Signal-driven Shipwreck Segmentation and Probabilistic Uncertainty Estimation with Sidescan Sonar for Autonomous Underwater Navigation

Lakshmi Annamalai
Centre for Artificial Intelligence and Robotics
Defence Research and Development Organization
Bangalore, India
lakshmia@alum.iisc.ac.in

***Abstract*—Underwater Autonomous Vehicles (UAVs) have enabled efficient exploration of difficult-to-access underwater environments. However, these environments are filled with hazards such as shipwrecks, posing significant risks to the safe navigation of UAVs. Side Scan Sonar (SSS) plays a critical role in segmenting these shipwrecks and avoiding potential accidents. However, UAVs operate under extreme resource limitations, necessiating ultra-lo-computation segmentation models to avoid mission failures due to overheating etc. Existing deep learning approaches for SSS images demand huge GFLOPs, infeasible for real-time UAV inference and curtailing mission endurance. To mitigate this, we introduce PingSEG, a low-compute framework achieving ¡ 1 GLOP, by operating on one-dimensional SSS signals (pings) rather than two-dimensional SSS images. To compensate for the loss of spatial information, we propose a Temporal Belief Update (TBU) module, which boosts the segmentation accuracy, by propagating history embedded in previous segmentation masks to current estimation. The effectiveness of PingSEG has been validated through extensive experiments on the publicly available AI4Shipwrecks side scan sonar dataset. In summary, PingSEG provides a computation and memory-efficient shipwreck segmentation for resource-starved underwater vehicles.**

***Index Terms*—Side scan sonar, semantic segmentation, underwater autonomous vehicles**

I. Introduction

Underwater Autonomous Vehicles (UAVs) have revolutionized ocean exploration, enabling detailed and efficient surveys of vast and often inaccessible underwater environments. Underwater environments are inherently hazardous, with obstacles such as shipwrecks posing significant risks to UAV operations. By detecting and mapping shipwrecks, UAVs can plan and execute safer navigation paths, avoiding collisions and potential damage. This capability not only increases the operational lifespan of UAVs but also ensures the safety and success of their missions. Shipwreck segmentation also extends to other underwater applications such as resource exploration, enhancing the safety of other marine vehicles, maritime archaeology *etc* [31] [32] [33] [34] [35].

Side scan sonar, an advanced sonar technology, plays a pivotal role in detecting and mapping shipwrecks. Side scan sonar (SSS) emits sound waves, also known as acoustic pulses, as a fan-shaped beam that sweeps the ocean floor. When these waves encounter objects or the ocean floor, they are reflected back to the SSS. In SSS processing, the received data, also known as pings, are converted into two-dimensional images, which are further processed to segment shipwrecks. Lightweight methods such as thresholding, region-growing fall short on accuracy, failing to segment fragmented wrecks under acoustic reverberation. Hence, deep learning techniques are required to enhance the reliability of shipwreck segmentation. However, UAV deployments faces acute computational bottlenecks. Traditional deep learning segmentation modles require $10 - 100$ GFLOPs per $\sim 5112 \times 512$, causing thermal throttling, frame drops and premature shutdowns. High-compute models also amplify power draw during peak loads, reducing endurance by 50%.

This work addresses these challenges by reframing 2D image segmentation as 1D ping segmentation, thereby enabling 24+ hour UAV deployments. The proposed low-latent and low-compute shipwreck segmentation, referred to here as PingSEG, operates on one-dimensional side scan sonar signals (pings) rather than two-dimensional sonar images. PingSEG employs a one-dimensional convolutional encoder-decoder architecture followed by a segmentation uncertainty estimation module. The concept is developed based on the hypothesis that the nature of the reflection depends on the material and shape of the object. Hard and rough surfaces, such as shipwrecks, produce strong reflections, while soft or flat surfaces produce weaker reflections. The frequency profile of the reflected wave also depends on several factors, including material composition and the size and shape of the object.

However, direct segmentation from ping data is challenging due to the absence of spatial information. Therefore, the paper provides a post-processing module known as the Temporal Belief Update (TBU) module, which indicates the uncertainty of a samples "shipwreck" probability. This module leverages exponential integration to combine segmentation probability information over time, effectively incorporating spatial context into mask estimation and mitigating the negative impact of the loss in spatial information.

In summary, the advantages of PingSEG are as follows: i) one-dimensional convolution is effective in extracting mean-

ingful features based on the patterns within the sequence, thereby making it an appropriate choice for time-series ping data. Moreover, one-dimensional convolution operates in a single dimension, resulting in fewer computations and memory requirements, ii) PingSEG's ability to process ping data directly without converting pings into images not only reduces latency but also enhances efficiency and responsiveness in real-time applications. This capability makes it well-suited for tasks such as shipwreck segmentation, where immediate analysis of data streams is critical to avoid potential accidents of UAVs.

II. Related Work

In this section, a survey on conventional semantic segmentation networks is provided, followed by a survey on the networks tailored to side scan sonar images.

A. Conventional Segmentation Networks

Long et al. [5] pioneered the use of Fully Convolutional Networks (FCNs) for semantic image segmentation. By replacing the fully connected layer with a convolutional layer, they framed semantic image segmentation as a pixel-wise classification problem. This breakthrough inspired the development of numerous convolutional-based semantic segmentation networks for various applications, including general image segmentation [20] [21] [22] [23] [24] [25], biomedical image segmentation [27] [28], and satellite image segmentation [29] [30].

One notable network is U-Net [6], which adopts an encoder-decoder architecture. The principles of U-Net were further extended in the DeepLab series, from DeepLabv1 to DeepLabv3+ [7] [8] [9] [10]. DeepLabv3+ achieved impressive performance by integrating the encoder-decoder structure with atrous spatial pyramid pooling. Despite its success, DeepLabv3+ requires a large number of data samples to achieve high segmentation accuracy due to its complex structure and numerous parameters. More recently, transformer-based architectures have gained popularity in the field of semantic segmentation [26].

B. Segmentation Networks of Side Scan Sonar Image

The authors in [1] introduced a side-scan sonar image target detection method using the FCN model. They utilized a deep convolutional neural network (CNN), derived from FCN, to segment side-scan sonar (SSS) images. They subsequently applied Markov Random Field (MRF) for post-processing to refine the segmentation results. In a different approach, [2] proposed an unsupervised CNN model for segmenting synthetic aperture sonar images.

Meanwhile, [3] presented ECNet, composed of an encoder-decoder network. A single-stream deep neural network with multiple side outputs was integrated to optimize edge segmentation. Furthermore, [4] developed a real-time sonar image segmentation algorithm based on ENet and MRF, focusing on enhancing efficiency without compromising segmentation accuracy.

III. Proposed Solution

This section presents the proposed solution by first introducing the network architecture of PingSEG. It is followed by the details of the proposed novel pixel-level uncertainty score estimation, derived based on temporal information.

A. Network Architecture

The proposed solution comprises a U-Net network architecture consisting of an encoder and decoder. Given the N samples of a ping, $x(n)$, $n = 0 \dots N$, the goal is to estimate accurate segmentation mask, with minimal latency and computational complexity.

The input sequence $x(n), n = 0 \dots N$ is fed into the encoder $\phi(x(n))$, which extracts features using four blocks with channels 8, 16, 32, 64. Each block consists of two convolutional layers with kernel size 1×3 and ReLU activation in between. The first convolution layer in each block takes as input the output from the previous block, while the second convolution layer takes the output of current block's first convolution layer.

The decoder, $\psi(\phi(x(n)))$, is a mirror image of the encoder, performing upsampling and deconvolution on the extracted features. It is followed by a head consisting of a one-dimensional convolutional layer that produces the segmentation probability, $\mathbb{P}(n)$, for obstacle versus others. The model is trained using binary cross-entropy loss over 100 epochs with the Adam optimizer and a learning rate of 0.01.

B. Temporal Belief Update (TBU) module

Shipwreck segmentation using a side scan sonar signal is a critical application in underwater vehicles. Hence, it is mandatory to enhance the model's interpretability. Towards this, the paper proposes a novel module known as Temporal Belief Update (TBU) module, which acts as a measure of uncertainty regarding the given sample's "shipwreck" probability. The Temporal Belief Update (TBU) module achieves this by applying an exponential moving average to the U-Net segmentation probabilities, $\mathbb{P}$, of the current (t^{th}) and previous ($t-1^{th}$) ping. The updated probability estimate, $\mathbb{P}^+_t(n)$ of the current ping at time t, is calculated as follows:

$$\mathbb{P}^+_t(n) = (1-\alpha)\mathbb{P}_t(n) + \alpha\mathbb{P}^m_{t-1}(n)$$
$$\text{Where, } \mathbb{P}^m_{t-1}(n) = \frac{1}{2W}\sum_{i=n-W}^{n+W}\mathbb{P}^+_{t-1}(i) \tag{1}$$

Where W is the window length. The class of segmentation map of each sample, $\mathcal{S}_t(n)$, of current ping $x_t(n)$ is estimated as obstacle or others if $P^+_t(n) \geq \gamma$ or $P^+_t(n) < \gamma$ respectively, where γ is the threshold.

C. Intuitive Insight

Expanding the Eq. 1 at time instant $t = 3$, we get,

$$\mathbb{P}^+_3(n) = (1-\alpha)\mathbb{P}_3(n) + \alpha\frac{1}{2W}\sum_{i=n-W}^{n+W}\mathbb{P}^+_2(i) \tag{2}$$

Where, $\mathbb{P}_2^+(i)$ is given as follows,

$$\mathbb{P}_2^+(i) = (1-\alpha)\mathbb{P}_2(i) + \alpha\frac{1}{2W}\sum_{j=i-W}^{i+W}\mathbb{P}_1^+(j) \tag{3}$$

Substituting Eq. 3 into Eq. 2, we get,

$$\mathbb{P}_3^+(n) = (1-\alpha)\mathbb{P}_3(n) + \frac{\alpha}{2W}\sum_{i=n-W}^{n+W}\left[(1-\alpha)\mathbb{P}_2(i) + \frac{\alpha}{2W}\sum_{j=i-W}^{i+W}\mathbb{P}_1^+(j)\right] \tag{4}$$

Simplyfying the above equation, we get,

$$\mathbb{P}_3^+(n) = (1-\alpha)\mathbb{P}_3(n) + \frac{\alpha(1-\alpha)}{2W}\sum_{i=n-W}^{n+W}\mathbb{P}_2(i) + \left[\frac{\alpha}{2W}\right]^2\sum_{i=n-W}^{n+W}\sum_{j=i-W}^{i+W}\mathbb{P}_1^+(j) \tag{5}$$

Eq. 5 shows that the segmentation probability of t^{th} ping at sample n depends on segmentation probabilities of $n-W$ to $n+W$ samples of $(t-1)^{th}$ ping and $n-2W$ to $n+2W$ samples of $(t-2)^{th}$ ping, thereby capturing spatial information similar to 2D convolution operation of image segmentation networks. The weightage of segmentation probabilities reduces from $\frac{\alpha(1-\alpha)}{2W}$ to $\left[\frac{\alpha}{2W}\right]^2$ as we go from $(t-1)^{th}$ ping to $(t-2)^{th}$ ping.

D. Representation in $\mathcal{Z}$ domain

Considering a sample at W and ping 2, Eq. 1 becomes,

$$\mathbb{P}_t^+[W] = (1-\alpha)\mathbb{P}_t(W) + \alpha\mathbb{P}_{t-1}^m(W)$$
$$\text{Where, } \mathbb{P}_{t-1}^m[W] = \frac{1}{2W}\sum_{i=0}^{2W}\mathbb{P}_{t-1}^+(i) \tag{6}$$

$\mathbb{P}_{t-1}^m(W)$ is achieved by spatial averaging. In $\mathcal{Z}$ domain, this is represented as,

$$\mathbb{P}_{t-1}^m(\mathcal{Z}) = \frac{1}{2W}[\mathbb{P}_{t-1}^+(\mathcal{Z}) + \mathcal{Z}^{-1}\mathbb{P}_{t-1}^+(\mathcal{Z}) + \ldots + \mathcal{Z}^{-2W}\mathbb{P}_{t-1}^+(\mathcal{Z})] \tag{7}$$

The first equation becomes,

$$\mathbb{P}_t^+(\mathcal{Z}) = (1-\alpha)\mathbb{P}_t(\mathcal{Z}) + \alpha\mathbb{P}_{t-1}^m(\mathcal{Z}) \tag{8}$$

Substituting $\mathbb{P}_{t-1}^m(\mathcal{Z})$ into the above equation,

$$\mathbb{P}_t^+(\mathcal{Z}) = (1-\alpha)\mathbb{P}_t(\mathcal{Z}) + \frac{\alpha}{2W}[\mathbb{P}_{t-1}^+(\mathcal{Z}) + \mathcal{Z}^{-1}\mathbb{P}_{t-1}^+(\mathcal{Z}) + \ldots + \mathcal{Z}^{-2W}\mathbb{P}_{t-1}^+(\mathcal{Z})] \tag{9}$$

The second term indicates averaging across samples per ping and the equation as a whole indicates exponential averaging between samples of current ping and spatially averaged samples of previous ping.

IV. Experiments and Results

This section validates the utility of the proposed method through an ablation study and comparison study, which present the benefits of PingSEG as compared to state-of-the-art 2D segmentation methods.

A. Dataset

The experiments are conducted with a publicly available shipwreck dataset known as AI4Shipwrecks dataset [38]. It consists of 24 shipwreck sites collected from Thunder Bay National Marine Sanctuary in Lake Huron. The EdgeTech 2205 dual frequency $600/1600$ KHz ultra-high resolution side scan sonar was mounted on Iver3 AUV moving with a speed of 2.5 knots. AI4Shipwrecks dataset incorporates time-varying gain (TVG) to ensure that the returned sonar signal is not range-dependent. The samples of the dataset were labeled as two classes, shipwrecks, and others, by a team of three researchers.

B. Comparison with 2D Segmentation Methods

This section brings out the results of the experiments conducted to analyze the performance of the proposed PingSEG relative to the state-of-the-art 2D segmentation networks. Two sets of experiments were conducted to assess PingSEG in terms of both mean Intersection Over Union (mIOU) and compute efficiency. Computational efficiency is assessed based on two metrics: the number of parameters and GFLOPs (Giga Floating Point Operations) required to process an input data matrix of dimensions 480×640. The compute of the algorithms is estimated on Intel(R) Xeon(R) Gold 6226R CPU, 2.90GHz, 64 bit architecture. mIOU is estimated as the average IOU of shipwrecks and others.

The state-of-the-art conventional semantic segmentation networks included in the comparison are SegFormer [11], PSPnet [12], [13], DeepLabV3 (MobileNetV2) [14], [17], DeepLabV3+ (MobileNetV2) [15], [17], FCN (MobileNetV2) [16], [17], and MobileNetV3 [18].

Table I illustrates that PingSEG achieves significant computational efficiency advantages. It employs far fewer parameters compared to the state-of-the-art methods, resulting in lower computational requirements. This efficiency is particularly advantageous in scenarios with limited compute availability, such as those commonly encountered in side-scan sonar applications. It is noteworthy that training such smaller models requires minimal data compared to more complex architectures. This characteristic enhances PingSEG's suitability for applications in data-scarce environments, characteristic of the side-scan sonar domain.

However, it is noted that there exists a modest performance gap between PingSEG and traditional 2D segmentation approaches. This disparity arises primarily due to the inherent limitation of one-dimensional segmentation in capturing detailed spatial relationships and features that are more effectively addressed in 2D segmentation models. By highlighting these comparative results, this section projects

Method	Params (M)	GFLOPs	mIOU
SegFormer [11]	89.49	81.9700	0.75
PSPnet [12] [13]	30.88	46.5800	0.66
DeepLabV3 [14] [17]	87.91	18.3700	0.68
DeepLabV3+ [15] [17]	80.09	15.0200	0.69
FCN [16] [17]	49.09	9.4900	0.66
PingSEG_1	**0.016**	**0.439**	0.5
PingSEG	**0.016**	**0.439**	0.61

TABLE I: Comparison of the proposed method with 2D state-of-the-art semantic segmentation networks in terms of a) compute efficieny: number of parameters (in million) and GFLOPs and b) accuracy in terms of mIOU. The compute of the algorithms is estimated on Intel(R) Xeon(R) Gold 6226R CPU, 2.90GHz, 64 bit architecture. PingSEG_1 indicates the proposed framework without TBU module. The proposed PingSEG performs slightly lower than its 2D counterparts, but the computation efficiency achieved via PingSEG is significant, which is crucial for compute-constrained platforms.

both the strengths and limitatins of PingSEG in the context of underwater autonomous vehicle applications.

C. Qualitative Results

A critical aspect of evaluating side scan sonar image segmentation is assessing the visual quality of the results. To this end, we conducted a qualitative comparison of our approach with several state-of-the-art methods (Figure 5). Notably, our approach excels in segmenting data samples containing signals reflected from the shipwreck itself, accurately isolating these features from the surrounding background. However, we observed limitations when processing samples corresponding to the shadow regions of the shipwreck, where our approach struggles to correctly identify and isolate these areas.

V. Conclusion

This paper proposed PingSEG, a signal-driven low-compute solution for shipwreck segmentation using a side scan sonar signal. PingSEG includes a one-dimensional convolutional encoder-decoder architecture followed by a Temporal Belief Update (TBU) module. Experimental results demonstrated that PingSEG achieves comparable performance to 2D state-of-the-art semantic segmentation networks. Importantly, PingSEG's efficiency is highlighted by its significantly reduced computational demands in terms of the number of model parameters and GFLOPs. This directly addresses the resource constraints encountered in UAV applications. Furthermore, it has been shown empirically that the Temporal Belief Update (TBU) module enhances the performance of PingSEG by compensating for the lack of spatial information in one-dimensional signal segmentation. In conclusion, PingSEG offers a practical solution for shipwreck segmentation for safe navigation of UAVs, while effectively addressing the computational limitations. However, ping processing has limitations such as lesser spatial detail and contextual information, thereby leading to limited interpretive capabilities.

References

[1] Song Y., Zhu Y., Li G., Feng C., He B., Yan T, *Side scan sonar segmentation using deep convolutional neural network*, OCEANS, 2017.

[2] Chen J., Summers J. E, *Deep convolutional neural networks for semi-supervised learning from synthetic aperture sonar (SAS) images*, Proc. Meet. Acoust, 2017.

[3] Wu Wang, Rigall, Li, and Yan, *ECNet: Efficient convolutional networks for side scan sonar image segmentation*, Sensors.

[4] Li K., Yu F., Wang Q., Wu M., Li G., Yan T., He B, *Real-time segmentation of side scan sonar imagery for AUVs*, IEEE Underwater Technology, 2019.

[5] J. Long, E. Shelhamer, T. Darrell, *Fully convolutional networks for semantic segmentation*, IEEE Conference on Computer Vision and Pattern Recognition, 2015.

[6] O. Ronneberger, P. Fischer, T. Brox, *U-Net: convolutional networks for biomedical image segmentation*, Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015.

[7] L. Chen, G. Papandreou, I. Kokkinos, *et al.*, *Semantic image segmentation with deep convolutional nets and fully connected CRFs*, Proceedings of International Conference on Learning Representations, 2015.

[8] L. Chen, G. Papandreou, I. Kokkinos *et al.*, *Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs*, IEEE Trans. Pattern Anal. Mach. Intell. 40(4), 2017.

[9] L. Chen, G. Papandreou, F. Schroff and H. Adam, *Rethinking atrous convolution for semantic image segmentation*, arXiv: 1706.05587, 2017.

[10] L. Chen, Y. Zhu, G. Papandreou *et al.*, *Encoder-decoder with atrous separable convolution for semantic image segmentation*, arXiv:1802. 02611, 2018.

[11] Xie, Enze, et al. "SegFormer: Simple and efficient design for semantic segmentation with transformers." Advances in neural information processing systems, 2021.

[12] Wightman, Ross, Hugo Touvron, and Hervé Jégou. "Resnet strikes back: An improved training procedure in timm." arXiv preprint arXiv:2110.00476, 2021.

[13] Zhao, Hengshuang, et al. "Pyramid scene parsing network." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[14] Chen, Liang-Chieh, et al. "Rethinking atrous convolution for semantic image segmentation." arXiv preprint arXiv:1706.05587, 2017.

[15] Chen, Liang-Chieh, et al. "Encoder-decoder with atrous separable convolution for semantic image segmentation." Proceedings of the European conference on computer vision (ECCV), 2018.

[16] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition, 2015.

[17] Sandler, Mark, *et al.* "Mobilenetv2: Inverted residuals and linear bottlenecks." Proceedings of the IEEE conference on computer vision and pattern recognition, 2018.

[18] Howard, Andrew, *et al.* "Searching for mobilenetv3" Proceedings of the IEEE/CVF international conference on computer vision, 2019.

[19] MMSegmentation Contributors, OpenMMLab Semantic Segmentation Toolbox and Benchmark, "https://github.com/open-mmlab/mmsegmentation", 2020.

[20] Noh. H, Hong. S, Han. B, *Learning Deconvolution Network for Semantic Segmentation*, IEEE International Conference on Computer Vision (ICCV), 2015.

[21] Chen. L, Papandreou. G, Kokkinos. I, Murphy. K, Yuille. A.L, *DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs*, IEEE Trans. Pattern Anal. Mach. Intell, 2018.

[22] Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. arXiv e-prints 2015, arXiv:1511.00561

[23] Lin G, Milan A, Shen C, Reid I, *RefineNet: Multi-Path Refinement Networks for High-Resolution Semantic Segmentation*, IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[24] Chaurasia A, Culurciello E, *LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation*, arXiv:1707.03718, 2017.

[25] Li R, Liu W, Yang L, Sun S, Hu W, Zhang F and Li W, *DeepUNet: A Deep Fully Convolutional Network for Pixel-Level Sea-Land Segmentation*, IEEE JSTARS 2018.

[26] Bowen Cheng, Ishan Misra and *et al.*, "Masked-attention Mask Transformer for Universal Image Segmentation", IEEE Conference on Computer Vision and Pattern Recognition", 2022.

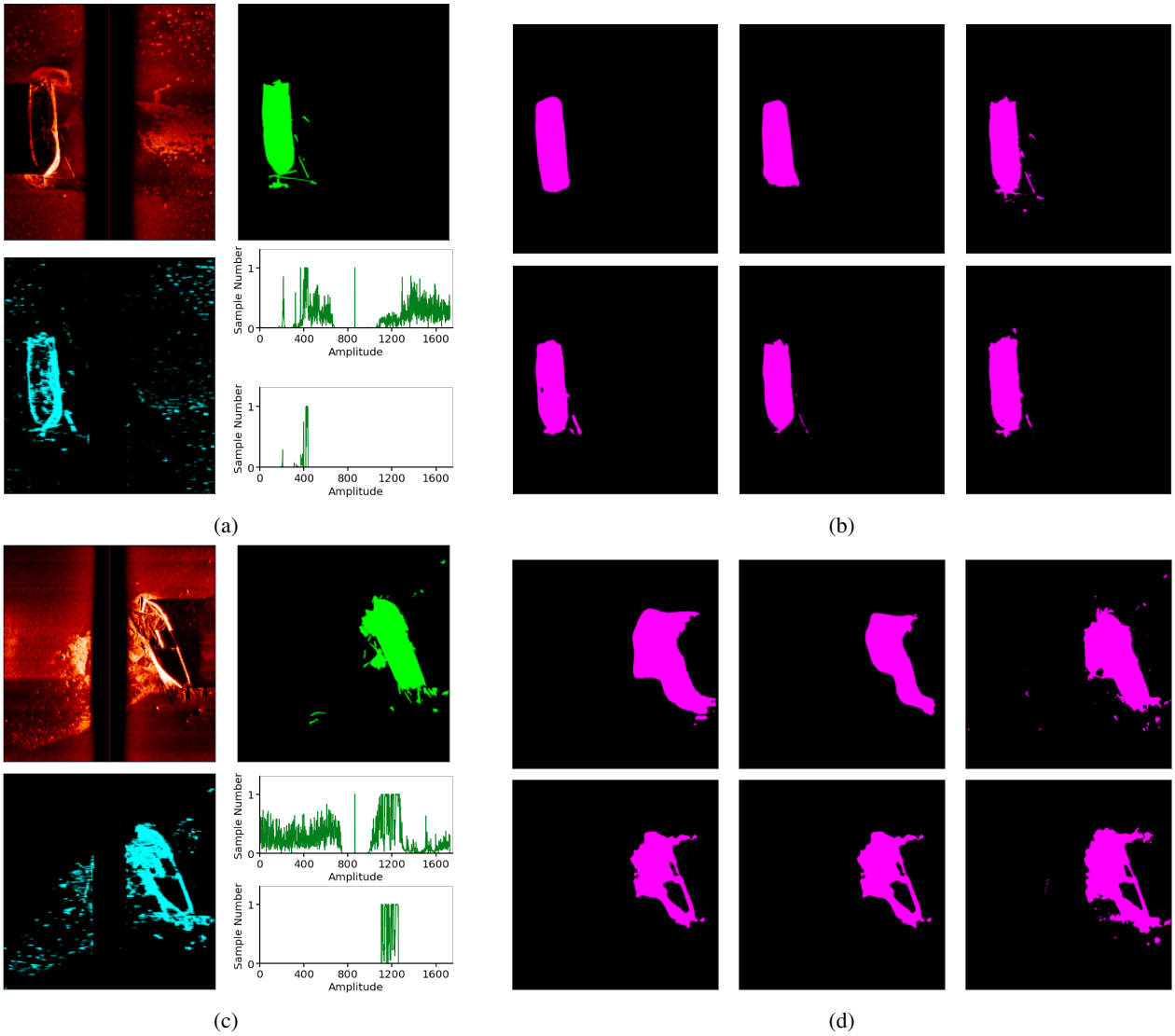

Fig. 1: Visual results of comparison of proposed method with state-of-the-art segmentation methods on two different samples from AI4Shipwrecks dataset (a,b and c,d). (a): Top left - original pings stacked as waterfall image. Top right - ground truth. Bottom left: Segmented pings stacked as waterfall image. Bottom right: Time series plot of raw ping (top) and segmented ping (bottom: Only samples segmented as shipwrecks are retained). (b) and (d) shows segmentation results of DeepLabV3 [14], FCN [16], Mask2Former [39], PSPnet [12], SegFormer [11], DeepLabV3+ [15] (from top left to bottom right).

[27] Ronneberger O, Fischer P and Brox T *U-Net: Convolutional Networks for Biomedical Image Segmentation*, Medical Image Computing and Computer-Assisted Intervention, 2015.

[28] Vigneault D.M, Xie W, and *et al.*, *Ω-Net (Omega-Net): Fully Automatic, Multi-View Cardiac MR Detection, Orientation, and Segmentation with Deep Neural Networks*, Med. Image Anal. 2017.

[29] Castagno J, Atkins E, *Roof Shape Classification from LiDAR and Satellite Image Data Fusion Using Supervised Learning*, Sensors, 2018.

[30] Zhang P, Ke Y, Zhang Z, and *et al.*, *Urban Land Use and Land Cover Classification Using Novel Deep Learning Models Based on High Spatial Resolution Satellite Imagery*, Sensors 2018.

[31] Odegard O, Hansen R.E, and *et al.*, *Archaeological use of Synthetic Aperture Sonar on deepwater wreck sites in Skagerrak*, J. Archaeolog. Sci., 2018.

[32] Cochrane G.R, Lafferty K.D, *Use of acoustic classification of sidescan sonar data for mapping benthic habitat in the Northern Channel Islands*, Cont. Shelf Res, 2002.

[33] Park C, Kim Y, and *et al.*, *Development of a 2 MHz Sonar Sensor for Inspection of Bridge Substructures*, Sensors, 2018.

[34] Barngrover C, Althoff A, Deguzman P and Kastner R, *A Brain–Computer Interface (BCI) for the Detection of Mine-Like Objects in Sidescan Sonar Imagery*, IEEE J. Oceanic Eng., 2016.

[35] Fatan M, Daliri M.R and Shahri A.M, *Underwater cable detection in the images using edge classification based on texture information*, Measurement, 2016.

[36] Chen Z, Duan Y and *et al.*, *Vision Transformer Adapter for Dense Prdeictions*, ArXiv:2205.08534, 2023.

[37] Wang J, Sun K *et al.*, *Deep High REsolution Representation Learning for Visual Recognition*, ArXiv:1908.07919, 2020.

[38] Advaith Sethuraman, Anja Sheppard *et al.*, *Machine Learning for Shipwreck Segmentation from Side Scan Sonar Imagery: Dataset and Benchmark*, arXiv: 2401.1454, 2024

[39] Cheng, Bowen, et al. "Masked-attention mask transformer for universal image segmentation." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.