Delft University of Technology

Delft Center for Systems and Control

Technical report 23-003

Towards establishing an automated selection framework for underwater image enhancement methods*

A. Ilioudi, B.J. Wolf, A. Dabiri, and B. De Schutter

If you want to cite this report, please use the following reference instead:
A. Ilioudi, B.J. Wolf, A. Dabiri, and B. De Schutter, "Towards establishing an automated selection framework for underwater image enhancement methods," *Proceedings of the OCEANS 2023 Limerick*, Limerick, Ireland, June 2023. doi:10.1109/OCEANSLimerick52467.2023.10244710

Delft Center for Systems and Control
Delft University of Technology
Mekelweg 2, 2628 CD Delft
The Netherlands
phone: +31-15-278.24.73 (secretary)
URL: https://www.dcsc.tudelft.nl

Towards establishing an automated selection framework for underwater image enhancement methods

Athina Ilioudi
Delft Center for Systems and Control
Delft University of Technology
Delft, Netherlands
a.ilioudi@tudelft.nl

Ben J. Wolf
Bernoulli Institute
University of Groningen
Groningen, Netherlands
b.j.wolf@rug.nl

Azita Dabiri
Delft Center for Systems and Control
Delft University of Technology
Delft, Netherlands
a.dabiri@tudelft.nl

Bart De Schutter
Delft Center for Systems and Control
Delft University of Technology
Delft, Netherlands
b.deschutter@tudelft.nl

***Abstract*—The majority of computer vision architectures are developed based on the assumption of the availability of good quality data. However, this is a particularly hard requirement to achieve in underwater conditions. To address this limitation, plenty of underwater image enhancement methods have received considerable attention during the last decades, but due to the lack of a commonly accepted framework to systematically evaluate them and to determine the likely optimal one for a given image, their adoption in practice is hindered, since it is not clear which one can achieve the best results. In this paper, we propose a standardized selection framework to evaluate the quality of an underwater image and to estimate the most suitable image enhancement technique based on its impact on the image classification performance.**

***Index Terms*—computer vision, underwater image enhancement, image processing**

I. Introduction

State-of-the-art image processing methods are gaining traction rapidly for underwater applications. Advanced techniques are emerging from the latest achievements in the field of computer vision. Enabled by the latest computer vision advancements, novel, sophisticated, underwater image processing methods have been developed, allowing underwater robotic vehicles to perceive their surrounding environment in order to achieve particularly complex tasks, such as litter detection, equipment inspection, fish monitoring, and tracking of ecosystems. Yet, an interesting paradox hovers over the development of underwater image processing techniques: the more approaches emerge, the less straightforward it is to determine which one is the most suitable to be implemented in each case.

More specifically, the fast pace of computer vision breakthroughs has accelerated the spread of image processing techniques in underwater applications in order to compensate for the degraded visibility conditions in the water. In general, the quality of underwater images is particularly low, due to various distortions caused by absorption, scattering, and other characteristics that effect the vision under the water. Due to the water, which is a particularly complex and inhomogeneous medium and also due to absorption and scattering of light, less light can enter a camera during image captioning, in comparison with an environment without water. An example to illustrate the degraded underwater visibility conditions is to consider that knowing the true color of an object underwater is almost not possible even for humans, since in these conditions, human vision is distorted as well.

Underwater image enhancement is a key technology in compensating for the poor visibility conditions and enabling underwater robots to perform complex tasks autonomously. However, the degradation variations among different water environments hinder the process of selecting the enhancement technique that is most suitable to be implemented in each case [9]. Furthermore, the currently existing underwater Image Quality Assessment (IQA) metrics, especially the ones that do not take into account a reference image (no-reference), have not yet been validated with respect to their relation to the performance of computer vision tasks. In other words, IQA metrics do not evaluate an enhancement method on the basis of the performance increase they yield for image classification, which is a fundamental objective of employing image enhancement approaches.

The aim of this work is to present a standardized analysis of the impact of image enhancement techniques on underwater

This work was supported in part by the European Union's Horizon 2020 Research and Innovation Programme under Grant 871295 (SeaClear), in part by the European Union under the Horizon Europe Programme (Grant Agreement 101093822 - SeaClear2.0), and in part by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme under Grant 101018826-CLariNet.

images as well as to propose an automated framework that evaluates the quality of an underwater image and estimates the best image enhancement technique based on the impact of each technique on the image classification performance. The potential advantages of this framework are also discussed. Our analysis is complementary to the existing works in the field of underwater image enhancement. The key differences are that (1) our focus is different and is primarily on integrating the impact on the image classification performance, (2) we introduce an automated selection approach for categorizing image enhancement techniques based on the image classification performance, and (3) we provide valuable insights about the image enhancement effect during pre-processing and post-processing of the training data. Specifically, this work includes five main steps:

- We implement three image enhancement techniques to raw underwater image data namely Fusion [2], as well as the neural networks WaterNET [5], and Cast-GAN [8], and we generate an image dataset for each implemented technique.
- We evaluate a convolutional neural network (CNN) that performs image classification on each one of the abovementioned generated datasets [15].
- We propose an image quality assessment method based on the image classification performance of the trained CNN.
- We analyze the quality of an underwater image in terms of comparing the performance of human-based image classification, algorithmic-based image classification, as well as, in terms of the respective quality rated by underwater image quality metrics.
- We train an additional supervised classification model that relates images, their IQA metrics, and the CNN's performance with image enhancement techniques.

The paper is organized as follows: Section II presents an overview of necessary preliminary knowledge. In Section III the proposed scheme is presented and analyzed. In Section IV the implementation of the proposed architecture is presented. Section V demonstrates the results and Section VI gives concluding remarks and an outlook to future work.

II. Preliminaries

This section presents some background and material with regards to evaluation metrics and methods that are analyzed within this study.

A. IQA Metrics

Performance validation and comparison among underwater image enhancement methods remains a largely unexplored research area. As an alternative to commonly used subjective tests, IQA metrics are implemented as objective measures to quantify perceptual quality [9]. Underwater IQA metrics measure color and contrast degradation. Then they quantify the perceived image quality via image attributes related to the degradation in water and they combine image attributes to mimic human preferences in the enhanced images. IQA metrics that are commonly used to assess the quality of enhanced underwater images are Underwater Image Quality Measure (UIQM) [11], Underwater Color Image Quality Evaluation (UCIQE) [16], and Colorfulness, Contrast, and Fog density index (CCF) [14]. However, it has been reported that underwater-specific IQA measures do not satisfactorily rate the quality of enhanced underwater images as it is not clear whether a high IQA metric value can guarantee high performance in the achievement of computer vision tasks [9].

B. Underwater Image Enhancement Methods

A number of image enhancement methods for underwater applications has been applied in this work, including both physics-based and neural-network-based ones. Fusion [2] is an effective technique that derives two images from a white-balanced version of the original, degraded input image and then it merges them based on a multi scale fusion algorithm. Cast-GAN [8] develops an underwater image enhancement method based on a generative adversarial network. Cast-GAN uses the trained generator to remove the color cast from underwater images, without distorting the color of water regions. Finally, WaterNET [5] proposes an enhancement framework based on a convolutional neural network. More specifically, in this technique, three versions of an underwater degraded image are generated by applying white balance [1], histogram equalization [17] and gamma correction [7] algorithms to it. Three confidence maps are extracted accordingly to these three versions of the original image by using a convolutional neural network. In this way, the neural network architecture learns three confidence maps that are subsequently used to combine the three versions into an enhanced result.

III. Proposed Selection Scheme

The selection pipeline is a classification model that relates an image with the best image enhancement technique to be implemented (Fig. 1). The approach is based on the Support Vector Machines (SVM) algorithm [3], which is a machine learning technique that maps the data to a higher dimensional space by using a so-called kernel function and searches for a hyperplane to distinctly separate the data points. The data points that are nearest to the hyperplane and are used to make the decision are called support vectors. Some examples of kernel functions for two data points (x_i, x_j) are the linear kernel:

$$K(x_i, x_j) = x_i^\intercal \ x_j + c, \tag{1}$$

the polynomial kernel:

$$k(x_i, x_j) = (\gamma x_i^\intercal \ x_j + c)^d, \ \gamma > 0, \tag{2}$$

and the radial basis function (RBF)

$$k(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}, \ \gamma > 0, \tag{3}$$

where γ, c are kernel parameters and d is the degree of the polynomial.

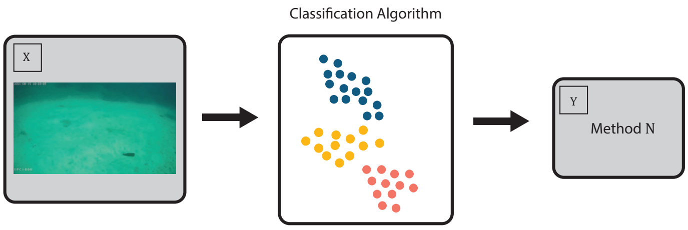

Fig. 1. Selection pipeline concept.

IV. METHODOLOGY

In this section, the implementation of the proposed framework is discussed. The dataset generation process, as well as the procedure followed for the CNN training and the implementation of the selection pipeline are explained.

A. Dataset Synthesis

In order to generate our dataset, two main aspects are taken into account:

- A broad range of underwater scenes, diverse characteristics of quality degradation, as well as a great variety of image content should be considered.
- The ultimate objective of this study is to assist in the task of underwater litter detection. Hence, the image data contain observations of 4 categories, namely animal, plant, Remotely Operated Vehicle (ROV), and trash.

The dataset is synthesized by real-world underwater images that were collected from different sea areas and by different cameras. More specifically, 300 images come from TrashCan public dataset [4]. TrashCan is sourced from the J-EDI (JAMSTEC E-Library of Deep-sea Images) dataset [12], by the Japan Agency of Marine Earth Science and Technology (JAMSTEC) and it contains videos from ROVs, largely in the sea of Japan. The visibility conditions of TrashCan images are considered as good. Furthermore, 300 additional images are generated from measurements in the sea area by Marseille, France. The visibility conditions in this area are worse compared to TrashCan.

Moreover, 600 images are taken from Dubrovnik, Croatia. The sea areas there have in general very clear visibility conditions. Two experiments were conducted with different hardware and camera equipment, leading to the collection of 300 images from each experiment. The two experiments were different in terms of various conditions such as, water conditions and equipment involved. As a result, the first 300 images showcase different characteristics in comparison to the remaining 300 images. We have thus constructed a diverse dataset of in total 1200 real-world underwater images, which is used in order to conduct our comprehensive study with regards to analyzing the image enhancement techniques.

We then implement the image enhancement techniques to the raw underwater images and we generate an image dataset for each implemented technique (Fig. 2). In total, we generate four datasets, i.e. raw, Fusion, Cast-GAN, and WaterNET datasets.

B. CNN Training

Considering our purpose to analyze underwater image enhancement methods with respect to their impact on the performance of computer vision tasks, we consider a deep-learning-based image classification approach, where the images are fed to a CNN, which in turn predicts their category. Our aim is to train a neural network capable of classifying each sample among the 4 available categories. It is a common practice to initialize the network's parameters using the weights as derived from training on another relatively similar task. In our case, we use the weights that are pretrained on the COCO (Microsoft Common Objects in Context) dataset [10] . COCO is a large-scale image dataset containing 328,000 images of everyday objects and humans. The dataset contains annotations that are widely used to train machine learning models to recognize, label, and describe objects.

Regarding the CNN architecture, YOLOv6 [6] is selected, which is one of the state-of-the-art CNN architectures. YOLOv6 is chosen since it is a real-time architecture with not only large accuracy but also fast inference speed. It is a single-stage object detection framework mostly focusing on industrial applications, with hardware efficient design and better performance than its predecessors [13].

In order to evaluate the image enhancement methods, the following two experiments are performed:

1) Training a CNN for each processed dataset: As a first step, we train YOLOv6 on each one of the four processed image datasets (Fig. 2). Having the four CNN networks, we aim to evaluate the impact of each enhancement technique on the feature space representation and its resulting effect on the learning process.

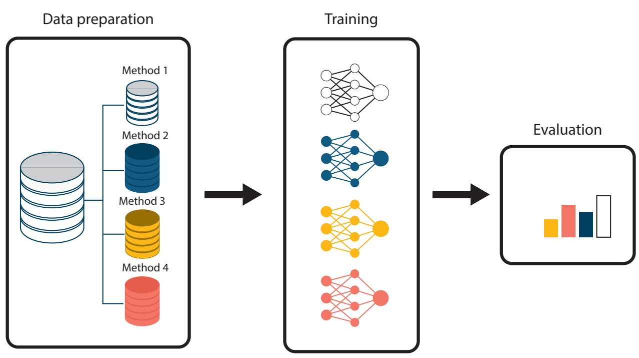

Fig. 2. Training a CNN architecture for each processed dataset.

2) Training a CNN on raw data: In the second experiment, we train YOLOv6 on raw TrashCan data and then we evaluate it on each of the four datasets. With this experiment, we aim to investigate whether a technique can operate as a post processing technique, allowing to employ pre-trained neural networks from other applications. This experiment aligns with enhancing human-based perception on image classification. However, neural-network-based perception vastly depends on the data that it has been trained on, and hence, we want to investigate whether shifting the data distribution might deteriorate the performance.

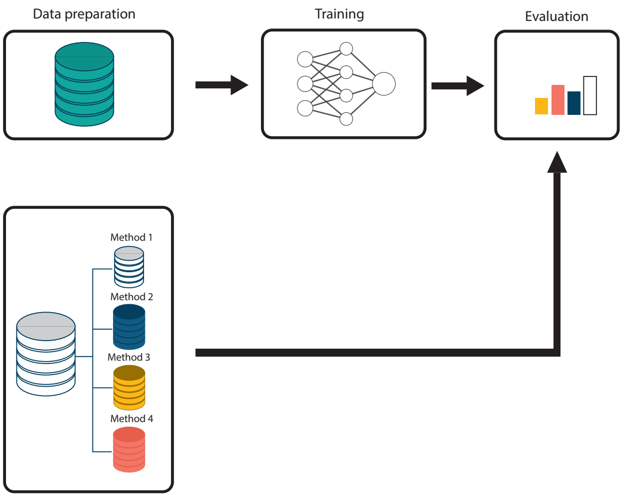

Fig. 3. Training a CNN architecture on raw image data and evaluating it on enhanced image data

C. Selection Framework

Before training the classification model, we implement data augmentation to the image samples, in order to enlarge our dataset. The original dataset contains around 230 images coming from the validation set of the previously described dataset. The augmentation that is performed includes the implementation of image transformations such as rotation, horizontal flip and transpose operations.

Based on the trained CNN architectures and the classification achieved per image, the multi-class SVM is applied to relate images with the best enhancement technique. The implementation is based on the scikit-learn library from Python programming language. We explore two variations of this architecture by adjusting the input features provided to the algorithm in order to learn the mapping between an image and the best enhancement approach: (i) In the first variation, the input features used are the pixel values corresponding to the hue (Hue), saturation (Saturation), and brightness (Value/Brightness) channels. (ii) In the second case, we perform a training of the SVM architecture by adopting the IQA metrics (UCIQE, UIQM, and CCF) as features for the classification task. In this way we perform a faster training than in the first architecture, since the three IQA values require much less computational effort to be processed than the pixel values of the HSV image channels. The dataset is divided using the ratio of 80:20, where 80 % is for training and 20 % is for testing.

V. Experimental Results

Our aim is to construct a mapping between a given image and the respective most suitable image enhancement method, considering the classification performance. For this reason, the predicted classification score and the respective difference with the ground truth per image is extracted. In other words, the predicted labels for each image are compared with the ground truth. The result is also compared with the predicted labels of the image processed by the other methods. For the images whose labels correspond to the correct ground truth values, the confidence score is compared with the other enhancement methods in order to validate which one ensures the highest confidence. An image enhancement technique is regarded as the best performing one per each image if it yields the highest confidence, provided that this confidence is greater than the confidence of the second best performing technique with a threshold of 2%. If this threshold difference is not met, then both of the highest scoring techniques are considered as best performing ones. The results are depicted in Fig. 4 and Fig. 5. More specifically, Fig. 4 depicts the number of images on which each enhancement method reaches the highest classification score while considering the 2% difference threshold. This figure corresponds to the first experiment presented in Section III. We can conclude that the method Fusion outperforms the rest methods, while WaterNet is the second best performing method respectively. Finally, Cast-GAN performs the worst among the four methods.

In addition, Fig. 4 depicts the results obtained from the second experiment in Section III. From this figure, it can be observed that the implementation of no enhancement (i.e. just raw image data) appears to perform the best, compared to any other image enhancement method. This can be explained by the fact that the training of the CNN is done on solely raw data. As a result, the network has learned to best distinguish the features belonging to each category, when the data are in raw form. WaterNet is the second best performing approach, followed by Cast-GAN. It is also noticeable that Fusion achieves a low performance at this experiment, contrary to the previous one where it outperformed the other methods.

Furthermore, it should be noted that the score using IQA metrics is not necessarily proportional to the classification score. As shown in Fig. 7, the Cast-GAN appears to outperform the other methods regarding the UCIQE metric. Nevertheless, this technique is not one of the top performing ones with regards to the classification performance. At the same time, Fig. 6 depicts the performance score per image among the examined methods in a descending order. As can be seen, some methods might reach quite a high classification score for only a few images, while achieving quite a low performance for the majority of the images.

The classification results are presented in Table I where the F1-score metric for the two SVM architectures is presented. Note that in this case, the dataset size limits the possibility to achieve higher performance in this four-class classification problem. Hence, in our future work we will consider larger size datasets.

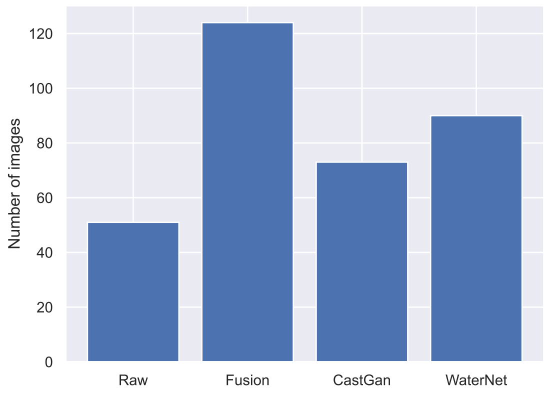

Fig. 4. Image enhancement methods evaluation based on four CNN architecture trained on each processed image dataset.

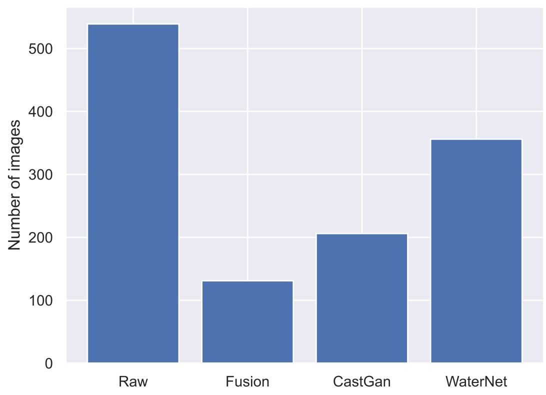

Fig. 5. Image enhancement methods evaluation based on a CNN architecture trained on raw image data.

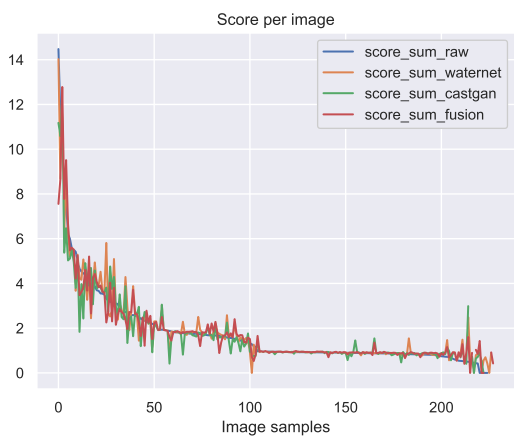

Fig. 6. Classification score among methods.

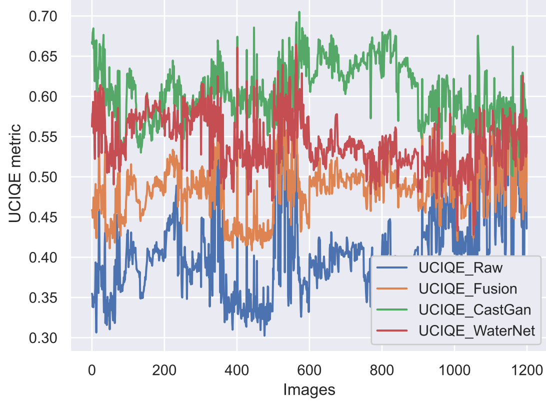

Fig. 7. UCIQE metric per image for each method.

TABLE I
PERFORMANCE OF CLASSIFICATION

f1-score	
SVM Classifier - HSV channels as features	41 %
SVM Classifier - IQA metrics as features	46 %

VI. Conclusions And Future Work

In this paper a framework is introduced towards automating the selection process of the most appropriate underwater image enhancement method to be implemented on an image and the impact of these methods on the performance of classification task is explored. The results show that there is a distinction between how the human eye perceives the image enhancement to more easily classify underwater objects and how a trained neural network increases its classification performance by the implementation of various techniques. A selection pipeline is also introduced based on a supervised classification model that performs a mapping between images, their IQA metrics, and their likely optimal enhancement technique.

Future work could contain the expansion of the dataset to further validate the proposed scheme. An additional important step should concern the implementation and analysis of more image enhancement methods. Finally, an interesting future research direction should concern the implementation of enhancement methods on parts of an image instead of the total image. This could be achieved if, for example, an image is split into a number of areas, and for each area the most appropriate image enhancement technique is implemented accordingly, instead of applying solely one technique on the total image. In this way, if only parts of an image are used, and in each of these parts a different technique is applied, even higher performance could be achieved, since the texture of an image usually differs along it and hence, a technique can have a varying performance among the different image areas.

References

[1] C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert. Enhancing underwater images and videos by fusion. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 81–88, 2012.

[2] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert. Color balance and fusion for underwater image enhancement. *IEEE Transactions on Image Processing*, 27(1):379–393, 2018.

[3] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[4] J. Hong, M. Fulton, and J. Sattar. Trashcan: A semantically-segmented dataset towards visual detection of marine debris, 2020.

[5] 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*, 29:4376–4389, 2020.

[6] C. Li, L. Li, H. Jiang, K. Weng, Y. Geng, L. Li, Z. Ke, Q. Li, M. Cheng, W. Nie, Y. Li, B. Zhang, Y. Liang, L. Zhou, X. Xu, X. Chu, X. Wei, and X. Wei. Yolov6: A single-stage object detection framework for industrial applications, 2022.

[7] C. Li, S. Tang, J. Yan, and T. Zhou. Low-light image enhancement based on quasi-symmetric correction functions by fusion. *Symmetry*, 12(9), 2020.

[8] C. Y. Li and A. Cavallaro. Cast-GAN: Learning to remove colour cast from underwater images. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 1083–1087, 2020.

[9] C. Y. Li and A. Cavallaro. On the limits of perceptual quality measures for enhanced underwater images. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 4148–4152, 2022.

[10] T. Lin, M. Maire, S. J. Belongie, L. D. Bourdev, R. B. Girshick, J. Hays, P. Perona, D. Ramanan, P. Doll'a r, and C. L. Zitnick. Microsoft COCO: Common Objects in Context. *CoRR*, abs/1405.0312, 2014.

[11] K. Panetta, C. Gao, and S. Agaian. Human-visual-system-inspired underwater image quality measures. *IEEE Journal of Oceanic Engineering*, 41(3):541–551, 2016.

[12] T. Sasaki, S. Azuma, S. Matsuda, A. Nagayama, M. Ogido, H. Saito, and Y. Hanafusa. JAMSTEC E-library of Deep-sea Images (J-EDI) Realizes a Virtual Journey to the Earth's Unexplored Deep Ocean. In *AGU Fall Meeting Abstracts*, volume 2016, pages IN53C–1911, Dec. 2016.

[13] F. Sultana, A. Sufian, and P. Dutta. *A Review of Object Detection Models Based on Convolutional Neural Network*, pages 1–16. Springer Singapore, Singapore, 2020.

[14] Y. Wang, N. Li, Z. Li, Z. Gu, H. Zheng, B. Zheng, and M. Sun. An imaging-inspired no-reference underwater color image quality assessment metric. *Computers & Electrical Engineering*, 70:904–913, 2018.

[15] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick. Detectron2. https://github.com/facebookresearch/detectron2, 2019.

[16] M. Yang and A. Sowmya. An underwater color image quality evaluation metric. *IEEE Transactions on Image Processing*, 24(12):6062–6071, 2015.

[17] K. Zuiderveld. Contrast limited adaptive histogram equalization. In P. S. Heckbert, editor, *Graphics Gems*, pages 474–485. Academic Press, 1994.