PIRIE, C. and MORENO-GARCIA, C.F. 2021. Image pre-processing and segmentation for real-time subsea corrosion inspection. In *Iliadis, L., Macintyre, J., Jayne, C. and Pimenidis, E. (eds.). Proceedings of the 22nd Engineering applications of neural networks conference (EANN2021), 25-27 June 2021, Halkidiki, Greece*. Proceedings of the International Neural Networks Society (INNS), 3. Cham: Springer [online], pages 220-231. Available from: https://doi.org/10.1007/978-3-030-80568-5_19

Image pre-processing and segmentation for real-time subsea corrosion inspection.

PIRIE, C. and MORENO-GARCIA, C.F.

2021

Image Pre-processing and Segmentation for Real-Time Subsea Corrosion Inspection

Craig Pirie[1] and Carlos Francisco Moreno-Garcia[1[0000-0001-7218-9023]]

[1]Robert Gordon University, Garthdee Road, Aberdeen, UK, AB10 7GJ

c.pirie4@rgu.ac.uk
c.moreno-garcia@rgu.ac.uk

Abstract. Inspection engineering is a highly important field in the Oil & Gas sector for analysing the health of offshore assets. Corrosion, a naturally occurring phenomenon, arises as a result of a chemical reaction between a metal and its environment, causing it to degrade over time. Costing the global economy an estimated US $2.5 Trillion per annum, the destructive nature of corrosion is evident. Following the downturn endured by the industry in recent times, the need to combat corrosion is escalated, as companies look to cut costs by increasing efficiency of operations without compromising critical processes. This paper presents a step towards automating solutions for real-time inspection using state-of-the-art computer vision and deep learning techniques. Experiments concluded that there is potential for the application of computer vision in the inspection domain. In particular, Mask R-CNN applied on the original images (i.e. without any form of pre-processing) was found to be most viable solution, with the results showing a mAP of 77.1%.

Keywords: corrosion, inspection, subsea, segmentation, real-time recognition

1 Introduction

Oil & Gas companies have come into a period of hardship in recent times following the decline of fossil-fuel prices[1]. This has led the industry to look for ways to save money and increase the efficiency of their operations without compromising the safety of essential processes. Consequently, it has highlighted the need for the inspection process to be assisted by modern technology to leverage the abilities of the machine to minimise human input and boost productivity for the engineer. Implementing the proposed methods should save companies time and money.

Currently, the most common subsea inspection methods used in this industry include smart pigging with video cameras and deploying a Remotely Operated Vehicle (ROV) that feed live video footage back to the controller [1]. An engineer can then examine the captured footage for signs of corrosion and implement the appropriate measures to correct the occurrence. This can be a laborious and tedious task, increasing both cost

[1] https://oilandgasuk.co.uk/oil-gas-uk-figures-show-impact-of-oil-price-downturn-on-jobs/

and the risk of corrosion being overlooked due to human error. Such a task is well suited for automation, and image recognition may stand as the solution for this problem. Advancements made in deep neural networks since the turn of the century have enhanced the reliability and capabilities of image recognition. Systems are getting smarter, with the ability to make predictions quicker than ever before. New techniques have been developed to progress from detecting only presence of an object in an image, to localise objects, classify multiple instances of objects and segment such objects. These methods have seen success in applications such as in disease diagnosis [2], real-time detection of helmets worn by construction workers [3], self-driving cars [4] and underwater imagery analysis [5]. A system that could detect, in real-time, the pixel-perfect location of each instance of corrosion in an image would be ideal for this scenario. Thus, this paper will investigate the application of state-of-the-art image recognition algorithms in the corrosion inspection domain. A huge hurdle exists in the application of computer vision underwater however, being the diminished quality of imagery below the water surface. The aquatic environment is extremely noisy, where problems such as light absorption and scattering can limit visibility by up to a few metres [6]. These factors, together with issues such as low contrast, diminished colours and blur, may suffocate the performance of subsea image recognition systems [7]. The application of appropriate image pre-processing techniques will also be investigated in this paper and the effect they have on the accuracy of the trained recognition models will be evaluated.

2 Related Work

Before working with the images, related literature [5] suggests pre-processing them in a way that the areas of interest can be highlighted. In this section, we will firstly discuss the most popular approaches to that aim. Afterwards, we will briefly discuss the particularities and challenges posed by real-time detection and instance segmentation.

2.1 Pre-Processing

Contrast Correction. Enhancement of contrast is an essential player in the improvement of visual quality for computer vision. In conditions such as those in the underwater environment, where overall luminescence is insufficient, the details of the images or video features will be obscured [8]. Global Histogram Equalisation (GHE) is one of the most common techniques for improving the effects of contrast in an image [9]. For context, an image histogram is a visual representation of the colour intensities within an image [10]. This equalization method involves redistributing grey levels within an image to achieve a uniform histogram [11]. It does this by efficiently distributing the most frequent contrast intensities across the histogram. However, GHE does not perform well when the contrast is not uniform across the image. When the picture contains patches significantly lighter or darker than others, the contrast enhancement is sub-optimal [12]. Marine imagery is not perfect when it comes to the balance of contrast, therefore GHE is assumed to be unsuitable for the underwater domain. Adaptive Histogram Equalization (AHE) is a step forward to solve the aforementioned problems.

This new method differs from GHE as it generates several histograms for different locations within the image [13]. AHE enhances each pixel by transforming the histogram of sections in the image by comparing grey levels of that pixel with those in its surroundings [12]. Although this technique solves the non-uniform contrast problem, it is common for this algorithm to amplify noise within an image, thus degrading the quality of the display [14].

Contrast Limited Adaptive Histogram Equalisation (CLAHE) was originally designed for the improvement of medical images where contrast is insufficient [15]. The success of applying this algorithm in the subaqueous domain was documented by Hitam et al. [16]. CLAHE builds on from AHE by introducing the concept of contrast limiting to combat the noise amplification problem. This concept limits the magnification of noise in an image by 'clipping' the histogram at a user-defined value [17]. A lower clipping value outputs a smoother image but with contrast enhancement becoming minimal. Whereas a higher clipping value offers a more detailed image, the picture becomes distorted due to higher levels of noise. Clipping, therefore, gives the power to the user to decide trade-off between contrast and noise. Enhancing contrast in a coloured image using histogram equalisation is a more complex task than that of grey images, due to the three colour components Red, Green & Blue [18]. To produce an enhanced colour image, CLAHE can be applied to each colour component individually, then once combined, the desired results are achieved such as in Figure 1 below.

Fig. 1. Original image (Left) and image with CLAHE Applied (Right).

Colour Normalisation. Underwater images usually have a blue or green tint due to the presence of a high percentage of blue pixels, followed by green, and lastly, red [19]. Unfortunately, achieving a high-quality image requires an equal distribution of these colours across the image – a luxury not provided by the underwater environment [20]. This is due to the absorption properties of water, where light of increasingly longer wavelengths is absorbed at deeper depths underwater. Without correcting the image defects caused by this problem, there may be difficulty training image recognition models using these images [21]. One of the simplest colour equalisation algorithms available is called 'Grey World'. It assumes that the average of all red, green and blue pixels in an image is grey [22]. By taking the average of the three components independently in the colour-casted image, this method can find the illumining colour (i.e., the colour tinting the image). Then, it uses the difference between the illumining and grey values to scale each of the RGB components to fit the original assumption [23]. Ancuti et al. [24] explored the use of this algorithm in the subaquatic scenario. They found that the method worked reasonably well for bringing balance to reasonably distorted underwater images and can improve performance of segmentation tasks. However, as the image quality deteriorates, so too does the effectiveness of the strategy. Their investigations also uncovered that this algorithm was prone to welcoming artefacts of red into the image, due to over-compensation of the low levels of red pixels.

Fig. 2. Original image (Left) and image with 'Gray World' Applied (Right).

Colour constancy theory is the notion that a colour should appear the same, even under different coloured illuminations [25]. For example, under both conditions of white light from a Light Emitting Diode (LED), or red light during a sunset, an apple to humans will always appear to be red. Retinex theory [26] explains the ability for the human eye to achieve colour constancy through the retina and cortex's perception of luminance and colour. Success has been noted in mimicking Retinex theory to counter the effects of illumination caused in digital images. Its application in the subsea domain has been proven to be successful [27], and NASA have implemented Retinex processing at near real-time speeds [28]. Retinex algorithms profit from the chance of removing undesirable illumination effects and enhancing image edge [29]. Single Scale Retinex is the most basic algorithm in the Retinex chain. In SSR the illumination is estimated by applying a linear low-pass filter for an input colour image. The enhanced image is then produced by subtracting the 2D convolution of Gaussian surround function and original image of i^{th} component [30].

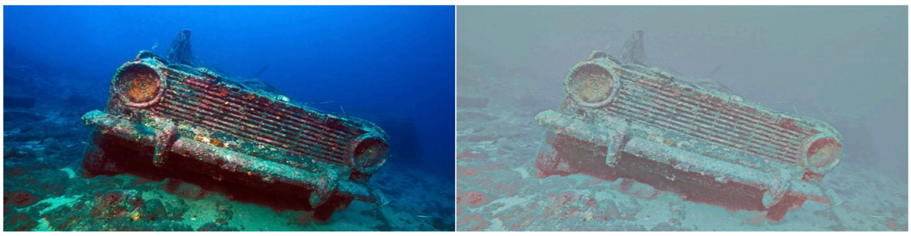

Fig. 3. Original image (Left) and image with Retinex Applied (Right).

2.2 Real-time Detection

Humans can simply glance at an image and instantly recognise objects within, while understanding how they interact with each other. The human visual system is both fast and accurate, allowing us to do complicated tasks, such as driving, without considerable conscious thought. Processing time is limited for real-time applications such as ROV inspection, further increasing the difficulties of object detection [31]. Replicating the speed and performance of the human visual system in computer algorithms could bring a host of benefits such as increasing the autonomy of underwater robots for the inspection domain [32]. The Faster Region Based-Convolutional Neural Network (Faster R-CNN) [33] is the *de facto* real-time object detection algorithm. Faster R-CNN is an ensemble method, combining a Region Proposal Network (RPN) and a Fast R-CNN network. The RPN is a Deep Convolutional Neural Network that proposes regions of interest. Fast R-CNN then uses these regions to make predictions using bounding boxes to classify objects. Both of the networks share the same convolutional layers, greatly reducing the computational costs of producing region proposals [34]. As a result, the network almost achieves real- time prediction speeds, at 5 fps.

YOLO is a true real-time algorithm, which excels in super-fast recognition of objects [35]. An acronym for 'You Only Look Once', this algorithm vastly outperforms the standard region-based networks in terms of prediction speed. The name derives from the algorithm's ability to need just one glance at an image to be able to identify and locate objects within it, bringing us closer to the replication of human vision. YOLO has obtained classification speeds of up to 67 frames-per-second in newer versions [36]. This type of network is very good at generalising. For example, when trained on natural images, the network can learn to detect these objects in artwork [33]. One of the major drawbacks of YOLO, however, is the difficulty it has detecting small objects that are grouped together [36]. This should not be a problem for corrosion as it would be okay to class several, small patches of rust as one instance. YOLO is a relatively simple concept to comprehend. The first step in this method is to resize the image to a grid of dimensions $S \times S$. Within each grid square, the network draws several bounding boxes that differ in shape. A prediction is then made on each bounding box using a single convolutional neural network. Predictions above a certain threshold value are used to classify the objects within the images [33].

2.3 Instance Segmentation

Humans are efficient machines for localising patterns and grouping them into meaningful fragments. Using the power of the modern computer, scientists seek to mimic this grouping ability that humans possess, giving rise to the field of image segmentation [37]. Semantic image segmentation offers precise region boundaries, where basic box boundaries may be insufficient. For example, this technique is heavily used in the self-driving car industry. It would be dangerous if cars drove according to an approximate boundary of a road or obstacles, such as from a bounding box. Precise outlines of objects provided by segmentation algorithms are vital here to ensure cars know exactly how objects relate to each other to drive safely around their surroundings [38]. Instance segmentation offers both the pixel-wise classification abilities of semantic segmentation and the ability to detect different instances of the same class of object such as in classic boundary-box object detection. This means that where semantic segmentation can only display the presence of a particular class, instance segmentation can further split this up to tell us more about the frequency of a class [39]. Mask R-CNN [40], which stands for Mask Regional Convolutional Neural Network, is the de facto method for instance segmentation and combines the segmentation abilities of the Fully Convolutional Network (FCN) [41] with the fast object recognition capability of Faster R-CNN [32]. This new concept extends Faster R-CNN by adding a branch to its network to predict segmentation 'masks' for each Region of Interest (RoI), in parallel with the task of classifying the objects and drawing boundary boxes. A small FCN is applied to each RoI to handle the generation of segmentation masks. According to the original paper, this additional step in the network only adds a small additional overhead to the Faster R-CNN, allowing it to still run at 5 fps. Instance segmentation would maximise information gained from predictions during inspection as a pixel-perfect outline of each individual instance of corrosion can be drawn. By using Mask R-CNN, this may be achieved in almost real-time standards.

3 Methodology

Our experimental framework consists of a pipeline where we will test all possible combinations between pre-processing methods and instance segmentation algorithms to discover the best approach. In addition, we will test all segmentation methods using the original images (i.e., no pre-processing). For pre-processing, CLAHE, Gray World and Retinex were used. In the segmentation phase, YOLO, Mask R-CNN and a standard CNN were chosen to explore the effectiveness of image classification, bounding-box object recognition and instance segmentation. Figure 4 depicts the workflow taken during experimentation.

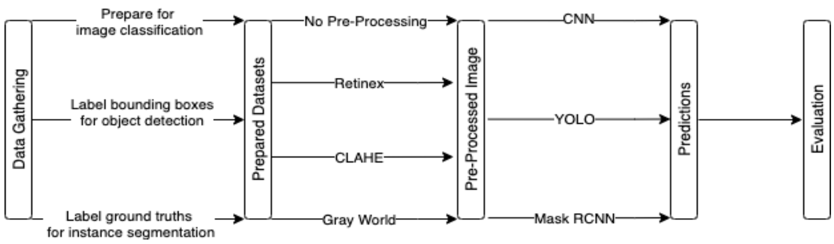

Figure. 4. Workflow representation.

No publicly available datasets were found for corrosion, let alone for subsea inspection. Search engines also highlighted the scarcity of data surrounding subsea inspection domain, as very minimal images were returned for queries of this type. Therefore, a new dataset[2] was generated using publicly available sources for general corrosion, not limited to that of the subsea domain. A Python script was used to scrape online image repositories on Google and Bing using corrosion-related search terms. Each scraped image was automatically downloaded on the naïve assumption that it was relevant. Manual pruning was then conducted to remove any erroneous samples. An additional set was generated using the same method but for the negative class, 'no rust', and using

[2] The dataset can be downloaded from https://drive.google.com/drive/folders/1dbOVdg5x75brUAwuI2X6voIMEwJzfiYX?usp=sharing

search terms such as 'metallic objects. This was necessary for CNN, but not for Mask R-CNN and YOLO as they do not make use of the negative class.
For CNN, images were annotated by placing them inside 'corrosion' and 'no corrosion' directories. A Python tool called 'labelImg' was used to draw bounding-boxes around objects in each image in YOLO format. VGG Image Annotator (VIA) is an HTML tool that was used to draw polygonal outlines of each object for Mask R-CNN. The final training set was composed of 1,272 images in total with a 1,106 corrosion and a 166 no-corrosion split. A train-test split of 70:30 was used for all 3 classifiers. An under-water corrosion validation set was compiled of 24 images, downloaded manually from Google Images. The same annotation methods were repeated as for the training set. Each set was duplicated once for each pre-processing method applied. The parameters used in the three instance segmentation algorithms are:
CNN. Batch_size = 128, Epochs = 15, IMG_HEIGHT=150, IMG_WIDTH=150, Conv2D(16,3,padding='same',activation='relu',input_shape=(IMG_HEIGHT, IMG_WIDTH,3)), MaxPooling2D(), Dropout(0.2), Conv2D (32,3, padding='same', activation='relu'), MaxPooling2D(), Dropout(0.2), Flatten(), Dense(512,activa-tion='relu'), Dense(1)
YOLO. Classes=1, Filters=18, Batch =24, Subdivisions=8
Mask R-CNN. Detection_min_conf = 0.9, IMGS_PER_GPU=2, NUM_CLASSES=2 (1 for back-ground, 1 for rust), STEPS_PER_EPOCH=1

4 Results

Tables 1 and 2 below refer to the calculated precision returned during testing of each classifier on each dataset. For CNN this is standard precision, but for YOLO and Mask R-CNN, the mean Average Precision (mAP) is used. Moreover, we measured the total time to predict a sample, yielding 0.05 (CNN), 0.69 (YOLO) and 12.63 (Mask R-CNN) seconds respectively[3]. Finally, in Figure 5 we show an example of the behaviour of all combinations in a single image.

Table 1. Precision of Classifiers (%).

Dataset	CNN	YOLO	Mask R-CNN
Surface	90.9	7.1	57.0
Underwater	75.0	9.0	77.1

Table 2. Precision of Pre-Processing Methods (%).

Pre-Processing	CNN	YOLO	Mask R-CNN
None	75.0	9.0	77.1
Gray World	75.0	14.2	69.8
Retinex	16.7	0.3	44.7
CLAHE	70.8	9.3	66.6

[3] Experiments were run on a MacBook Pro with a 2.3 GHz Dual-Core Intel Core i5 processor, 8 GB 2133 MHz LPDDR3 memory and an Intel Iris Plus Graphics 640 1536 MB graphics card.

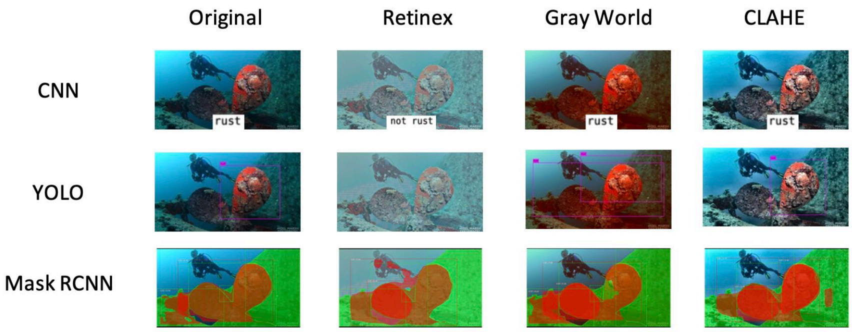

Fig. 5. Sample predictions for each pre-processing method and classification algorithm

CNN proved to have learned considerably well on the corrosion dataset, obtaining a precision of 90.9% on the surface validation set. It was shown that CNN could transfer the knowledge gained from the surface dataset into the underwater dataset, claiming a precision of 75%. The drop-in precision was expected, as the subsea data was not included in the training images due to lack of availability, meaning that CNN was not familiar with the aquatic environment. CNN is also shown to meet expectations in terms of prediction speed, due to the absence of regional layers that are present in object detection algorithms. The main drawback of the CNN was that no information regarding the location and frequency of object is included in predictions – meaning more work for the engineer.

True real-time object recognition was shown to be unachievable using YOLO for corrosion inspection. When tested against both the surface and sub-surface datasets, the model appears to have struggled to learn patterns associated with rust. This is likely due to the problem noted in [36] where the algorithm can have difficulty detecting small, close together objects. It may also be the case that due to corrosion being a property of an object, rather than an object in its own right, YOLO struggles to learn features connected with rust. When the confidence threshold was adjusted down, YOLO began to output predictions of some degree of precision. However, it was inconsistent, and the threshold value had to be altered manually for each individual image before a prediction was shown, defeating the purpose of using the algorithm for automation. What the algorithm failed to achieve in prediction ability, it made up for in prediction times. The algorithm met expectancies by classifying objects in less than a second, making it the quickest of both object detection algorithms and bringing the project the closest to real-time recognition.

Performance was unmatched to the Mask R-CNN algorithm, which achieved a promising 57% mAP value. Mask R-CNN defies expectations as it appears to improve performance when tested on the underwater dataset, reaching an admirable 77.1% mAP precision. Inspecting the sample prediction above, it is clear that the model can cope with corrosion in the foreground and can segment most of the propeller, but the background corrosion is mostly ignored. Instance segmentation, as demonstrated by the Mask R-CNN, is shown to be the most useful of the studied algorithms for corrosion inspection. The machine can now inform the inspecting engineer of not only the presence of rust or its location but of the extent of which corrosion covers an area in an image. This means that the user can focus less on the laborious task of finding corrosion and instead divert their attention immediately on the more expert task of implementing the measures to treat the instance of rust found. The drawback of this method, however, was that prediction times were an order of magnitude slower than that of YOLO and CNN, which may be an issue for ROV automation. The Gray World pre-processing technique shows to hold some degree of promise for increasing the precision of the trained networks. To the human eye, Gray World is shown to demonstrate stark mitigation of the undesirable effects of the underwater domain. It can be seen that the blue tint that distorts the image is reduced, making it more natural in appearance. Although, the over-compensation effect described by [24] was observed as some images were noticed to have an unnatural red tint. Numerically, the algorithm shows that, on average, it increases model precision by 18.7%. However, when taking a closer look, it is shown to have no effect on CNN overall and actually harms the prediction process of Mask RCNN by 10.5%.

Retinex was predicted to be the best performing pre-processing technique but the converse was found to be true. Visually, Retinex carries out an outstanding task of nullifying the effects of the aquatic environment on the image. Nonetheless, this does appear to be at the cost of the overall quality of the image. The product becomes obviously artificial, with a dramatic increase in contrast and a loss of colour depth, making it harder to identify rust within. Further testing supports the human evaluation of Retinex, with each network demonstrating an average loss of predictive ability when coupled with the algorithm. CLAHE was another algorithm thought to potentially boost the prediction capability of the trained neural networks. Using this method was found to not affect the colour of the overall image, but a considerable improvement was made regarding the image contrast. Altering the image in this way brings out more details in the image, making the edge of objects clearer and more definitive. Visibility within an image is subsequently improved, as detail in the background can now be identified more clearly. The machine interpretation of CLAHE was different, as it was found that applying this technique reduced prediction performance by 1.2%. Only YOLO suggests having benefited from this type of pre-processing, whereas CNN and Mask RCNN show a decline in precision.

5 Conclusion & Future Work

This paper demonstrates that computer vision techniques can learn to recognise corrosion in the underwater environment and the results are encouraging. The research conducted compares three state-of-the-art deep learning algorithms - CNN, YOLO and Mask R-CNN, to identify which is most appropriate for assisting the corrosion inspection process. Ultimately, the investigation found Mask R-CNN as the most suitable computer vision algorithm to recognise instances of corrosion in underwater images. The selected method obtained a precision of 77.1% mAP in a time of 12.63s when tested in the subsea domain. To support the machine at test-time, three established pre-processing techniques were trialled, and their assistance was documented. Enhancing images according to the Gray World assumption in pre-processing before prediction found to hold the most promise, with an average net increase of 18.7% in mAP across the four models. However, on the selected deep learning network, Mask RCNN, no pre-processing technique was found to help, and Gray World was seen to reduce the effectiveness of Mask RCNN by 10.5%. The outcome of this work may be appealing to the industry to alleviate pressures experienced as a result of the downturn. However, more work must be done to further understand how image recognition can be applied to inspection domains and to determine their true performance. The scarce supply of underwater data meant that it was impossible to train models on true domain-specific knowledge, and alternative measures had to be introduced. In future research, a repository of underwater images should be compiled, perhaps with the assistance of Generative Adversarial Networks (GANs) to create more data. The experiments should be then repeated using the underwater dataset during training in order to properly evaluate the ability of the machine to learn in this context. Pre-processing experiments should also be re-run, with the potential for the techniques to also be optimised and used during training.

References

1. Yang Y, Khan F, Thodi P and Abbassi R 2021 Corrosion induced failure analysis of subsea pipelines
2. R. Anantharaman, M. Velazquez and Y. Lee, "Utilizing Mask R-CNN for Detection and Segmentation of Oral Diseases," 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, 2018, pp. 2197-2204, doi: 10.1109/BIBM.2018.8621112.
3. F. Wu, G. Jin, M. Gao, Z. HE and Y. Yang, "Helmet Detection Based On Improved YOLO V3 Deep Model," 2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC), Banff, AB, Canada, 2019, pp. 363-368, doi: 10.1109/ICNSC.2019.8743246.
4. Badue C, Guidolini R, Carneiro R, Azevedo P, Cardoso V, Forechi A, Jesus L, Berriel R, Paixão T, Mutz F, de Paula Veronese L, Oliveira-Santos T and De Souza A 2021 Self-driving cars: A survey
5. Szymak P, Piskur P, Naus K. The effectiveness of using a pretrained deep learning neural networks for object classification in underwater video. Remote Sens. 2020;12(18):1–19.
6. Andreas Arnold-Bos, Jean-Philippe Malkasse, Gilles Kervern. A preprocessing framework for automatic underwater images denoising. European Conference on Propagation and Systems, Mar 2005, Brest, France. ffhal-00494314
7. Stéphane Bazeille, Isabelle Quidu, Luc Jaulin, Jean-Philippe Malkasse. Automatic underwater image pre-processing. CMM'06, Oct 2006, Brest, France. ffhal-00504893
8. S. Huang, F. Cheng and Y. Chiu, "Efficient Contrast Enhancement Using Adaptive Gamma Correction with Weighting Distribution," in IEEE Transactions on Image Processing, vol. 22, no. 3, pp. 1032-1041, March 2013, doi: 10.1109/TIP.2012.2226047.
9. M. Abdullah-Al-Wadud, M. H. Kabir, M. A. Akber Dewan and O. Chae, "A Dynamic Histogram Equalization for Image Contrast Enhancement," in IEEE Transactions on Consumer Electronics, vol. 53, no. 2, pp. 593-600, May 2007, doi: 10.1109/TCE.2007.381734.
10. A. W. Setiawan, T. R. Mengko, O. S. Santoso and A. B. Suksmono, "Color retinal image enhancement using CLAHE," International Conference on ICT for Smart Society, Jakarta, Indonesia, 2013, pp. 1-3, doi: 10.1109/ICTSS.2013.6588092.

11. Ponraj, D & Jenifer, M & Poongodi, P. & Manoharan, Samuel. (2011). A Survey on the Preprocessing Techniques of Mammogram for the Detection of Breast Cancer. Journal of Emerging Trends in Computing and Information Sciences. 2.
12. J. A. Stark, "Adaptive image contrast enhancement using generalizations of histogram equalization," in IEEE Transactions on Image Processing, vol. 9, no. 5, pp. 889-896, May 2000, doi: 10.1109/83.841534.
13. J. B. Zimmerman, S. M. Pizer, E. V. Staab, J. R. Perry, W. McCartney and B. C. Brenton, "An evaluation of the effectiveness of adaptive histogram equalization for contrast enhancement," in IEEE Transactions on Medical Imaging, vol. 7, no. 4, pp. 304-312, Dec. 1988, doi: 10.1109/42.14513.
14. Badue C, Guidolini R, Carneiro R, Azevedo P, Cardoso V, Forechi A, Jesus L, Berriel R, Paixão T, Mutz F, de Paula Veronese L, Oliveira-Santos T and De Souza A 2021 Self-driving cars: A survey
15. S. M. Pizer, R. E. Johnston, J. P. Ericksen, B. C. Yankaskas and K. E. Muller, "Contrast-limited adaptive histogram equalization: speed and effectiveness," [1990] Proceedings of the First Conference on Visualization in Biomedical Computing, Atlanta, GA, USA, 1990, pp. 337-345, doi: 10.1109/VBC.1990.109340.
16. M. S. Hitam, E. A. Awalludin, W. N. Jawahir Hj Wan Yussof and Z. Bachok, "Mixture contrast limited adaptive histogram equalization for underwater image enhancement," 2013 International Conference on Computer Applications Technology (ICCAT), Sousse, Tunisia, 2013, pp. 1-5, doi: 10.1109/ICCAT.2013.6522017.
17. Reza, A.M. Realization of the Contrast Limited Adaptive Histogram Equalization (CLAHE) for Real-Time Image Enhancement. *The Journal of VLSI Signal Processing-Systems for Signal, Image, and Video Technology* **38,** 35–44 (2004). https://doi.org/10.1023/B:VLSI.0000028532.53893.82
18. J. Han, S. Yang and B. Lee, "A Novel 3-D Color Histogram Equalization Method with Uniform 1-D Gray Scale Histogram," in IEEE Transactions on Image Processing, vol. 20, no. 2, pp. 506-512, Feb. 2011, doi: 10.1109/TIP.2010.2068555.
19. Ahmad Shahrizan Abdul Ghani, Nor Ashidi Mat Isa, Enhancement of low quality underwater image through integrated global and local contrast correction, Applied Soft Computing, Volume 37, 2015, Pages 332-344, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2015.08.033.
20. K. Iqbal, M. Odetayo, A. James, Rosalina Abdul Salam and Abdullah Zawawi Hj Talib, "Enhancing the low quality images using Unsupervised Colour Correction Method," 2010 IEEE International Conference on Systems, Man and Cybernetics, Istanbul, Turkey, 2010, pp. 1703-1709, doi: 10.1109/ICSMC.2010.5642311.
21. Y. Y. Schechner and N. Karpel, "Clear underwater vision," Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004., Washington, DC, USA, 2004, pp. I-I, doi: 10.1109/CVPR.2004.1315078.
22. B. Henke, M. Vahl and Z. Zhou, "Removing color cast of underwater images through non-constant color constancy hypothesis," 2013 8th International Symposium on Image and Signal Processing and Analysis (ISPA), Trieste, Italy, 2013, pp. 20-24, doi: 10.1109/ISPA.2013.6703708.
23. Chikane, Varsha and Chiou-Shann Fuh. "Automatic White Balance for Digital Still Cameras." *J. Inf. Sci. Eng.* 22 (2006): 497-509.
24. C. O. Ancuti, C. Ancuti, C. De Vleeschouwer and P. Bekaert, "Color Balance and Fusion for Underwater Image Enhancement," in IEEE Transactions on Image Processing, vol. 27, no. 1, pp. 379-393, Jan. 2018, doi: 10.1109/TIP.2017.2759252.
25. Foster, Does colour constancy exist? Trends in Cognitive Sciences, Volume 7, Issue 10, 2003, Pages 439-443, ISSN 1364-6613, https://doi.org/10.1016/j.tics.2003.08.002.

26. Land, Edwin H. “The Retinex Theory of Color Vision.” *Scientific American*, vol. 237, no. 6, 1977, pp. 108–129., www.jstor.org/stable/24953876. Accessed 10 Mar. 2021.
27. X. Fu, P. Zhuang, Y. Huang, Y. Liao, X. Zhang and X. Ding, "A retinex-based enhancing approach for single underwater image," 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 2014, pp. 4572-4576, doi: 10.1109/ICIP.2014.7025927.
28. Hines G, Rahman Z, Jobson D and Woodell G Single-Scale Retinex Using Digital Signal Processors *Citeseerx.ist.psu.edu*
29. B. Xie, F. Guo and Z. Cai, "Improved Single Image Dehazing Using Dark Channel Prior and Multi-scale Retinex," 2010 International Conference on Intelligent System Design and Engineering Application, Changsha, China, 2010, pp. 848-851, doi: 10.1109/ISDEA.2010.141.
30. Vishwakarma, Anish & Mishra, Agya. (2012). Color Image Enhancement Techniques: A Critical Review. Indian J. Comput. Sci. Eng. 3.
31. Z. Deng, H. Sun, S. Zhou, J. Zhao and H. Zou, "Toward Fast and Accurate Vehicle Detection in Aerial Images Using Coupled Region-Based Convolutional Neural Networks," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 10, no. 8, pp. 3652-3664, Aug. 2017, doi: 10.1109/JSTARS.2017.2694890.
32. Ren S, He K, Girshick R and Sun J Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks *arXiv.org*
33. J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.
34. Q. Fan, L. Brown and J. Smith, "A closer look at Faster R-CNN for vehicle detection," 2016 IEEE Intelligent Vehicles Symposium (IV), Gothenburg, Sweden, 2016, pp. 124-129, doi: 10.1109/IVS.2016.7535375.
35. W. Lan, J. Dang, Y. Wang and S. Wang, "Pedestrian Detection Based on YOLO Network Model," 2018 IEEE International Conference on Mechatronics and Automation (ICMA), Changchun, China, 2018, pp. 1547-1551, doi: 10.1109/ICMA.2018.8484698.
36. Juan Du 2018 *J. Phys.: Conf. Ser.* 1004 012029.
37. Hongyuan Zhu, Fanman Meng, Jianfei Cai, Shijian Lu, Beyond pixels: A comprehensive survey from bottom-up to semantic image segmentation and cosegmentation, Journal of Visual Communication and Image Representation, Volume 34, 2016, Pages 12-27, ISSN 1047-3203, https://doi.org/10.1016/j.jvcir.2015.10.012.
38. T. Pohlen, A. Hermans, M. Mathias and B. Leibe, "Full-Resolution Residual Networks for Semantic Segmentation in Street Scenes," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 3309-3318, doi: 10.1109/CVPR.2017.353.
39. Romera-Paredes B., Torr P.H.S. (2016) Recurrent Instance Segmentation. In: Leibe B., Matas J., Sebe N., Welling M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science, vol 9910. Springer, Cham. https://doi.org/10.1007/978-3-319-46466-4_19
40. Kaiming He, Georgia Gkioxari, Piotr Dollar, Ross Girshick; Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2961-2969
41. Hyeonwoo Noh, Seunghoon Hong, Bohyung Han; Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1520-1528