TOLIE, H.F., REN, J., CHEN, R. and ZHAO, H. 2024. Protecting visual data privacy in offshore industry via underwater image inpainting. In *Proceedings of the 9th International conference on image, vision and computing 2024 (ICIVC 2024), 15-17 July 2024, Suzhou, China*. Piscataway: IEEE [online], pages 281-286. Available from: https://doi.org/10.1109/ICIVC61627.2024.10837433

Protecting visual data privacy in offshore industry via underwater image inpainting.

TOLIE, H.F., REN, J., CHEN, R. and ZHAO, H.

2024

2024 the 9th International Conference on Image, Vision and Computing

Protecting Visual Data Privacy in Offshore Industry via Underwater Image Inpainting

Hamidreza Farhadi Tolie
National Subsea Centre
Robert Gordon University
Aberdeen, UK
h.farhadi-tolie@rgu.ac.uk

Jinchang Ren*
National Subsea Centre
Robert Gordon University
Aberdeen, UK
j.ren@rgu.ac.uk
*Corresponding author

Rongjun Chen
School of Computer Science
Guangdong Polytechnic Normal University
Guangzhou, 610556, China
crj321@163.com

Huimin Zhao
School of Computer Science
Guangdong Polytechnic Normal University
Guangzhou, 610556, China
zhaohuimin@gpnu.edu.cn

***Abstract*—Leveraging advanced artificial intelligence (AI) methodologies offers the advantage of incorporating multiple expert viewpoints, thereby facilitating a more comprehensive inspection of underwater infrastructure. However, the implementation of AI techniques in subsea tasks is hindered by the lack of extensive and diverse datasets required for effective training and inference of the AI models, emphasizing the vital need for enhanced data sharing practices within the offshore sector. The sensitive textual information within underwater survey data, such as site geolocations, water depths, mission-specific details, timestamps, and third-party data, necessitate a balanced approach to data privacy. To address this, we propose the integration of cutting-edge text detection and image inpainting techniques. These methodologies enable the identification and subsequent removal of textual regions from images while preserving the quality and natural appearance of the images. Experimental results validate the efficacy of our proposed approach in simultaneously preserving the visual quality and protecting the privacy. The removal of detected textual regions from images demonstrates less distortions, underscoring the potential of this methodology for application in offshore industry settings. This study contributes to the ongoing discourse regarding data privacy in underwater surveys, offering a viable solution to balance information sharing with confidentiality concerns.**

***Index Terms*—Underwater image, underwater data privacy, text detection, image inpainting**

I. Introduction

The advent of artificial intelligence (AI) technologies and their increased usages in various domains have attracted the offshore industry to use the potential of the AI for automating their operations including infrastructure inspection [1], [2] and maintenance [3], [4] as well as anomaly detection for fault diagnosis [5]. AI methods are generally trained to learn the mapping between the sensed data from the environment and their relative subjective assignments. These mappings may involve categorizing objects [6], identifying anomalies or defects [7], or analyzing scalar values from the measurements. Subjective assignments, especially in specialized fields like subsea infrastructure anomalies, are evaluated by field experts. This allows end users to leverage multiple expert perspectives, resulting in saving of the time and cost. However, the effective and robust training of such AI models requires a large and diverse amount of data, which is not usually available.

In the course of a subsea survey mission, various types of data are recorded and subsequently analyzed and annotated by experts. These recorded data predominantly consist of optical and sound navigation and ranging (SONAR) images or videos, which can be utilized by AI models to support data analysis and decision making. As such data often contains rich and sensitive textual information illustrating the location and depth of survey sites, mission's timestamps and third-party details etc., setting a barrier and preventing industries from freely sharing the data for safety and privacy concerns.

To mitigate such concerns in properly addressing the rather infrequent anomaly cases with the subsea infrastructures even with low visibility/contrast and non-uniform color cast [8], our study explores secure underwater visual data sharing by employing image inpainting techniques to remove textual information from images while preserving the quality and visual consistency. The major contributions of this research are highlighted as follows:

- Comparison of various text detection methods for effective extraction of the textual regions within a given underwater image.
- Removing the textual regions using state-of-the-art image inpainting methods, where various inpainting methods are analyzed for effective reconstruction of the removed textual regions for data privacy.
- To derive the best workflow for image-inpainting based text removal from underwater images.

Jinchang Ren is also with the School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou, 610556, China.

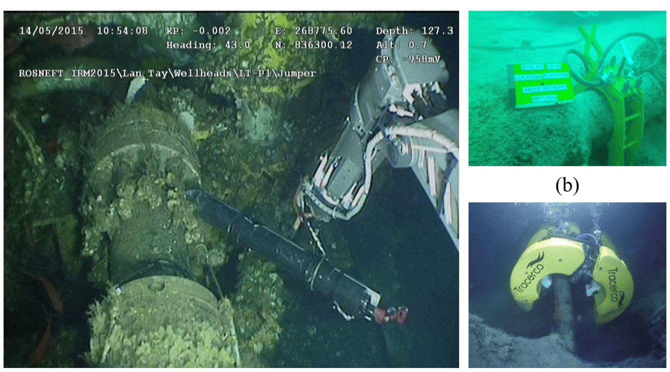
The rest of the manuscript is organized as follows. Section II briefly reviews the related work, Section III presents the proposed methodology. Section IV provides both the qualitative and quantitative analyses . Finally, Section V draws the conclusion with future work discussed.

II. Related work

In the context of underwater survey, privacy concerns are predominantly associated with the insertion of textual information in the data. To protect the privacy of industrial data, the texts need to be identified and then removed whilst the removed text regions are reconstructed using the background pixels for visual consistency and coherency, i.e. image inpainting. Relevant techniques in this pipeline is reviewed in detail as follows.

A. Text detection

Text detection has a critical role in extracting the texts from images, contributing widely across various industries. Text detection is the main component in optical character recognition (OCR) for locating the position and range of the texts [9], where image enhancement by noise removal is also focused for more accurate recognition [10]. Segmentation is then used to distinguish the graphics from text, followed by feature extraction from the textual regions. With these features, characters can be recognized in machine-learning models, where post-processing can also be applied to refine the results by correcting the misspellings in extracted text.



(a) (b) (c)

Fig. 1. Sample underwater images containing inserted textual information (a) and manufacturer details (b) and (c).

Given the primary focus of our research on detecting textual regions in underwater visuals, we focus on text detection and segmentation methods developed for natural scene images within and beyond the conventional OCR methodologies. As seen in Fig. 1, the underwater images may contain inserted textual information and also some physically written details about the equipment and infrastructures' manufacturer or contractors. Therefore, employing the text detectors for natural scenes allows us to not only identify inserted textual information within images but also to detect the written textual data, such as names and model information, presented in various graphical forms and shapes as labels on remotely operated vehicles (ROV) or infrastructures. Employing these approaches enables us to improve the privacy protection for both mission-related data and third-party information.

The traditional natural-scene-image text detectors mainly use hand-crafted features including but not limited to maximally stable extremal regions (MSER) [11] and stroke width transform (SWT) [12]. As these text detectors rely on certain assumptions about the characteristics of the textual regions, they have limited adaptability to variations in text appearances such as font families and the texts orientation in an image. Moreover, they are less robust to the noise and complex backgrounds. Fig. 2 illustrates the detected textual regions using the SWT [12] method for the sample image shown in Fig. 1 panel (a). As seen, there are many incorrect and missing detection in the image, which demonstrates the low performance of hand-crafted feature-based conventional methods on complex underwater images.

Fig. 2. Detected textual regions using the SWT [12] technique from the sample image in Fig. 1(a).

With the emerging deep learning techniques, these issues have been well addressed, using the well-known object detection or segmentation architectures such as Faster R-CNN [13] and fully convolutional networks (FCN) [14] to learn diverse features from the data. This has allowed them to better adapt to various text appearances and become more robust to the noise and complicated backgrounds. Zhou *et al.* proposed the efficient and accurate scene text (EAST) detector [15] to detect texts in various orientations, sizes, and aspect ratios, where FCN [14] was utilized to support multi-scale feature extraction. Baek *et al.* proposed the character region awareness for text (CRAFT) detection [16], which detects the texts by exploring each character and the affinity between the characters. Using the convolutional neural networks (CNN) to extract features, CRAFT predicts the textual regions by introducing the quadrangles rather than bounding boxes to enclose the individual characters, which can also predict the link between the adjacent characters to form the final detection regions.

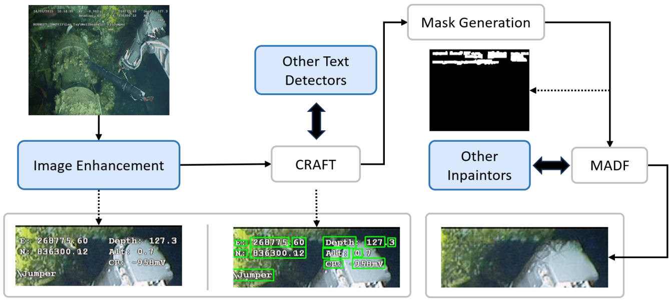


Fig. 3. Framework of the proposed method for underwater data privacy protection.

B. *Image inpainting*

Inpainting is an image processing technique used to fill in or replace missing information within an image. Inpainting algorithms can be divided into four main categories: statistical-, partial differential equation (PDE)-, exemplar-, and the deep learning-based methods. The statistical-based methods estimate the most-likely values for the missing pixels based on the statistical properties of the surrounding pixels using the probability distributions. These methods struggle to deal with the images with complex textures and patterns like the natural scene images as accurately estimating the exact distribution of the data is very challenging [17]. The PDE-based methods fill the missing areas by propagating the information from the known parts of the image using the smoothness priors [18] [19], though, corporation of smoothness priors may lead to a blurred appearance in the filled regions. The third category, exemplar-based methods, inpaint the missing regions using information from the matched similar regions [20]. For the fourth category, deep learning-based methods learn the patterns and structures within the image and then perform the inpainting by generating visually coherent content for the missing areas, by taking the surrounding context into account [21] [22] [23].

Although the existing methods, especially the deep learning-based methods, have made remarkable advancements in producing visually pleasant content for image inpainting, they rarely consider the mask information derived from text detection. This mask-unawareness makes them to treat all missing regions equally in the process of feature extraction despite of their various shapes and sizes. To tackle this issue and improve the visual quality of the inpainted images, Zhu *et al.* proposed the mask-aware dynamic filtering (MADF) [24] method to learn multi-scale features from the missing regions.

Generating mask images to fully cover textual regions is not always straightforward, particularly when characters may not have been fully detected due to existing distortions in the image. To address this challenge, we propose employing set of morphological operations to enhance coverage of text within the image. This approach aims to yield more natural-looking inpainting results, especially for underwater images.

III. The Proposed Method

In this section, we present the methodology we have utilized for inpainting-based text removal for barrier-free data sharing and privacy protection of underwater images. As seen in the framework shown in Fig. 3, the proposed method has four main modules: image enhancement, text detector, mask image generator, and the inpainting module, detailed below.

A. *Pre-processing: image enhancement*

Although underwater images belong to the category of natural scene images, their color distribution is often distorted, resulting in a predominant blueish or greenish appearance [8]. While employing natural scene text detectors can largely identify textual regions within these images, additional pre-processing is essential to enhance the color distribution and address other degradations such as noise, low contrast, loss of sharpness, and lens aberrations resulting from placing cameras in underwater housings. Improving the quality of underwater images results in a more natural appearance, enabling the effective application of pre-trained natural scene text detectors on these images. It is important to note that this pre-processing, i.e., underwater image enhancement, differs from the methods employed within text detection processes. To enhance the quality of the underwater images as an additional pre-processing step, we have utilized our recently proposed deep learning-based method namely, deep inception and channel-wise attention modules (DICAM) [25], which improves the color cast of images and addresses the blurriness, low contrast, and loss of content.

B. *Natural scene text detectors*

To achieve more effective extraction of textual information and mitigate the limitations of conventional text detectors, we employed pre-trained deep learning methods, including EAST and CRAFT to identify textual regions in underwater images. The technical distinctions between these two are noteworthy: EAST directly identifies textual regions within an image, while CRAFT initially identifies individual characters and then links them together to form textual regions. Furthermore, EAST is limited to predicting multi-oriented texts, whereas CRAFT surpasses in predicting both curved and arbitrarily shaped texts in addition to multi-oriented ones. Aligning with our objectives and the need to protect the third-party information, we opted for the CRAFT detector as the main detector in our approach to remove the mentioned information wherever visible in the images.

C. *Mask generator*

In order to effectively remove textual regions within images using state-of-the-art inpainting methods, we have designed a sequence of morphological operations to generate the mask image for each detected bounding box. This process begins with the application of a dilation operation, which expands the boundaries of the textual regions to enhance the inpainting effect.

$$DI = roi \oplus se \tag{1}$$

In this equation, DI represents the dilated image, roi is the region of interest (bounding box), se denotes the structuring element, and $\oplus$ indicates the dilation operation. For our study, we have empirically selected a square structuring element with a size of 3×3.

Subsequently, to address any missing text regions that may result from transmission issues or low-quality screen recordings, we employ a morphological closing operation using the same structuring element:

$$mask = DI \bullet se = (roi \oplus se) \ominus se \tag{2}$$

Here, $\bullet$ denotes the morphological closing operation, and $\ominus$ signifies the erosion operation. This step ensures that gaps within the text regions are filled. To further refine the extraction of the textual regions, we apply the dilation operation once more, as described in Eq. 1, using the same structuring element. By developing this specific sequence of morphological operations, we have enhanced the accuracy and effectiveness of text region detection and removal, thus contributing to the improvement of inpainting methods.

D. Inpainting

To remove the detected textual information, we explored the application of three inpainting methods including the two PDE-based approaches, namely Navier-Stroke (NS) [18] and Telea [19], along with the MADF mask-aware filtering method [24]. Note that the NS and Telea methods are accessible within the OpenCV library, while MADF is available online at https://github.com/MADF-inpainting/Pytorch-MADF. MADF has been trained over three different datasets, among them we have chosen the pre-trained model on the Places2 [26] dataset that contains a wide diverse range of images from 365 scene categories.

IV. Experimental Results

In this section, we discuss the experimental results to validate the effectiveness of the employed text detectors and the applied inpainting settings. The experiments involved text detection and inpainting processes carried out on 126 underwater images sourced from several videos available on Google and YouTube. Sample images used in our experiments along with the results are illustrated in Figure 4. It's worth noting that these images were collected at various resolutions and qualities, aiming to assess the generalization ability of the introduced method.

A. Qualitative performance analysis: text detectors

We conducted a qualitative analysis of the performance of the EAST and CRAFT detectors both with and without the application of any image enhancement methods. The outcomes are presented in Figures 5-6. In each figure row, the original image, the text detection result, and the result after applying enhancement are shown for comparison.

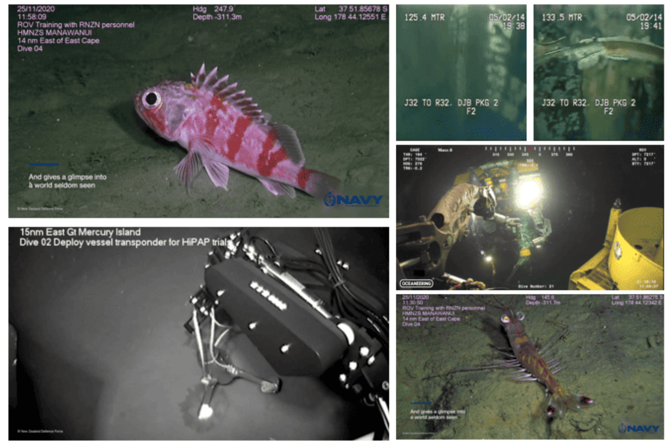


Fig. 4. Sample underwater images collected from online sources.

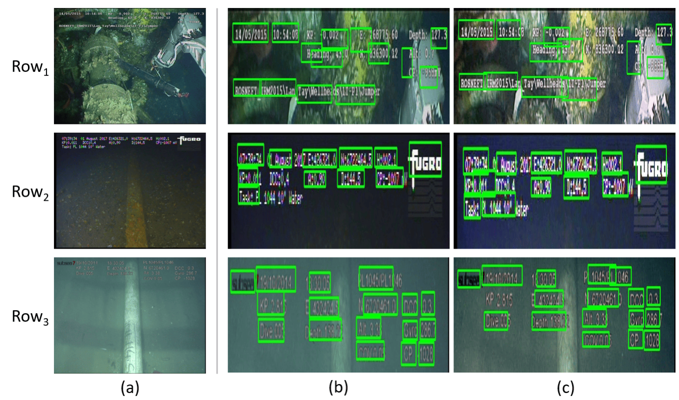


Fig. 5. Sample underwater images with their corresponding textual regions detected using the EAST detector on the raw (a) and enhanced (b) images.

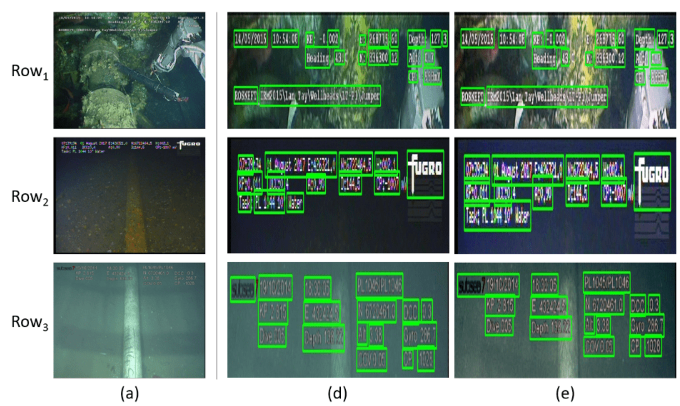


Fig. 6. Sample underwater images with their corresponding textual regions detected using the CRAFT detector on the raw (a) and enhanced (b) images.

For better visual effect, we cropped the top part of the images where textual information is present. In Figure 5, it is evident that the EAST detector leaves some textual regions undetected across all sample images. Even with image enhancement, the impact on improving the results appears minimal. Conversely, for the CRAFT detector, as shown in

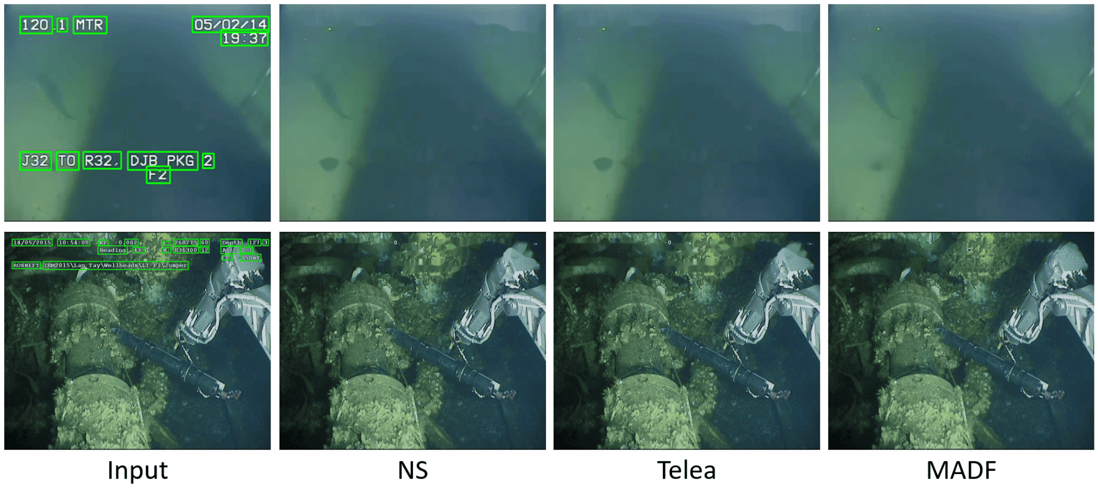


Fig. 7. Sample underwater images with their corresponding inpainted images generated by the NS, Telea, and MADF methods, from left to right, respectively.

Figure 6, there is a notable improvement in detection performance for the same group of sample images. Applying image enhancement to CRAFT can not only enhance the detection performance but also improve the connectivity of the textual regions. Additionally, in the second row of both images, featuring arbitrary-shaped text (i.e., the Fugro logo), CRAFT demonstrates full detection capability, showcasing its superiority in handling diverse text shapes.

B. Qualitative performance analysis: Inpainting methods

In addition to evaluating the text detectors, we have also illustrated the results of applying the NS, Telea, and MADF inpainting methods to reconstruct the removed text regions. Figure 7 presents the results of inpainted sample images using the aforementioned methods. As seen, unlike the NS and Telea methods, MADF does not form blurring effects in the inpainted image. This in particular helps to preserve the quality and naturalness of the acquired underwater images.

C. Quantitative performance analysis

In addition to the previously discussed qualitative analysis, we employed the underwater image quality measure (UIQM) [27] and naturalness image quality evaluator (NIQE) [28] to assess the quality and naturalness of both the original and the inpainted images with respect to the EAST and CRAFT text detectors. The results, as presented in Table I, highlight that MADF demonstrates the highest performance based on the UIQM and NIQE metrics, in particular when CRAFT has been utilized as the main text detector.

Overall, the proposed textual data protection strategy not only safeguards confidential information within images but also preserves the image quality. Notably, the results from the NIQE metric indicate that the combination of the CRAFT text detector and MADF inpainting enhances the naturalness of underwater images, imparting a more realistic appearance compared to the original one.

Moreover, to validate the impact of the enhancement on the proposed strategy (CRAFT + MADF), we also reported the UIQM and NIQE metrics results in Table II. The results show that the applied enhancement technique reduces the naturalness of the inpainted image, however, it has significantly increased the UIQM metric from 1.80 to 2.03. This is mainly because the current enhancement methods are designed to improve the contrast, brightness, color richness of the images rather than improving the readability of textual information. Meanwhile, the NIQE was degraded from 4.89 to 4.58. This reduction can be attributed to the fact that the NIQE method heavily relies on the statistical properties of the image. When enhancements are applied, these properties shift towards improved contrast, brightness, and richer color cast, thereby altering the basis upon image quality score obtained by the NIQE metric.

TABLE I

Performance comparison of the inpainting methods in terms of image quality evaluators. The best result is highlighted in bold.

Text Detector	EAST		CRAFT	
Inpainting Method	UIQM	NIQE	UIQM	NIQE
NS	1.60	4.5173	1.56	4.5965
Telea	1.61	4.5259	1.54	4.6230
MADF	**1.84**	4.7871	1.80	**4.8904**
Original Images	UIQM:	1.70	NIQE:	4.5976

TABLE II

Impact of enhancement on the proposed strategy.

Proposed Strategy	UIQM	NIQE
Enhancement + CRAFT + MADF	2.03	4.58

V. Conclusion

In this paper, we introduced a strategy aimed at protecting the data privacy in underwater visuals by leveraging cutting-edge text detectors and inpainting techniques. We evaluated the impact and efficacy of text detectors like EAST and CRAFT, along with inpainting methods such as NS, Telea, and MADF, both qualitatively and quantitatively. The findings distinctly showcased the superior performance of the CRAFT text detector and the MADF inpainting method. As a result, our strategy primarily relies on these two approaches to ensure effective text removal and inpainting based reconstruction for privacy protection. Through experimental validation, we highlighted the profound impact of inpainting on maintaining the image quality, enhancing naturalness, and, crucially, securing data privacy within the underwater visuals.

Our future work will concentrate on refining existing text detectors and OCR methodologies to facilitate automatic indexing and retrieval of underwater visuals. This endeavour will necessitate modifications to existing enhancement methods aimed at improving readability within textual regions. Additionally, we plan to explore the implications of selective text removal as a strategy for targeted privacy protection.

Acknowledgement

This work is partially supported by the SeaSense project, funded by the Net Zero Technology Centre, UK, Guangdong Provincial Department of Education, Innovation and Strengthening University Project (2022ZDJS015), and Guangdong Provincial Department of Education Special Project (CXTD016).

References

[1] M. Ho, S. El-Borgi, D. Patil, and G. Song, "Inspection and monitoring systems subsea pipelines: A review paper," *Structural Health Monitoring*, vol. 19, no. 2, pp. 606–645, 2020.

[2] W. Tang, D. Flynn, and V. Robu, "Sensing technologies and artificial intelligence for subsea power cable asset management," in *2021 IEEE International Conference on Prognostics and Health Management (ICPHM)*, pp. 1–6, IEEE, 2021.

[3] M. Fumagalli and E. Simetti, "Robotic technologies for predictive maintenance of assets and infrastructure [from the guest editors]," *IEEE Robotics & Automation Magazine*, vol. 25, no. 4, pp. 9–10, 2018.

[4] G. Arcangeletti, L. Gambella, E. Aloigi, A. Radicioni, M. Novello, F. Bacati, and S. Shaiek, "Subsea processing systems predictive maintenance enabled by condition monitoring system," in *Offshore Technology Conference*, p. D011S013R002, OTC, 2020.

[5] P. Wu, C. A. Harris, G. Salavasidis, A. Lorenzo-Lopez, I. Kamarudzaman, A. B. Phillips, G. Thomas, and E. Anderlini, "Unsupervised anomaly detection for underwater gliders using generative adversarial networks," *Engineering Applications of Artificial Intelligence*, vol. 104, p. 104379, 2021.

[6] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 11, pp. 3212–3232, 2019.

[7] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM computing surveys (CSUR)*, vol. 41, no. 3, pp. 1–58, 2009.

[8] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, "An underwater image enhancement benchmark dataset and beyond," *IEEE Transactions on Image Processing*, vol. 29, pp. 4376–4389, 2019.

[9] S. Vijayarani and A. Sakila, "Performance comparison of ocr tools," *International Journal of UbiComp (IJU)*, vol. 6, no. 3, pp. 19–30, 2015.

[10] Y. Alginahi *et al.*, "Preprocessing techniques in character recognition," *Character recognition*, vol. 1, pp. 1–19, 2010.

[11] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and vision computing*, vol. 22, no. 10, pp. 761–767, 2004.

[12] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," in *2010 IEEE computer society conference on computer vision and pattern recognition*, pp. 2963–2970, IEEE, 2010.

[13] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.

[14] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440, 2015.

[15] X. Zhou, C. Yao, H. Wen, Y. Wang, S. Zhou, W. He, and J. Liang, "East: an efficient and accurate scene text detector," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 5551–5560, 2017.

[16] Y. Baek, B. Lee, D. Han, S. Yun, and H. Lee, "Character region awareness for text detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9365–9374, 2019.

[17] Levin and Zomet, "Learning how to inpaint from global image statistics," in *Proceedings Ninth IEEE international conference on computer vision*, pp. 305–312, IEEE, 2003.

[18] M. Bertalmio, A. L. Bertozzi, and G. Sapiro, "Navier-stokes, fluid dynamics, and image and video inpainting," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, pp. I–I, IEEE, 2001.

[19] A. Telea, "An image inpainting technique based on the fast marching method," *Journal of graphics tools*, vol. 9, no. 1, pp. 23–34, 2004.

[20] A. Criminisi, P. Pérez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Transactions on image processing*, vol. 13, no. 9, pp. 1200–1212, 2004.

[21] R. A. Yeh, C. Chen, T. Yian Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do, "Semantic image inpainting with deep generative models," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5485–5493, 2017.

[22] Y. Zeng, J. Fu, H. Chao, and B. Guo, "Learning pyramid-context encoder network for high-quality image inpainting," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1486–1494, 2019.

[23] Y. Ren, X. Yu, R. Zhang, T. H. Li, S. Liu, and G. Li, "Structureflow: Image inpainting via structure-aware appearance flow," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 181–190, 2019.

[24] M. Zhu, D. He, X. Li, C. Li, F. Li, X. Liu, E. Ding, and Z. Zhang, "Image inpainting by end-to-end cascaded refinement with mask awareness," *IEEE Transactions on Image Processing*, vol. 30, pp. 4855–4866, 2021.

[25] H. F. Tolie, J. Ren, and E. Elyan, "Dicam: Deep inception and channel-wise attention modules for underwater image enhancement," *Neurocomputing*, vol. 584, p. 127585, 2024.

[26] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 million image database for scene recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 6, pp. 1452–1464, 2017.

[27] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures," *IEEE Journal of Oceanic Engineering*, vol. 41, no. 3, pp. 541–551, 2015.

[28] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal processing letters*, vol. 20, no. 3, pp. 209–212, 2012.