ENHANCED FOUNDATION MODEL-BASED INTERACTIVE SEGMENTATION FOR BOREHOLE IMAGE DATA

Imane Baho*
SLB, France
ibaho@slb.com

Eya Ghamgui*
SLB, France
eghamgui@slb.com

Abderaouf Boudia
SLB, UVSQ, France
aboudia2@slb.com

Franco Marchesoni
SLB, ENS-PS, France
fmarchesoni@slb.com

Josselin Kherroubi
SLB, France
jkherroubi@slb.com

***Abstract*—Corrosion detection in borehole data is challenging due to the complexity of the data and the difficulty in separating corrosion from the surrounding background. Although artificial intelligence models offer promising automation for this process, their effectiveness is impeded by the limited availability of labeled data. Generating accurate segmentation labels remains a laborious, resource-intensive and time-consuming, diverting experts' attention from other important tasks. Interactive image segmentation offers a practical solution by minimizing user input, thus lightening the annotation workload. In addition, the advent of large foundation models in interactive image segmentation has further streamlined this task across various domains. One state-of-the-art model, Segment Anything Model (SAM) [1], is celebrated for its robust segmentation capabilities. However, when applied to corrosion detection datasets, SAM fails to provide good performance. Attempts to adapt SAM using weak supervision techniques, such as WeSAM [2], have resulted in only marginal improvements. In this paper, we improve the performance of interactive segmentation models to facilitate more precise and efficient labeling of corrosion datasets. We propose two key improvements to the WeSAM method: improving the clicking procedure and the prompt encoding. These enhancements resulted in over 13-point gain in the Intersection over Union (IoU) metric, significantly improving the effectiveness of the corrosion detection labeling process.**

***Index Terms*—Corrosion Detection, Borehole Data, Interactive Image Segmentation, Weakly Supervised Learning, Domain Adaptation.**

I. Introduction

Evaluating the condition of the wellbore is a crucial stage in a well's life-cycle. This step involves the use of various advanced logging tools like electromagnetic, ultrasonic, and mechanical devices to inspect the pipes. These devices provide a geological view of the open-hole subsurface and assess well integrity in cased-hole wells, focusing on the quality of the casing and cement. Particularly, ultrasonic tools are non-invasive corrosion diagnostic and monitoring instruments, essential for well-integrity analysis [3]. They contain emitter/receiver transducers, which emit ultrasonic waves transmitted and reflected through pipe and formation boundaries. These reflected waves are used to measure the distance between the ultrasonic tool and the pipe, from which we derive both pipe radius and thickness maps. The resulting visualizations represent metal distribution relative to a nominal value, highlighting areas of metal excess or loss. Excess metal indicates regions with additional material, while metal loss signifies areas with less material.

Moreover, the pipes, which are manufactured using a mold, exhibit homogeneous patterns within small sections called joints, which are connected by collars. These manufacturing patterns and collars serve as the background for any defects or anomalies. Corrosion in the pipe manifests as metal loss, and may occur in either the inner or outer wall, or both. The distinction between manufacturing patterns and actual defects, while precisely characterizing defect properties, is a complex task, making manual labeling even more challenging. Thus, generating labels for large volume of data is not only time-consuming but also requires specialized expertise to avoid errors. In addition, the subjective nature of human interpretation further complicates the process. To overcome these challenges, automated annotation methods are crucial. Interactive image segmentation stands out as an effective solution, combining expert input with AI-driven methods to facilitate labeling, speed up the overall annotation process, and increase adoption.

A. Interactive Image Segmentation

Interactive image segmentation is a process in which users provide input to guide a model in accurately segmenting specific regions of an image. The human input can take various forms, including clicks in the foreground or background, text instructions to guide the segmentation, or bounding boxes around objects of interest. Many powerful interactive segmentation approaches have recently appeared [4–13], often starting from encoders pre-trained on large image dataset. However, state-of-the-art methods perform badly when evaluated on out-of-domain datasets such as remote sensing [14] or geological data

To overcome this, a more generalizable vision foundation model, such as SAM [1], has been developed, inspired by the success of the language foundation models pre-trained on the large text dataset. This approach allows a promptable segmentation to generate valid segmentation masks, showing its ability to enable zero-shot segmentation using prompts. Many variants of the SAM model have appeared to achieve more memory-efficient and higher-quality segmentation, as in

*These two authors contributed equally to this work.

[15, 16]. Aside from the higher quality data, one key factor contributing to the high quality of segmentation in SegNext [16] is the dense representation and fusion of visual prompts. These methods are designed to be generalist, but they perform poorly in specific industrial domains, highlighting the need for further exploration into domain adaptation.

B. Domain Adaptation

Traditional machine learning models rely on the assumption that training and test data are drawn from the same distribution, a condition that is often violated in real-world applications due to domain shift. While collecting new labeled instances to address domain shift can be costly and labor-intensive, leveraging publicly available annotated data offers a practical alternative. However, this can result in domain shift. To address this issue, domain adaptation techniques (DA) [17, 18] have been developed to minimize the distributional differences between the source and target domains.

Among various DA approaches, Source-Free Domain Adaptation (SFDA) [19, 20] has become a key area of research, especially when access to source data is restricted due to privacy concerns, storage limitations, or confidentiality issues. For example, SHOT [19] applies self-training on unlabeled target data to alleviate domain shift without requiring source data. Furthermore, methods such as teacher-student frameworks [21] facilitate mutual learning through self-training architectures, iteratively refining the model using only target data.

Limited access to target data is another major issue faced by industry due to the privacy and expensive costs. To address this, weakly supervised SFDA [19, 20] addresses limitations by incorporating weak supervision, such as bounding boxes, points, or coarse masks. This approach enables the model to learn effectively even with limited labels. Recently, [22] uses weak labels to align source-target features, achieving competitive performance compared to supervised learning methods. Building upon these advancements, WeSAM has been proposed to adapt SAM using weak supervision. This method applies a self-training strategy to adapt SAM to the target distribution.

In this paper, we introduce an enhanced interactive foundation model for corrosion detection. We summarize our contributions as follows:

- Strategic click-based approach simulating human interaction by using iterative prediction errors to refine the learning process.
- Effective prompt encoding mechanism within the model's architecture to enhance its adaptability.

II. Methodology

In this section, we provide an overview of WeSAM followed by our proposed approach, that integrates the high-quality segmentation and the weakly supervised domain adaptation techniques.

A. Overview of WeSAM: Weakly supervised Segment Anything Model

WeSAM is an adaptation approach designed to improve the generalization of segmentation models like SAM across domain shifts. It employs a self-training strategy with weak supervision, generating pseudo-labels for the target domain to enhance model performance. The architecture includes three encoder networks (anchor, student, teacher) with shared weights between the student and teacher. The anchor network, frozen, retains knowledge from the source domain. Different augmentations are applied to the networks, producing three feature maps. The decoder generates foreground masks from various prompts, including bounding boxes, points, and coarse segmentation masks. The prompt encoding used is identical to the one used for SAM. Adaptation objectives include updating the student/teacher networks, anchor loss regularization, and contrastive loss. All training prompts are derived from the ground-truth segmentation mask to simulate human interactions as weak supervision. The point prompt is created by randomly selecting 5 positive points within the mask and 5 negative points outside it, and the coarse segmentation mask is generated by fitting a polygon to the ground-truth mask.

B. Our Contribution

1) Contribution 1: Clicking Procedure: To minimize the effort required for human labeling, we focus exclusively on using points as input prompts for the model. We adopt the iterative click simulation strategy from RITM [12]. Based on ground truth and model prediction, we automatically generate click-based inputs for training, simulating human interaction. This method generates clicks sequentially, with each new click strategically placed in regions of the model's previous prediction that are erroneous or uncertain. By focusing on these regions, the model iteratively refines its predictions, enabling more accurate and efficient learning from a reduced number of user interactions. In our experiments, we iteratively select 10 points based on previous errors, offering a more targeted approach than the original method, which randomly selects 5 positive and 5 negative points.

2) Contribution 2: Prompt Encoding: Recognizing the limitations of SAM in terms of high-quality segmentation, SegNext [16] was introduced as a new approach that outperformed current state-of-the-art methods. Inspired by this approach, we introduce the dense prompt encoding of SegNext in the WeSAM architecture. The visual prompts (in our case clicks or masks) are encoded using a 3-channel dense map. Clicks are encoded as a binary disk with a predefined radius: positive clicks (foreground) are encoded in the first channel; negative clicks (background) are encoded in the second channel. The mask is encoded in the third channel, because of the ambiguous information that could come from the false positives and negatives of the first initial segmentation masks. This will allow the user to improve the next segmentation output based on the previous one. The visual prompts are then combined with image feature maps. These feature maps are passed

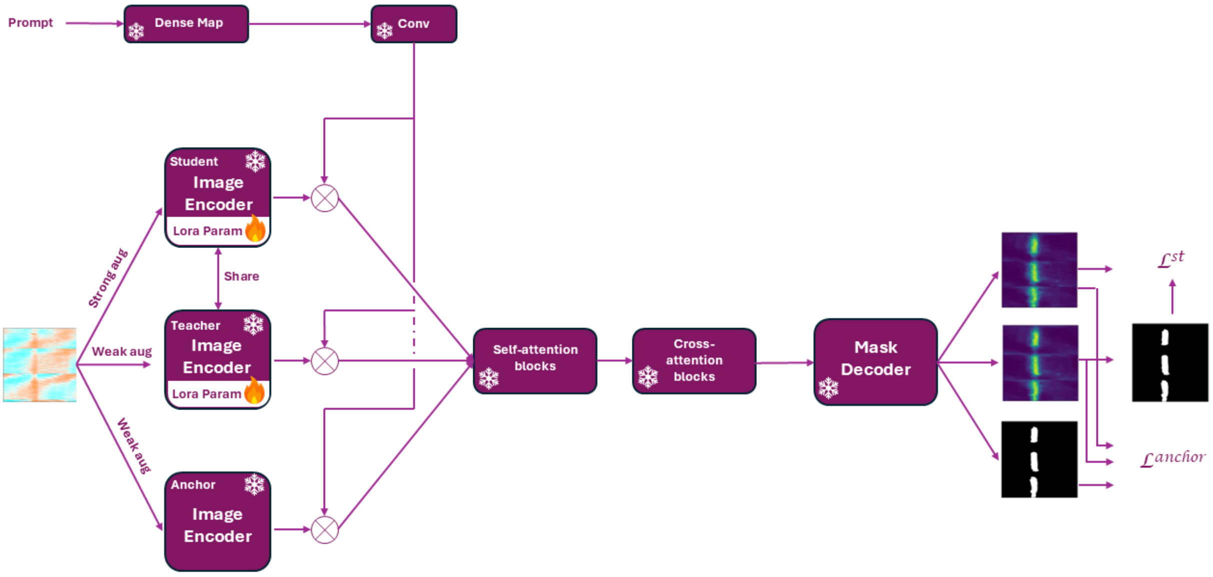

Fig. 1: Illustration of the proposed architecture.

through self-attention and cross-attention blocks, followed by a mask decoder. The figure 1 illustrates the new architecture.

III. Experiments

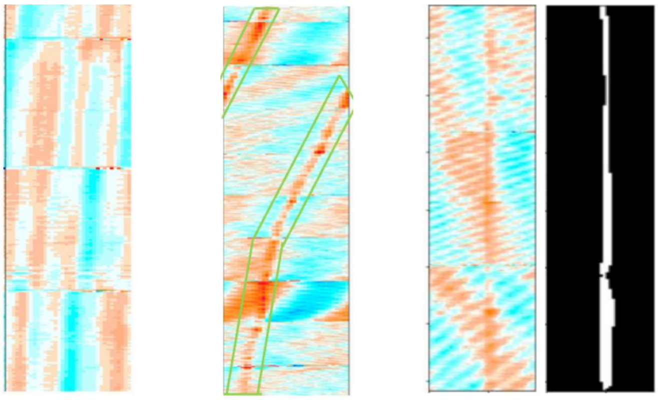

Fig. 2: Illustration of the THBK dataset. Left: defect-free image, center: axial groove image, right: THBK image and its corresponding binary mask [23].

A. THBK Dataset

In this paper, we evaluate our method on the THBK dataset (Azimuthal Variation of Pipe Thickness) [23], which is acquired through ultrasonic measurements. The main objective is to segment the axial long defects, referred to as grooves. These defects represent metal loss either in the inner or outer wall of the pipe, as depicted by the red patterns in Figure 2. The dataset consists of 180 grayscale images along with their corresponding masks, as shown in Figure 2. The acquisition process may cause some data points to be corrupted. Therefore, data cleaning and pre-processing are required to train the model. In addition, we apply various data augmentation techniques, which are divided into two categories: weak augmentation (vertical and/or horizontal flips) and strong augmentation (posterization, sharpening, random adjustments in brightness and contrast, and customized contrast shift). One of the biases a model might develop is a preference for cases where there is a high contrast between the target class (grooves) and the background. This is why, a customized contrast shift was applied to create greater overlap between the groove and background.

B. Evaluation Metrics

We evaluate our experiments using the intersection over union (IoU) after each user's input i.e. click which gives insight into the model's learning curve and its ability to rapidly improve segmentation accuracy with incremental user input.

IV. Results & Discussion

	1 click	10 clicks	20 clicks
SAM	5.02	43.81	49.68
WeSAM	2.64	51.57	52.15
Ours 1	2.51	54.67	56.79
SegNext	8.55	52.32	73.83
Ours 2	**14.14**	**65.34**	**83.95**

TABLE I: Quantitative comparison of the IoU metric (%) for 1, 10, and 20 clicks across all methods.

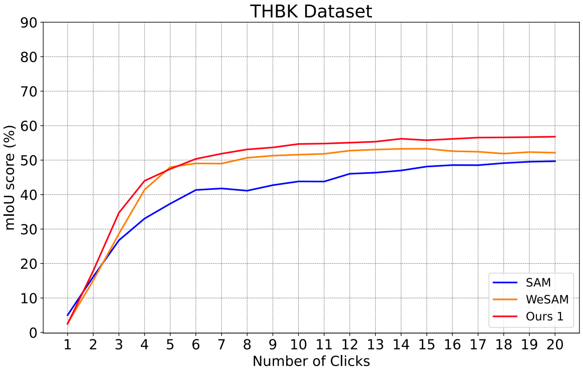

Fig. 3: Quantitative results of SAM, WeSAM and our first contribution.

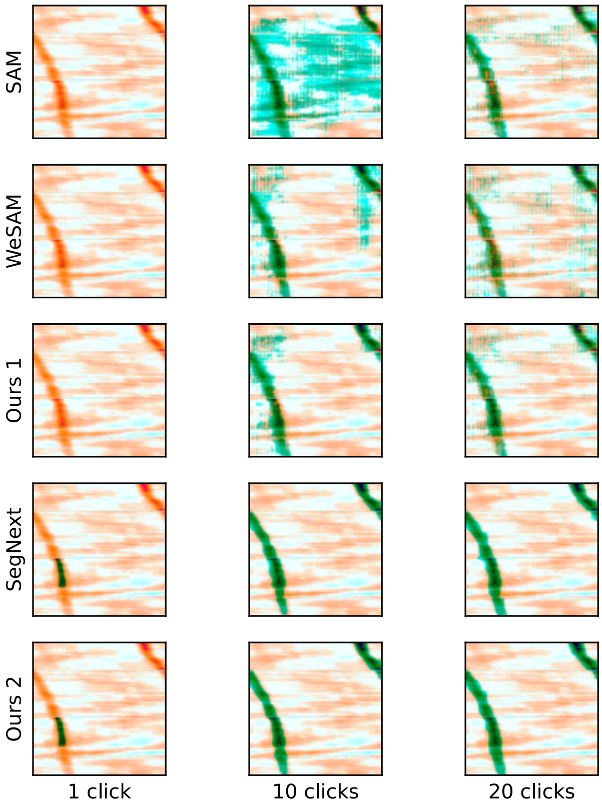

Fig. 4: Qualitative comparison across all methods.

	1 click	10 clicks	20 clicks
No custom contrast shift	12.65	60.03	81.00
Custom contrast shift	**14.14**	**65.34**	**83.95**

TABLE II: Quantitative comparison of the IoU (%) for 1, 10 and 20 clicks for our approach with/without the customized data augmentation.

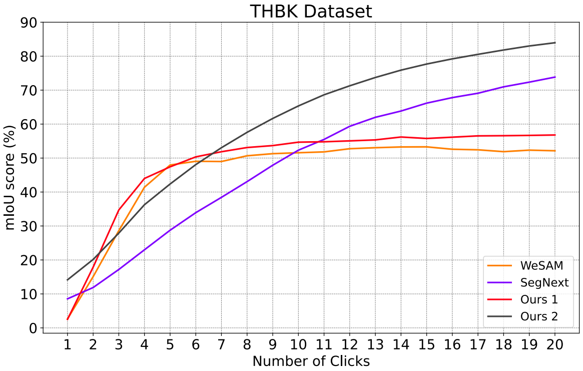

Fig. 5: Quantitative results of WeSAM, SegNext and our global approach.

In this section we present our results for the two contributions. To be noted that we refer to our first contribution by **Ours 1** and the combination of both contributions by **Ours 2**.

The experimental results, illustrated in Figure 3, demonstrate the effectiveness of our first contribution, outperforming both SAM and WeSAM, validating that the clicking procedure during training is an important aspect. Additionally, our global approach (combining both contributions) surpasses SegNext for all the clicks, as shown in Figure 5. Furthermore, according to table I, our approach achieves the best performance compared to all other methods.

Qualitative results, illustrated in Figure 4, support these findings, as SAM consistently lags behind in all experiments, struggling with false positives and false negatives. In contrast, our first contribution provides improvement performance over WeSAM. Moreover, our global approach (combining both contributions), alongside SegNext, demonstrates superior refinement of segmentation masks over time, with our method achieving better overall performance. Notably, our approach reduces false positives more effectively, enhancing the quality of the segmentation and ensuring more precise model outputs compared to SegNext.

We also examined the importance of the customized contrast shift III-A, which enhances performance, as shown in Table II. This adjustment reduces the likelihood of the model relying solely on cases where the target class has high contrast against the background, thereby improving its robustness to varying data distributions.

To summarize, our contribution significantly enhances corrosion detection for borehole images. By incorporating a click-based procedure during training, we effectively mimic human interactions, guiding the model more efficiently. Additionally, the dense representation and fusion of visual prompts emerge as crucial design choice that contribute to high-quality annotation.

References

[1] 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," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 4015–4026.

[2] H. Zhang, Y. Su, X. Xu, and K. Jia, "Improving the generalization of segmentation foundation model under distribution shift via weakly supervised adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 23 385–23 395.

[3] P. Khalili and P. Cawley, "Relative ability of wedge-coupled piezoelectric and meander coil emat probes to generate single-mode lamb waves," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 65, no. 4, pp. 648–656, 2018.

[4] X. Chen, Z. Zhao, Y. Zhang, M. Duan, D. Qi, and H. Zhao, "Focalclick: Towards practical interactive image segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 1300–1309.

[5] H. Ding, S. Cohen, B. Price, and X. Jiang, "Phraseclick: toward achieving flexible interactive segmentation by phrase and click," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*. Springer, 2020, pp. 417–435.

[6] Z. Lin, Z. Zhang, L.-Z. Chen, M.-M. Cheng, and S.-P. Lu, "Interactive image segmentation with first click attention," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 13 339–13 348.

[7] Z. Lin, Z.-P. Duan, Z. Zhang, C.-L. Guo, and M.-M. Cheng, "Focuscut: Diving into a focus view in interactive segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 2637–2646.

[8] ——, "Knifecut: Refining thin part segmentation with cutting lines," in *Proceedings of the 30th ACM International Conference on Multimedia*, 2022, pp. 809–817.

[9] Q. Liu, Z. Xu, Y. Jiao, and M. Niethammer, "isegformer: interactive segmentation via transformers with application to 3d knee mr images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2022, pp. 464–474.

[10] Q. Liu, M. Zheng, B. Planche, S. Karanam, T. Chen, M. Niethammer, and Z. Wu, "Pseudoclick: Interactive image segmentation with click imitation," in *European Conference on Computer Vision*. Springer, 2022, pp. 728–745.

[11] Q. Liu, Z. Xu, G. Bertasius, and M. Niethammer, "Simpleclick: Interactive image segmentation with simple vision transformers," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 22 290–22 300.

[12] K. Sofiiuk, I. A. Petrov, and A. Konushin, "Reviving iterative training with mask guidance for interactive segmentation," in *2022 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2022, pp. 3141–3145.

[13] S. Zhang, J. H. Liew, Y. Wei, S. Wei, and Y. Zhao, "Interactive object segmentation with inside-outside guidance," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 12 234–12 244.

[14] F. Marchesoni-Acland, T. Magne, F. Rekbi, and G. Facciolo, "On the domain generalization capabilities of interactive segmentation methods," *Image Processing On Line*, vol. 14, pp. 25–40, 2024.

[15] L. Ke, M. Ye, M. Danelljan, Y.-W. Tai, C.-K. Tang, F. Yu *et al.*, "Segment anything in high quality," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[16] Q. Liu, J. Cho, M. Bansal, and M. Niethammer, "Rethinking interactive image segmentation with low latency high quality and diverse prompts," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 3773–3782.

[17] G. Wilson and D. J. Cook, "A survey of unsupervised deep domain adaptation," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 11, no. 5, pp. 1–46, 2020.

[18] X. Liu, C. Yoo, F. Xing, H. Oh, G. El Fakhri, J.-W. Kang, J. Woo *et al.*, "Deep unsupervised domain adaptation: A review of recent advances and perspectives," *APSIPA Transactions on Signal and Information Processing*, vol. 11, no. 1, 2022.

[19] J. Liang, D. Hu, and J. Feng, "Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation," in *International conference on machine learning*. PMLR, 2020, pp. 6028–6039.

[20] B. Chidlovskii, S. Clinchant, and G. Csurka, "Domain adaptation in the absence of source domain data," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 451–460.

[21] Y. Huang, L. Yang, and Y. Sato, "Weakly supervised temporal sentence grounding with uncertainty-guided self-training," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 18 908–18 918.

[22] A. Das, Y. Xian, D. Dai, and B. Schiele, "Weakly-supervised domain adaptive semantic segmentation with prototypical contrastive learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 15 434–15 443.

[23] S. Benslimane, J. Kherroubi, K. Singh, J.-L. Le Calvez, T. Berard, and M. Lemarenko, "Automated corrosion analysis with prior domain knowledge-informed neural networks," in *SPWLA Annual Logging Symposium*. SPWLA, 2022, p. D051S017R003.