Artificial Intelligence Research and Development
A. Cortés et al. (Eds.)

doi:10.3233/FAIA220344

On the Importance of Color Pre-Processing for Object Detection in Submarine Images

Jose-Luis LISANI [a,1], Ana Belén PETRO [a], Catalina SBERT [a], Amaya ÁLVAREZ-ELLACURÍA [b], Ignacio A. CATALÁN [b], and Miquel PALMER [b]

[a] *Universitat Illes Balears (UIB), Spain*
[b] *IMEDEA (UIB-CSIC), Spain*

Abstract. When training a neural network for object detection a great deal of effort is usually devoted to augment the training dataset. The rationale behind this process is that augmentation increases the generalization capability of the network. However, little attention has been paid to the application of image enhancement techniques as a pre-processing step of the training task. In this paper we show, in the context of fish detection in submarine images, that the application of classical color enhancement methods may improve significantly the performance of the well known Mask R-CNN object detector.

Keywords. image enhancement, object detection, CNN, submarine images

1. Introduction

The use of image enhancement techniques is common in many vision tasks. These techniques are used to improve the contrast, the brightness and the color of the images, usually as a pre-processing step to a further analysis. As an example, we see in Figure 1 one original subaquatic image and the result of its processing with the Retinex algorithm [1]. We observe that the objects in the scene (in particular the fish) are more easily distinguishable in the processed image than in the original. In general, enhancement helps humans in detection tasks. The question arises whether the same is true for deep learning-based detection algorithms. In this paper we seek to find an answer to this question in the particular case of fish detection in submarine images.

Some recent papers [2,3] have shown that the visual quality of an image is not necessarily correlated with the accuracy of an object detector that uses this image as input. We study in the current work how the performance of a popular CNN for object detection (Mask R-CNN [4]) is affected by the use of five representative underwater image enhancement algorithms.

The paper is organized as follows. Section 2 gives a short overview of the techniques used for the enhancement of underwater images and presents the five algorithms selected for the study. Section 3.1 describes the CNN used in the experiments, the set of images used for training and evaluation, and details the training parameters. In Section 3.2 an statistical analysis of the results is provided. Finally, some conclusions are drawn in Section 4.

[1]Corresponding Author: Jose-Luis Lisani, Universitat Illes Balears, Spain; E-mail: joseluis.lisani@uib.es

Figure 1. Original image (left) and result of processing with Retinex algorithm [1] (right).

2. Enhancement of Underwater Images

Underwater images suffer from color cast, low contrast and haze due to the different attenuation of the light wavelengths and to the scattering effect. These effects are depth-dependent.

Model-free or prior-based methods can be used to enhance these images. The former seek to improve the visual quality without taking the depth dependency into account while the later are based on physical image-formation models. Model-free methods are simple and aplicable to a wider type of images, but are not always able to correctly improve them. Prior-based approaches do not always obtain good results due to the use of over-simplified models or to the difficulty to estimate correctly the parameters of the model in a general case. In recent years, a new trend of enhancement methods has emerged, the deep-based (or data-driven) approaches. However, the lack of available training data limits its performance [5].

In our study we have selected two popular model-free methods (MSR [1] and Fusion [6]), and three prior-based methods (UDCP [7], ARC [8] and InfoLoss [9]). MSR (Multi-Scale Retinex) aims at removing global illumination changes by locally improving the contrast of the image. Fusion combines contrast enhanced and color corrected versions of the original image using a multi-scale strategy. Both UDCP (Underwater Dark Channel Prior) and ARC (Automatic Red-Channel) estimate the depth map of the image and use this information to restore the color balance, but while UDCP bases its estimation on the green and blue color channels, ARC uses the red channel. Finally, InfoLoss consists of two steps, first a dehazing method is applied after estimating the depth map of the image, and then a contrast enhancement algorithm is applied.

We have used our own implementations of MSR [10], Fusion, UDCP and InfoLoss, based on the descriptions provided in the original papers. For ARC we have used the online tool for underwater image processing `https://puiqe.eecs.qmul.ac.uk/`.

3. Experiments

We have trained a popular CNN for object detection (Mask R-CNN [4]) using different processed versions of the same original images. We have then compared the mAP values obtained with the trained networks on a common test set.

Mask R-CNN permits simultaneous detection, classification and instance segmentation of the image objects. The network consists of a Backbone for feature extraction (we use Resnet101 in our tests), a Region Proposal Network (RPN) and three output branches, for bounding box location, object classification and segmentation, respectively. We have used the implementation of the network available at https://github.com/matterport/Mask_RCNN.

3.1. Experimental Setting

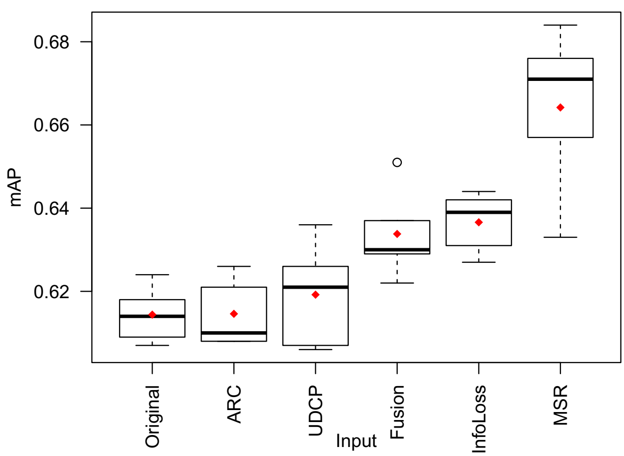
We have collected a dataset of 600 underwater images, coming from two different locations in the Mallorcan coast. All the fish in theses images have been manually segmented and the network have been trained to detect them. The dataset has been split into three sets: a training set (400 images, with 4252 annotated fish), a validation set (100 images, 917 fish, used for tuning the hyperparameters of the training) and a test set (100 images, 1004 fish, for evaluation of the results).

Six versions of the dataset have been used in the experiment: the original images and also the images processed with the five methods described in the previous section. The CNN has been trained and evaluated using these six datasets.

The following training strategy has been used: the upper layers of the network ('heads') have been trained for 30 epochs; the intermediate layers (fc3) have been trained for 30 additional epochs; finally, all the layers have been trained for other 30 epochs. In order to reduce the effect of the random nature of the minimization process in the obtained results, the above strategy has been repeated five times, and the mean average precision values (mAP) obtained on the test sets have been recorded.

3.2. Statistical Analysis of the Results

Figure 2 displays the mAP values obtained on the test set for each one of the trained networks, using as input the images pre-processed with the different methods.



Input	Average mAP
Original	0.61
ARC	0.61
UDCP	0.62
Fusion	0.63
InfoLoss	0.63
MSR	0.66

Figure 2. Boxplots of mAP values on the **test** set, obtained after pre-processing the input images with different enhancement techniques. The table displays average mAP values, represented as red dots in the figure.

Visually one can observe that almost all pre-processing methods improve the performance obtained with the un-processed images. In particular, MSR obtains a 5% increase on average.

In order to check if these differences are statistically relevant we perform a one way ANOVA test (Analysis of Variance) to the obtained values. The ANOVA result is significant ($F = 12.79, df = 5, p < 0.0001$), thus at least one group is significantly different from the rest. Additionally, we apply the Tukey Honest Significant Differences post-hoc test to obtain pairwise comparisons of the methods. The results (p-values) of the test are displayed in Table 1. We observe that significant differences are obtained when comparing MSR with the rest.

	Original	ARC	UDCP	Fusion	InfoLoss	MSR
Original	-	1.0	0.99	0.14	0.07	**< 0.001**
ARC	-	-	0.99	0.15	0.07	**< 0.001**
UDCP	-	-	-	0.41	0.23	**< 0.001**
Fusion	-	-	-	-	0.99	**< 0.01**
InfoLoss	-	-	-	-	-	**< 0.05**

Table 1. P-values corresponding to the pairwise comparison of the methods using the Tukey test. Statistically meaningfull differences are marked in bold type.

4. Conclusions

The obtained results show that the performance of an object detection network can be increased by preprocessing the input images (both during the training and the inference steps) using classical enhancement methods. In particular, for the case of underwater images, the use of the Multi-Scale Retinex method permits to significantly increase the mAP value by a 5% on average, with respect to the original un-processed images. As a continuation of this work we shall investigate how the use of augmentation techniques, both on the original and pre-processed images, may affect the detection results.

Acknowledgements

This work has been sponsored by the Comunitat Autonoma de les Illes Balears through the Direcció General de Política Universitària i Recerca with funds from the Tourist Stay Tax Law ITS 2017-006 (PRD2018/26).

References

[1] Jobson DJ, Rahman Z, Woodell GA. A multiscale retinex for bridging the gap between color images and the human observation of scenes. IEEE Trans on Image Processing. 1997.

[2] Liu R, Fan X, Zhu M, Hou M, Luo Z. Real-World Underwater Enhancement: Challenges, Benchmarks, and Solutions Under Natural Light. IEEE Trans on Circuits and Sys for Video Tech. 2020;30(12):4861-75.

[3] Chen L, Jiang Z, Tong L, Liu Z, Zhao A, Zhang Q, et al. Perceptual Underwater Image Enhancement With Deep Learning and Physical Priors. IEEE Trans on Circuits and Sys for Video Tech. 2021;31(8):3078-92.

[4] He K, Gkioxari G, Dollr P, Girshick R. Mask R-CNN. In: IEEE Int. Conf. on Computer Vision; 2017. p. 2980-8.

[5] Anwar S, Li C. Diving deeper into underwater image enhancement: A survey. Signal Processing: Image Communication. 2020;89:115978.

[6] Ancuti CO, Ancuti C, C DV, Bekaert P. Color balance and fusion for underwater image enhancement. IEEE Trans on Image Processing. 2018;27(1):379-93.

[7] Drews Jr P, do Nascimento E, Moraes F, Botelho S, Campos M. Transmission Estimation in Underwater Single Images. In: 2013 IEEE Int. Conf. on Computer Vision Workshops; 2013. p. 825-30.

[8] Galdran A, Pardo D, Picón A, Alvarez-Gila A. Automatic Red-Channel underwater image restoration. Journal of Visual Communication and Image Representation. 2015;26:132-45.

[9] Li CY, Guo JC, Cong RM, Pang YW, Wang B. Underwater Image Enhancement by Dehazing With Minimum Information Loss and Histogram Distribution Prior. IEEE Trans on Image Processing. 2016;25(12):5664-77.

[10] Petro AB, Sbert C, Morel JM. Multiscale Retinex. Image Processing On Line. 2014:71-88.