentation.

Model (+ FTL)	$\beta = 1.0$				$\beta = 1.5$				$\beta = 2.0$			
	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.7$	$\alpha = 0.9$	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.7$	$\alpha = 0.9$	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.7$	$\alpha = 0.9$
UNet	0.7346	0.8322	0.8737	0.8632	0.7434	0.8223	0.8741	0.8717	0.7547	0.8072	0.8614	0.8669
ET-UNet	0.7891	0.8377	0.8740	0.8672	0.7513	0.8479	0.8711	0.8741	0.7927	0.8493	0.8572	0.8646
Guided ET-UNet	0.7506	0.8359	0.8763	0.8679	0.7461	0.8336	0.8721	0.8768	0.7582	0.8109	0.8557	0.8687
S-Guided ET-UNet	0.7346	0.8276	0.8758	0.8654	0.7706	0.8227	0.8755	0.8766	0.7777	0.8173	0.8659	0.8711

Table III: Ablation Study of Proposed Models: Each cell shows the test performance using Dice Score metric.

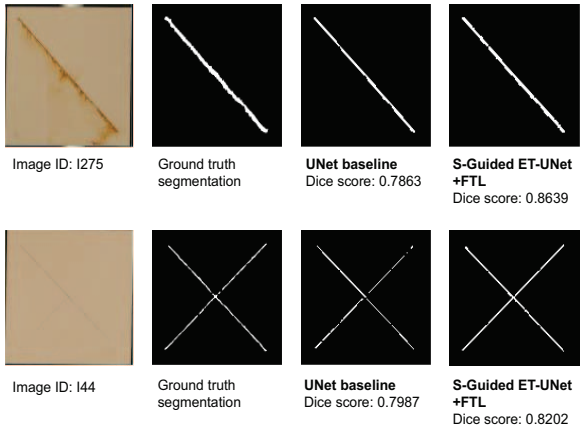

Figure 4: Case study: Challenging corrosion panels segmented by UNet and our proposed S-Guided ET-UNet + FTL model.

improvement considering that UNet is commonly used in the field of image segmentation with relative small data sets.

Moreover, from Weighted Increase performance for handling the difficult-to-segment corrosion, ET-UNet outperforms UNet without our strategies, but they indeed make this improvement larger. In addition, from the loss plots in Figure 3, the models with FTL tend to have smaller total loss values. Also, our best-performing model shows smooth training and validation losses – denoting its robustness.

Ablation Study. Since in Table II, we found the class-balanced loss, (7), shows a significant improvement in each proposed model for our edge-related task. This strategy forces the model to learn minority pixels better by penalizing false negatives or false positives while predicting segmentation or its edge. It helps us to solve the prediction challenge on the small yet crucial areas of corrosion using the limited image pairs – via controlling α and β in the loss. In TableIII, we show that our proposed segmentation models are able to be locally optimized in terms of α and β.

Case study. Providing visual examples of the segmentation predictions from the baseline models and proposed strategies can be a useful way to explain why our strategies are necessary for corrosion segmentation when leveraging the popular UNet deep learning architecture. In Figure 4, we display test results of the baseline UNet using its best validation fold model along with the corresponding predictions from our best-performing model. This comparison highlights the benefits of our proposed strategies for challenging corrosion segmentation tasks. Overall, these visual examples highlight the potential impact of our solution for improving scientific corrosion segmentation.

VI. Conclusion and Future Work

In this paper, we propose to integrate edge-map information and class-balanced loss with UNet for scientific corrosion segmentation. Our DeepSC-Edge models from the proposed strategies exploit edge information to enhance the feature representation via Decoder and improve the prediction of segmentation to automate scientific corrosion assessment. Our experimental results demonstrate the effectiveness of our model to segment the publicly available corrosion dataset [1] paired with our expert-curated ground truth segmentation, especially on challenging hard-to-segment samples. Our strategies may inspire solutions for other image segmentation tasks that require edge information in UNet.

Our long-term goal is to speed up material discovery-related experimental workflows. The release of our expert-labeled

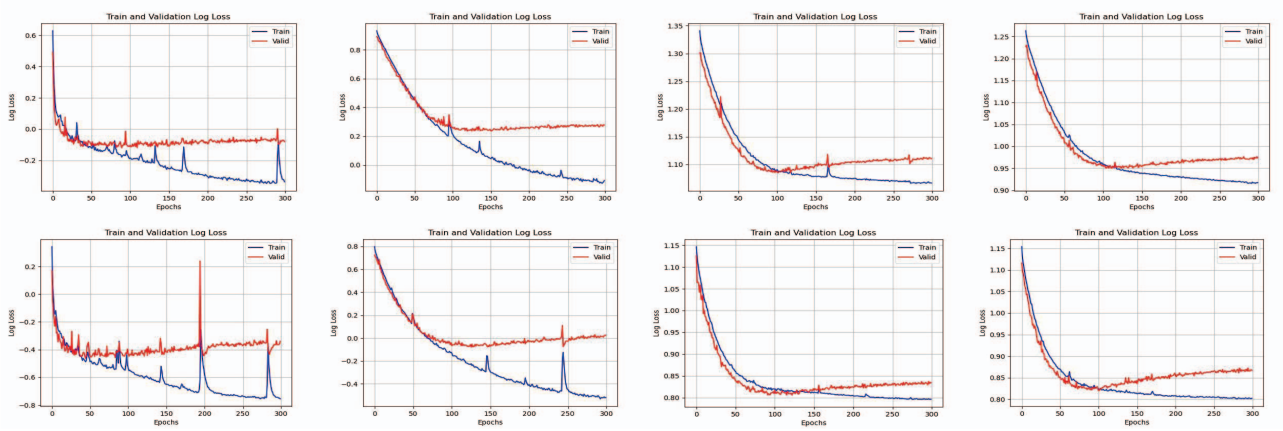

Figure 3: Loss plots on the best UNet validation fold: Top row, from left to right, shows UNet, ET-UNet, Guided ET-UNet, and S-Guided ET-UNet. Bottom row shows the corresponding models with the consideration of class-balanced loss – FTL.

segmentation corrosion data and our deep learning methods applied to this unique application domain will drive innovation of deep learning techniques such as UNet-based generative models, like Stable Diffusion [32], for scientific corrosion progression prediction and transfer learning from large models, like SAM [33], to this important domain – bridging computer vision and material science discovery.

VII. Acknowledgement

Research was sponsored by DEVCOM Army Research Laboratory under Cooperative Agreement W911NF-19-2-0112, W911NF-17-2-0227, W911NF1920108, and partially supported by DOE GANN P200A180088 Fellowship and NSF grants 1815866, 1910880, CCF-2006738, NRT-CEDAR 2021871. The views and conclusions in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of DEVCOM Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

[1] B. Yin, N. Josselyn, T. Considine, J. Kelley, B. Rinderspacher, R. Jensen, J. Snyder, Z. Zhang, and E. Rundensteiner, "Corrosion image data set for automating scientific assessment of materials," in *British Mach. Vis. Conf., BMVC 2021*.

[2] D. M. Bastidas, "Corrosion and protection of metals," *Metals*, vol. 10, no. 4, 2020. [Online]. Available: https://www.mdpi.com/2075-4701/10/4/458

[3] G. Koch, "Cost of corrosion," *Trends in oil and gas corrosion research and technologies*, pp. 3–30, 2017.

[4] C. Rudén and S. O. Hansson, "Registration, evaluation, and authorization of chemicals (reach) is but the first step–how far will it take us? six further steps to improve the european chemicals legislation," *Environmental health perspectives*, vol. 118, no. 1, pp. 6–10, 2010.

[5] N. E. C. Co and J. T. Burns, "Effects of macro-scale corrosion damage feature on fatigue crack initiation and fatigue behavior," *Intl. Journal of Fatigue*, vol. 103, pp. 234–247, 2017.

[6] T. Considine, D. Braconnier, J. Kelley, T. Braswell, C. Miller, B. Placzankis, and R. Jensen, "Data analytic prediction and correlation visualization of corrosion assessment for DoD sustainment," *Technical Report of US. CDCC Army Research Lab*, 2018, Unpublished.

[7] L. YING-YU and W. QUI-DONG, "Astm D1654: standard test method for evaluation of painted or coated specimens subjected to corrosive environments," *Annual Book of ASTM Standards. West Conshohocken*, 1992.

[8] J. I. Skar and D. Albright, "Corrosion behavior of die-cast magnesium in astm b117 salt spray and gm9540p cyclic corrosion test," *2003 Magnesium Technology*, vol. 1, p. 59, 2003.

[9] A. C. D.-. on Paint, M. Related Coatings, and Applications, *Standard Test Method for Evaluation of Painted or Coated Specimens Subjected to Corrosive Environments*. ASTM International, 2008.

[10] J. Li, K. Lim, H. Yang, Z. Ren, S. Raghavan, P.-Y. Chen, T. Buonassisi, and X. Wang, "Ai applications through the whole life cycle of material discovery," *Matter*, vol. 3, no. 2, pp. 393–432, 2020.

[11] J. M. J. Valanarasu, P. Oza, I. Hacihaliloglu, and V. M. Patel, "Medical transformer: Gated axial-attention for medical image segmentation," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24*. Springer, 2021, pp. 36–46.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[13] Z. Zhang, H. Fu, H. Dai, J. Shen, Y. Pang, and L. Shao, "Et-net: A generic edge-attention guidance network for medical image segmentation," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22*. Springer, 2019, pp. 442–450.

[14] H. Lin, B. Li, X. Wang, Y. Shu, and S. Niu, "Automated defect inspection of led chip using deep convolutional neural network," *Journal of Intelligent Manufacturing*, vol. 30, no. 6, pp. 2525–2534, 2019.

[15] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *arXiv:1506.01497*, 2015.

[16] Y.-J. Cha, W. Choi, G. Suh, S. Mahmoudkhani, and O. Büyüköztürk, "Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types," *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 9, pp. 731–747, 2018.

[17] Y. Bai, H. Sezen, and A. Yilmaz, "End-to-end deep learning methods for automated damage detection in extreme events at various scales," *arXiv preprint arXiv:2011.03098*, 2020.

[18] D. Vriesman, A. B. Junior, A. Zimmer, and A. L. Koerich, "Texture cnn for thermoelectric metal pipe image classification," in *2019 IEEE 31st Intl. Conf. on Tools with Artificial Intelligence (ICTAI)*. IEEE, 2019, pp. 569–574.

[19] M. Micikevičius, "Corrosion detection on steel panels using semantic segmentation models," Ph.D. dissertation, Vilniaus universitetas, 2023.

[20] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz *et al.*, "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.

[21] Q. Zuo, S. Chen, and Z. Wang, "R2au-net: attention recurrent residual convolutional neural network for multimodal medical image segmentation," *Security and Communication Networks*, vol. 2021, pp. 1–10, 2021.

[22] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation," *arXiv preprint arXiv:1802.06955*, 2018.

[23] G. Kong, H. Tian, X. Duan, and H. Long, "Adversarial edge-aware image colorization with semantic segmentation," *IEEE Access*, vol. 9, pp. 28 194–28 203, 2021.

[24] D. Wang, W. Dai, D. Tang, Y. Liang, J. Ouyang, H. Wang, and Y. Peng, "Deep learning approach for bubble segmentation from hysteroscopic images," *Medical & Biological Engineering & Computing*, vol. 60, no. 6, pp. 1613–1626, 2022.

[25] D. Cheng, G. Meng, S. Xiang, and C. Pan, "Fusionnet: Edge aware deep convolutional networks for semantic segmentation of remote sensing harbor images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 12, pp. 5769–5783, 2017.

[26] Y. Chen, A. Dapogny, and M. Cord, "Semeda: Enhancing segmentation precision with semantic edge aware loss," *Pattern Recognition*, vol. 108, p. 107557, 2020.

[27] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[28] A. B117, "Standard practice for operating salt spray (fog) apparatus," *ASTM International (1997 Edition)*, 2011.

[29] C. Rother, V. Kolmogorov, and A. Blake, ""grabcut" interactive foreground extraction using iterated graph cuts," *ACM transactions on graphics (TOG)*, vol. 23, no. 3, pp. 309–314, 2004.

[30] S. K. Sathua, A. Dash, and A. Behera, "Removal of salt and pepper noise from gray-scale and color images: an adaptive approach," *arXiv preprint arXiv:1703.02217*, 2017.

[31] N. Abraham and N. M. Khan, "A novel focal tversky loss function with improved attention u-net for lesion segmentation," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019, pp. 683–687.

[32] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 10 684–10 695.

[33] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo *et al.*, "Segment anything," *arXiv preprint arXiv:2304.02643*, 2023.