

RESEARCH

Open Access



# Underwater image segmentation in the wild using deep learning

Paulo Drews-Jr\* , Isadora de Souza, Igor P. Maurell, Eglén V. Protas and Silvia S. C. Botelho

\*Correspondence:  
paulodrews@furg.br  
Intelligent Robotics and  
Automation Group (NAUTEC),  
Federal University of Rio Grande  
(FURG), Rio Grande, Brazil

## Abstract

Image segmentation is an important step in many computer vision and image processing algorithms. It is often adopted in tasks such as object detection, classification, and tracking. The segmentation of underwater images is a challenging problem as the water and particles present in the water scatter and absorb the light rays. These effects make the application of traditional segmentation methods cumbersome. Besides that, to use the state-of-the-art segmentation methods to face this problem, which are based on deep learning, an underwater image segmentation dataset must be proposed. So, in this paper, we develop a dataset of real underwater images, and some other combinations using simulated data, to allow the training of two of the best deep learning segmentation architectures, aiming to deal with segmentation of underwater images in the wild. In addition to models trained in these datasets, fine-tuning and image restoration strategies are explored too. To do a more meaningful evaluation, all the models are compared in the testing set of real underwater images. We show that methods obtain impressive results, mainly when trained with our real dataset, comparing with manually segmented ground truth, even using a relatively small number of labeled underwater training images.

**Keywords:** Underwater images, Segmentation, Deep learning, Sensing

## Introduction

The segmentation of underwater images presents many applications in areas such as subsea inspection and biological research. Even a simple background subtraction, if it has a high accuracy, can be an important part of more complex tasks, such as animal counting [1, 2], image restoration, and robot obstacle avoidance [3, 4]. With that purpose, the segmentation method must be able to segment underwater images that are in the wild, and not in a controlled environment. As the main example, a technique that simply divide the input image in two classes, background and foreground, provides a valuable information for an algorithm that is responsible for an underwater robot obstacle avoidance, since it can show the regions of the image where there are possible objects to collide. However, underwater images exhibit some particular characteristics which make their handling more difficulty, including blurriness, reduced contrast, and distorted colors [5, 6]. Because of this, standard segmentation algorithms cannot be directly applied to

underwater images. Thus, the purpose of this paper is to explore two of the state-of-the-art deep learning segmentation architectures, together with restoration and fine-tuning techniques using underwater segmentation datasets also made available through this paper.

Convolutional neural networks (CNNs) are the current state of the art in image segmentation [7]. Thus, an evident solution to the problem of underwater image segmentation is the adaptation of state-of-the-art segmentation architectures to deal with underwater images. However, deep CNNs generally require more than thousands sample images to be properly trained. As the manual segmentation of images is a labor-intensive task, building a dataset as large as those usually used in other deep learning problems would take a considerable amount of time and resources. To overcome this problem, there are a few possible strategies.

The first one is to pre-train the network on the segmentation of non-underwater images and then fine tune it using a smaller dataset of manually labeled real underwater images. This approach is known as transfer learning [8]. Another approach is to use simulated data, which allows a larger number of training samples, but results in a less realistic dataset. Finally, we can try to pre-process the input to remove the effects of underwater degradation before segmenting it with a CNN trained with non-underwater images. We evaluated all these strategies in state-of-the-art image segmentation CNN architectures.

In this paper, we propose four datasets, mainly one composed of real manually annotated underwater images, to train deep CNN architectures to the task of underwater image segmentation. Furthermore, we present several deep learning solutions based on two state-of-the-art segmentation architectures, using different pre-processing and pre-training steps. All the setups are trained in each developed dataset, and after, the consequent models are evaluated using the ground truth of the real testing set as reference. To the best of our knowledge, we are the first work to use a CNN approach to underwater image segmentation in the wild. But the main contribution, which is what allow the use of CNN, is our dataset of real underwater images in the wild and their respective ground truths, which is made publicly available. We hope this dataset helps researchers to evaluate and improve underwater segmentation methods.

The remainder of the paper is organized as follows: the “[Related work](#)” section shows the related works in the areas of underwater image segmentation and image segmentation using CNNs; the “[Methodology](#)” section presents the proposed methodology; the “[Experimental results](#)” section evaluates the obtained results. Finally, we summarize the paper contributions and draw the future research directions in the “[Conclusion](#)” section.

## **Related work**

### **Deep learning-based segmentation**

In recent years, convolutional neural networks have become the state-of-the-art in the area of image segmentation, including high-level semantic segmentation. In [9], a texture segmentation and classification method based on features extracted by image classification CNNs trained on the ImageNet ILSVRC [10] dataset is proposed, achieving state-of-the-art performance on several datasets. Besides that, in [11], an end-to-end pixel-wise semantic segmentation using fully convolutional networks is performed. The main advantage of these models over standard CNNs is that they lack fully connected layers, allowing them to operate on inputs of variable size without the need to mod-

ify the network's architecture. Following, the Mask R-CNN [12], that is an extension of the Faster R-CNN [13], objects detection architecture, achieving state-of-the-art instance segmentation results. The current state of the art in the PASCAL VOC<sup>1</sup> semantic segmentation challenge is the DeepLab neural architecture [14], which has DeepLabv3+ [15] as the newest version, which outperforms other networks in the semantic segmentation task. The success of these architectures in their respective segmentation tasks leads us to believe that CNNs are the most promising approach to achieve good underwater segmentation results.

### **Underwater image segmentation**

Several approaches have been applied to the problem of underwater image segmentation. In [16], segmentation of underwater images technique is presented, using CLAHE histogram equalization followed by histogram thresholding. In [17], the underwater segmentation is performed by measuring the Mahalanobis distance between each pixel and the background color estimated from sample background images. In [18], a Particle Swarm Optimization (PSO) is used to maximize the entropy for underwater image segmentation. The same technique is adopted by [19] and [20], but using C-means to cluster the pixels. In [21], the underwater images are filtered with median filter, segmented them with K-Means clustering, and the image features are extracted using HOG, and then, used to classify the segments with an SVM classifier. Using a similar strategy, a novel solution [22] improve the selection of initial centroids of K-Means, which leads to better results, while increasing the computational cost. Also being a newer solution, in [23] an active contour strategy is used, minimizing an energy function to get the segmentation mask of the object in the underwater image. Already, in [24], a deep learning technique is used, in which a fully convolutional network is used to perform frame by frame fish segmentation in underwater videos. They use a weakly-labeled dataset of videos whose ground truth is derived from a motion-based background subtraction (BGS) technique [25] rather than manual annotations. The authors evaluate the precision and recall of their model in fish detection, but not the quality of the segmentation masks on a per pixel basis. In [26], a candidate object region is extracted from the image based on the presence of artificial light estimated from optical features. The region is segmented using parametric kernel graph cuts [27]. The main drawback of this method is to rely on the presence of artificial light in the image and therefore will not work properly in situations where the only source of illumination is natural light.

While these methods are able achieve to segment in certain situations, they still rely on heuristics or weakly labeled data. Inspired by the success of deep learning architectures in several difficult computer vision tasks, we aim to develop a more general solution, based on a reliable, manually labeled small set of training images.

### **Methodology**

The most straightforward approach to obtain powerful underwater image segmentation solution is to train a state-of-the-art segmentation CNN architecture with underwater images. The main obstacle to this approach, however, is that, to the best of our knowledge, no public adequate underwater segmentation dataset exists. The manual segmentation of

---

<sup>1</sup><http://host.robots.ox.ac.uk:8080/leaderboard/displaylb.php?challengeid=11&compid=6>

underwater images is a relatively simple, but labor-intensive process. Therefore, it would be extremely impractical to create a dataset as large as those usually used to train deep CNN architectures from scratch, as such models generally require thousands of samples to be properly trained.

We can circumvent this problem by pre-training the network using a large semantic segmentation dataset of non-underwater images and performing fine tuning using a much smaller dataset composed of manually segmented underwater images. The idea is low level features learned during the initial training help the network in the segmentation of underwater images. So, some datasets can be proposed, even with a relatively small number of images, to train deep learning segmentation models and to be used as a benchmark for comparison.

In the next sections, we present the proposed datasets created using both real and simulated images in the wild. Furthermore, we introduce the adopted neural architectures and the training process.

### Datasets

There are some datasets for specific underwater task, e.g., fish detection and classification<sup>2,3</sup>. However, these dataset are not compatible with our problem of underwater image segmentation in the wild, since they are focused in fishing. Thus, we created our datasets<sup>4</sup>.

*NAUTEC UWI Real*—Our *Real Underwater* dataset is composed of 700 underwater images in the wild collected from the Internet. The images were manually segmented in foreground and background pixels. We randomly use 400 images for training and 300 for testing. Three sample images from the dataset and their respective ground truth can be seen in Fig. 1. The dataset contains images acquired in several water conditions, illumination, and places, containing images in both benthic and pelagic zones without differentiating one from another. There are naturally and artificial lit images. Furthermore, divers, marine life, and many underwater objects are present in these images acquired in the wild. This dataset is available in an additional material of this work.

Manually segmenting underwater images is a labor-intensive, time-consuming task. Because of this, our real underwater dataset is relatively small. The use of simulated images can increase the amount of training data, which can be created by simulating the effects of underwater degradation on non-underwater images whose segmentation labels are available. These effects can be created according to the Jaffe-McGlamery optical model [29, 30], as adopted in [31]. We adopted a simplified version of the model where the forward scattering is neglected since the backscattering is the principal responsible by the image degradation [32]. We also use a set of real underwater image patch from a backscattering area that provide us medium parameters. These simulated effects are similar to those presented by Duarte et al. [33].

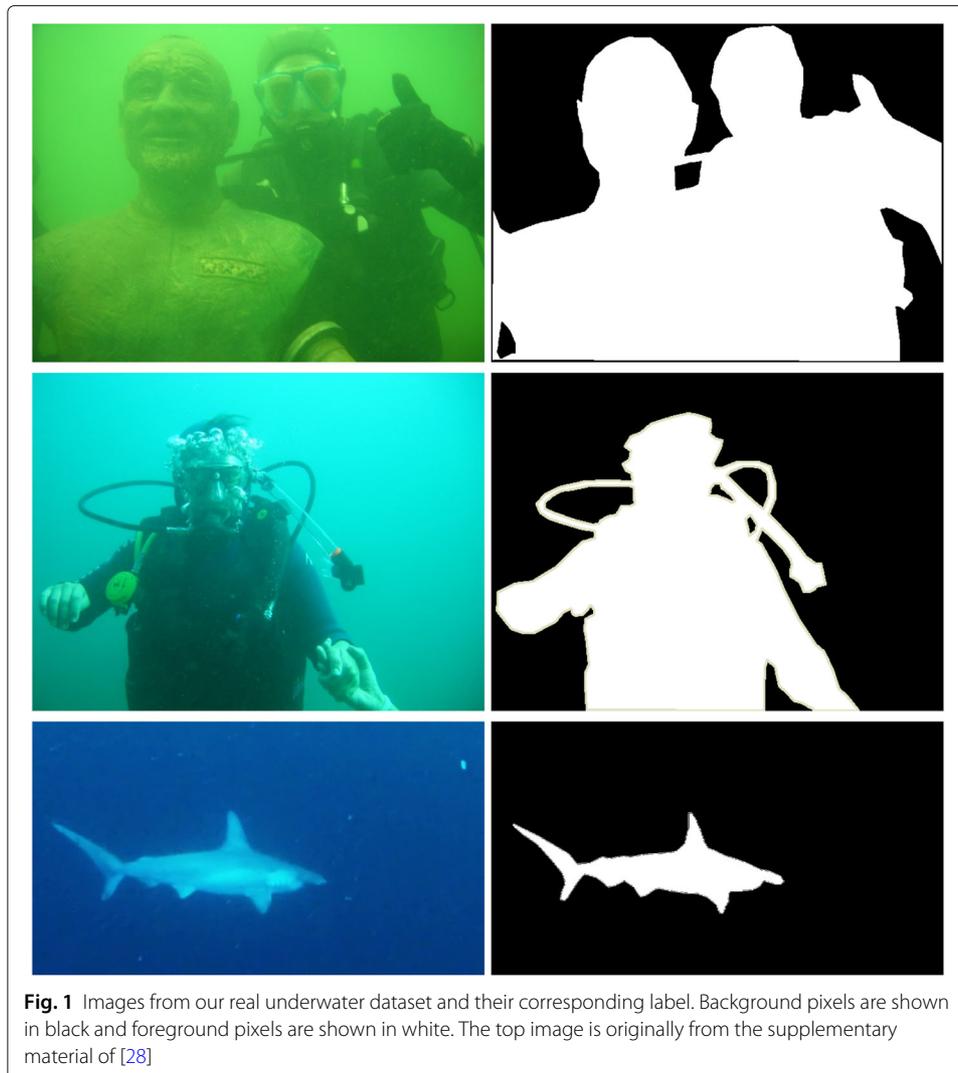
However, the model requires the availability of the image's depth map. While we believe outdoor scenes would be more adequate as they are closer to subsea images, we are forced to base our simulation on indoor images. To the best of our knowledge, there are no publicly available outdoor datasets with high-quality depth maps. Despite the obvious differences between indoor and underwater scenes, the network is expected to learn how

---

<sup>2</sup>NOAA dataset - <https://swfscdata.nmfs.noaa.gov/labeled-fishes-in-the-wild/>

<sup>3</sup>Ocean Networks Canada - <https://www.oceannetworks.ca/crowd-truthing-experiment>

<sup>4</sup>NAUTEC UWI1 datasets - <https://1drv.ms/f/s!ApAbq4Ufbfzjhze6ttiTxdpMg9i>



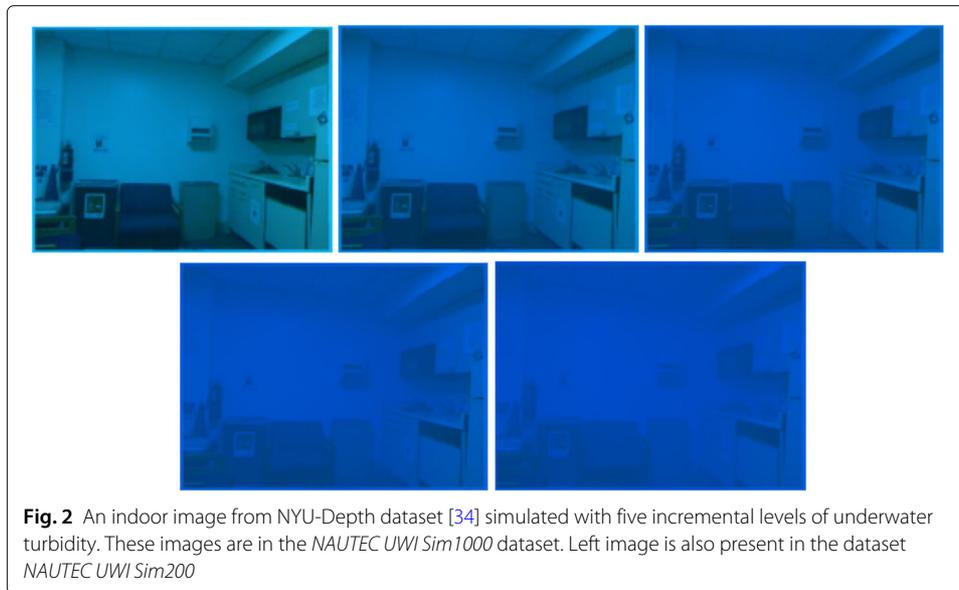
to perform the segmentation of objects obscured by underwater degradation in a more general way by using these data. This capability of the network to learn the attenuation effect generated by the light traveling in the water can be achieved using this simulated dataset [31].

We use NYU Depth V2 [34] as the basis of our simulated data. This dataset provides images with segmentation labels and high-quality depth maps. We modified the original segmentation labels by considering pixels labeled as wall, roof, floor, etc., as background and pixels labeled as objects as foreground. We create three additional datasets using this data:

*NAUTEC UWI Sim200*: This dataset is composed of 200 simulated images with relatively low turbidity. A simulated image of this dataset is present in the left of Fig. 2.

*NAUTEC UWI Mixed*: This dataset is the union of the Sim200 and Real Underwater datasets.

*NAUTEC UWI Sim1000*: This dataset is composed of 1000 simulated underwater images with four additional levels of increasing simulated underwater turbidity in rela-



tion to the *Sim200*. An image from this dataset with its five levels of turbidity is shown in Fig. 2.

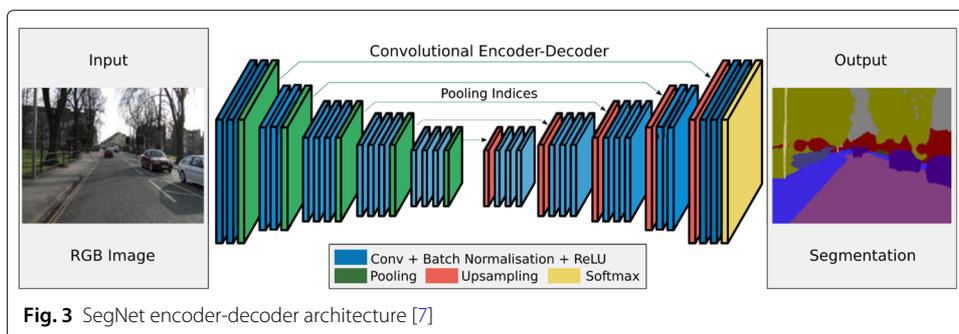
### Network architectures

Designing a deep learning architecture from scratch is an arduous and time-consuming task. Thus, we evaluated two well-known semantic segmentation architectures in this work: SegNet [7] and DeepLabv3+ [15]. The evaluated datasets for inland images have been previously described. We have used 10% of the training dataset for validation.

#### SegNet

The SegNet is a fully convolution encoder-decoder semantic segmentation architecture, as shown in Fig. 3. Its encoder network is topologically identical to the 13 convolutional layers in the VGG16 [35] image classification network. The main advantage of the SegNet over competing segmentation architectures is the reduction in memory use provided by its decoder network architecture. We choose to evaluate the SegNet because it is a classical image segmentation architecture based on CNNs.

We run 50,000 training iterations with a batch size of 5 using the Adam optimization algorithm [36] with a learning rate of  $1.5 \times 10^{-4}$ ,  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$  and  $\epsilon = 10^{-8}$ . The network is initialized using random weights. Furthermore, we also pre-trained the



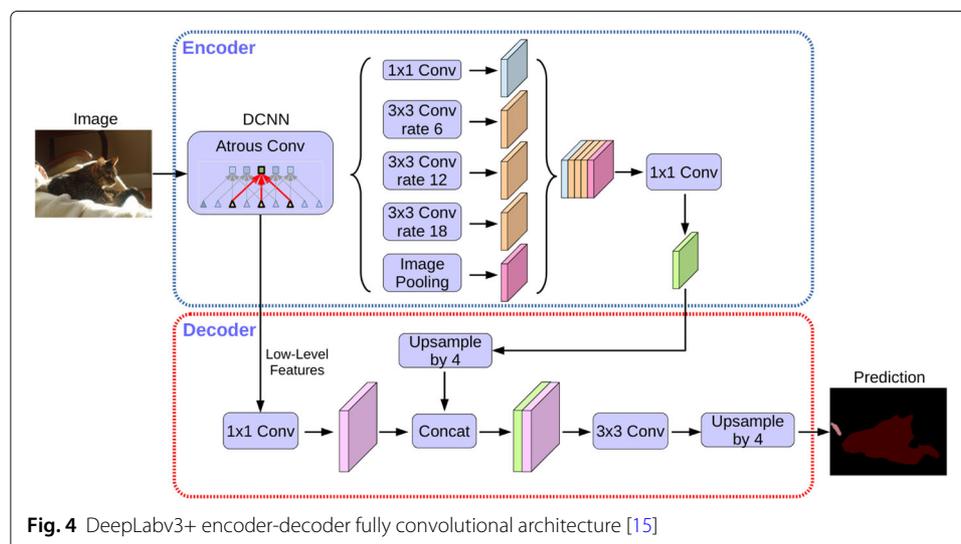
weights using the dataset PASCAL VOC dataset [37]. As our objective is to segment underwater images in foreground and background, we ignore the class information and simply consider pixels labeled as any of the 20 Pascal VOC classes to be foreground. We use 10,582 images for training and 300 for validation in the pre-training step. In this case, we resize the inputs to a width of  $480 \times 360$  pixels and train until the convergence.

The network parameters are also trained using underwater images. We evaluate the network using the previously described underwater datasets for 50,000 iterations using the same hyperparameters.

### DeepLabv3+

As the most deep learning segmentation architectures, DeepLabv3+ is an encoder-decoder network; moreover, it is fully convolutional and, currently, one of the top ranked architectures in the PASCAL VOC segmentation challenge. The main idea behind the architecture is to preserve spatial information by reducing the number of strided pooling operations. Furthermore, atrous convolutions compensate the reduction in the receptive field. Another adopted technique is the detection of objects at multiple scales by parallel atrous convolutions at different sampling rates. All these characteristics are showed in the DeepLabv3+ architecture, as shown in Fig. 4. We choose the DeepLab architecture because it is the state-of-the-art in semantic segmentation.

Training a large architecture such as Deeplab from scratch is difficult, specially when the amount of available data is limited. Differently from Segnet that is initialized using random weights, we start the training by initializing the model with Xception [38] backbone weights. We also pre-trained the model on the PASCAL VOC dataset for 20,000 iterations with a batch size of 8, randomly cropping the inputs to the size of  $513 \times 513$ . We employ common data augmentation methods, such as input scaling and mirroring. We use SGD with a momentum of 0.9 and polynomial learning rate decay with a base learning rate of  $10^{-4}$  and  $power = 0.9$ . Weight decay is set to  $4 \times 10^{-5}$ . Finally, we train the model using the previously described underwater datasets for 20000 iterations using the same training setup.



## Experimental results

We evaluate our models on the remaining 300 randomly selected images from the real underwater dataset that were not presented in the training step. The results are evaluated using the standard mean Intersection over Union (mIoU) metric. We also take the raw network output, with no additional post-processing.

For the sake of a fair comparison, we also evaluate the networks using a state-of-the-art underwater image restoration algorithm as a pre-processing step. Our idea is to reduce the effects of the water that makes the segmentation difficult. We adopted the Underwater Dark Channel Prior method (*UDCP*) [39] and the Underwater GAN (*UGAN*) [40] before segmenting the image with the models trained only on the PASCAL VOC dataset. Although the pre-processing using restoration methods sounds a promising idea, mainly with *UGAN*, the results are not competitive.

We do not present a comparison with underwater segmentation methods. All of them use classical methodologies that are unable to segment the underwater images in a proper way. We found the results of these methods to not be even remotely competitive to the results obtained using deep neural network.

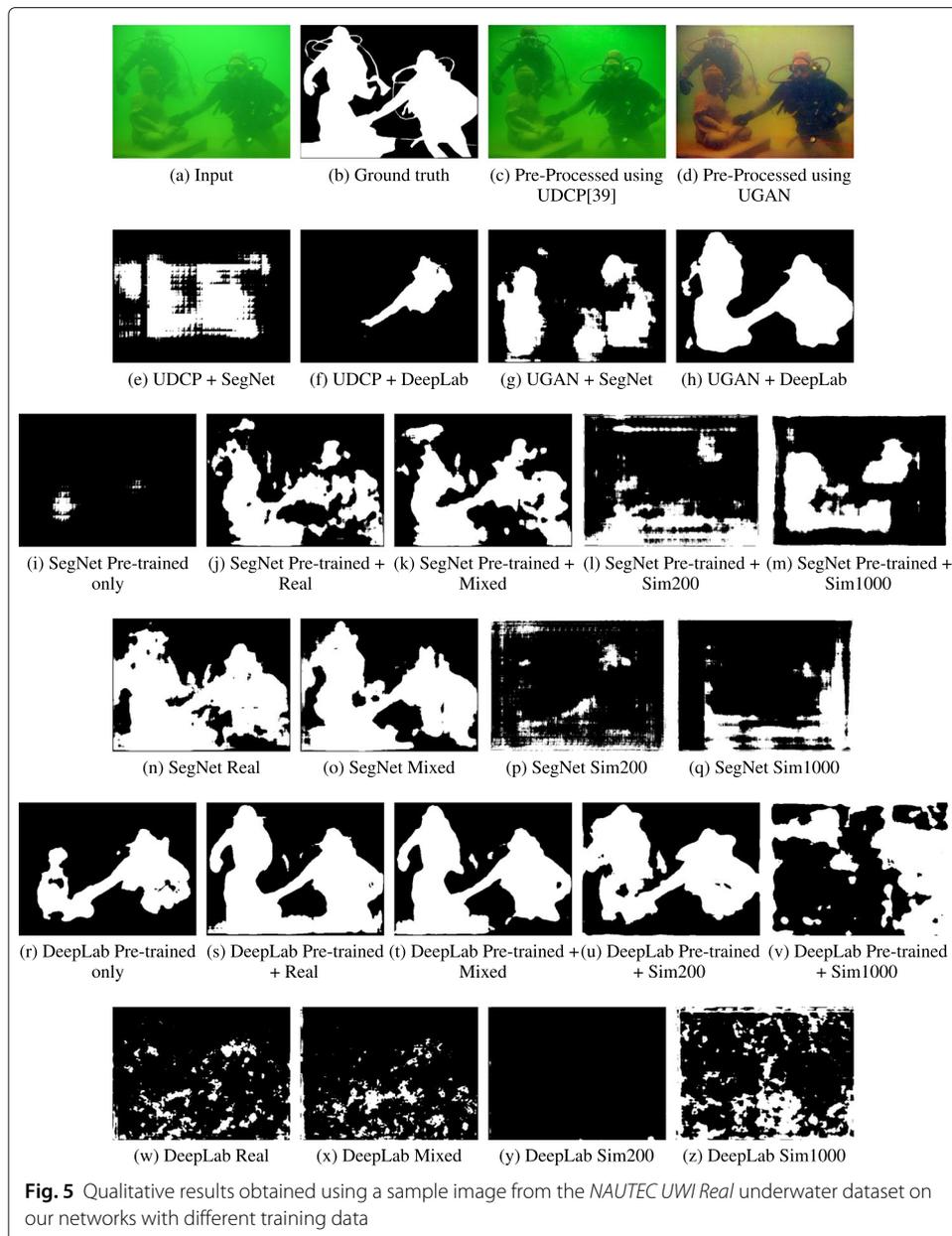
Table 1 shows the mIoU accuracy for all evaluated models. Figures 5, 6, and 7 show qualitative underwater segmentation results.

The results show a CNN approach is a viable approach to the task of underwater image segmentation even though a limited amount of training images. The best network is achieved using a DeepLab architecture trained using our *NAUTEC UWI Real* underwater

**Table 1** Results of our underwater segmentation models, averaged over the 300 test images

Architect.	Pre-processed	PASCAL Pre-trained	Training data	mIoU
SegNet	UDCP	Yes	-	0.415
SegNet	UGAN	Yes	-	0.421
SegNet	-	No	NAUTEC UWI Real	0.825
SegNet	-	No	NAUTEC UWI Mixed	0.805
SegNet	-	No	NAUTEC UWI Sim200	0.522
SegNet	-	No	NAUTEC UWI Sim1000	0.565
SegNet	-	Yes	-	0.444
SegNet	-	Yes	NAUTEC UWI Real	0.796
SegNet	-	Yes	NAUTEC UWI Mixed	0.795
SegNet	-	Yes	NAUTEC UWI Sim200	0.558
SegNet	-	Yes	NAUTEC UWI Sim1000	0.556
DeepLab	UDCP	Yes	-	0.434
DeepLab	UGAN	Yes	-	0.485
DeepLab	-	No	NAUTEC UWI Real	0.689
DeepLab	-	No	NAUTEC UWI Mixed	0.617
DeepLab	-	No	NAUTEC UWI Sim200	0.445
DeepLab	-	No	NAUTEC UWI Sim1000	0.481
DeepLab	-	Yes	-	0.444
<b>DeepLab</b>	<b>-</b>	<b>Yes</b>	<b>NAUTEC UWI Real</b>	<b>0.919</b>
DeepLab	-	Yes	NAUTEC UWI Mixed	0.909
DeepLab	-	Yes	NAUTEC UWI Sim200	0.762
DeepLab	-	Yes	NAUTEC UWI Sim1000	0.751

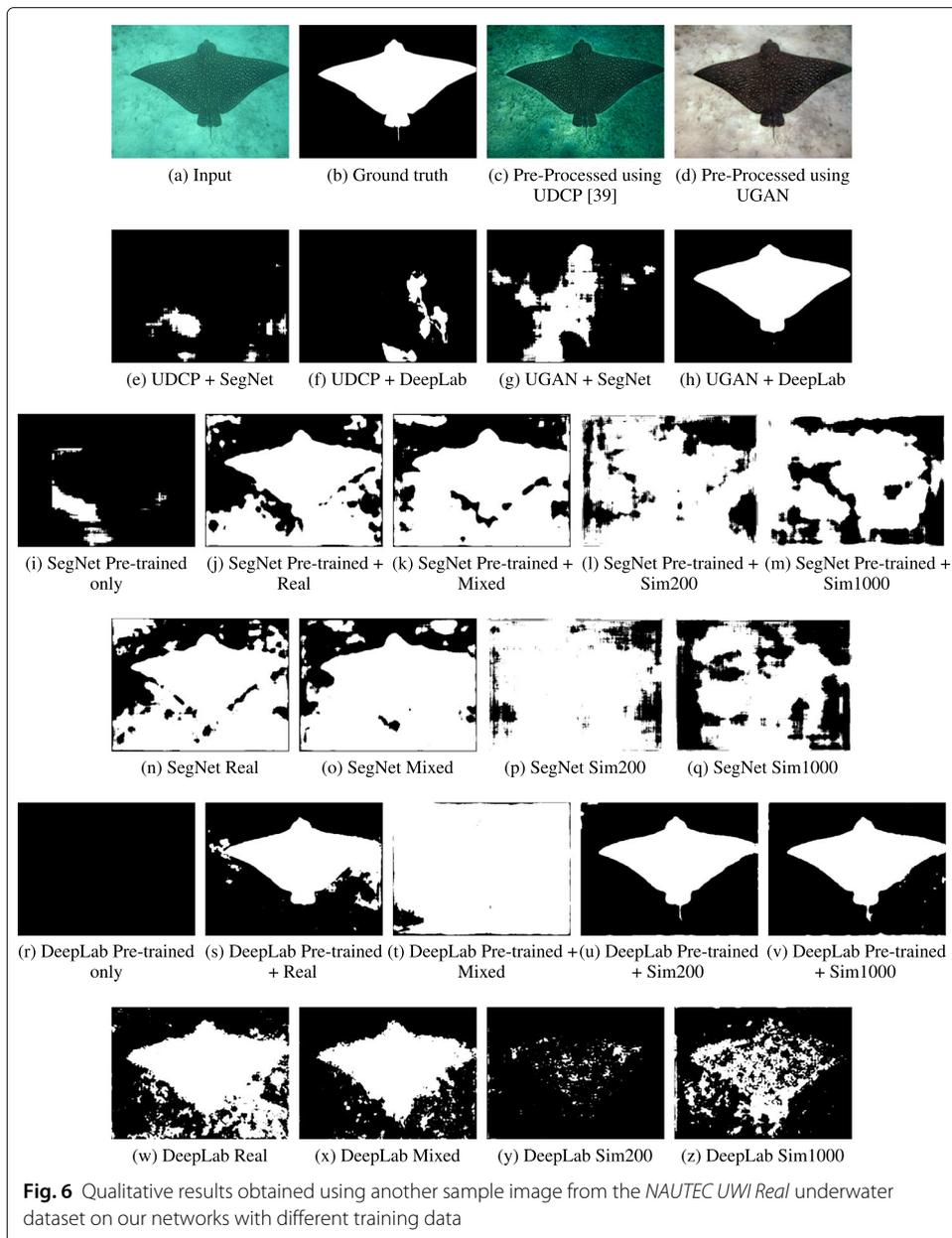
The model with the highest mIoU is highlighted in boldface



dataset with initial weights of a model pre-trained on the PASCAL VOC dataset. The network obtains  $\approx 91.9\%$  mIoU accuracy in a test set composed of 300 underwater images in the wild.

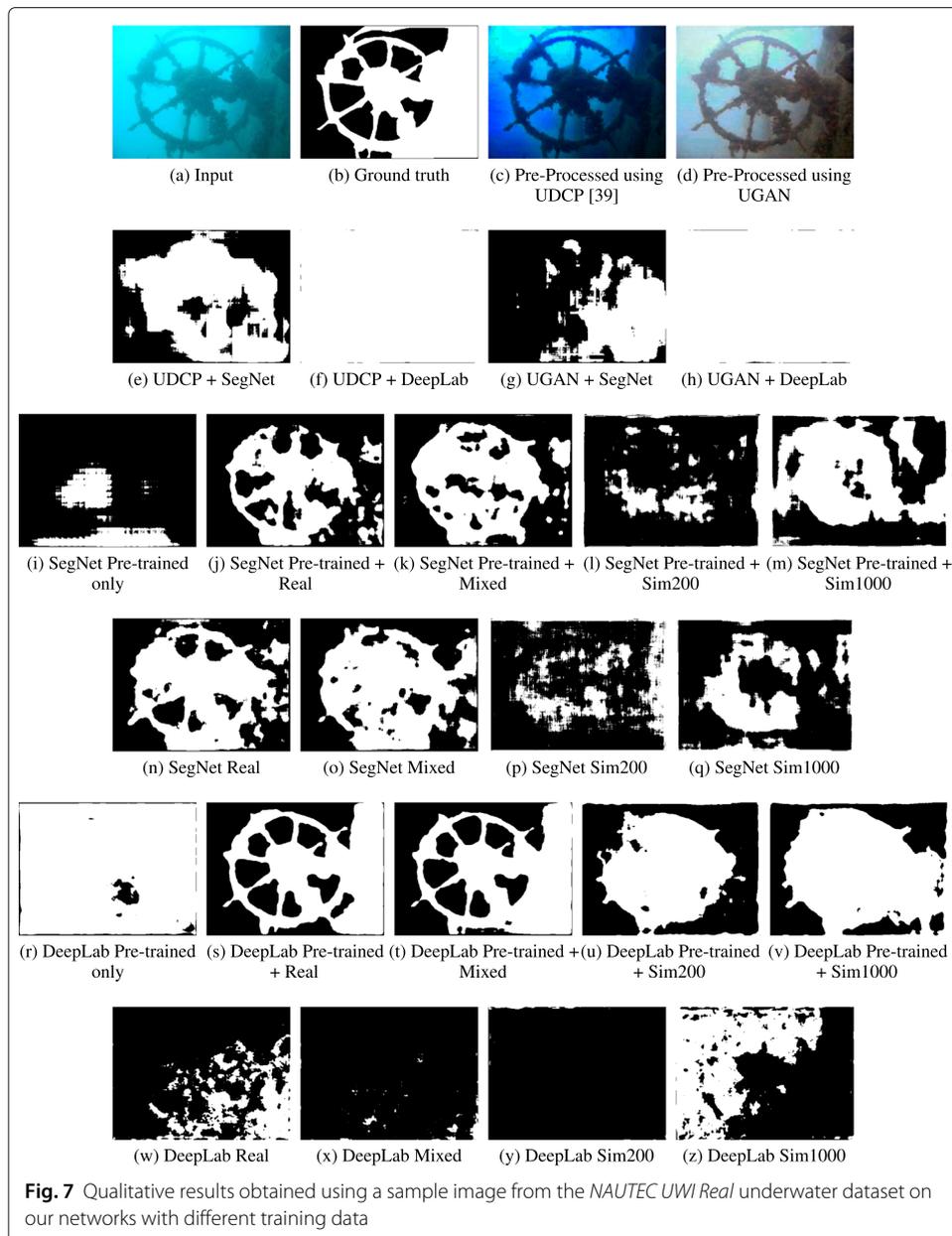
The segmentation performance is slightly reduced when the real dataset is augmented with simulated images. Models only trained with simulated data could not produce satisfactory results, but are still preferable over no fine-tuning at all.

The main reason is our simulated images are based on indoor scenes that is distinct from the real images of the testing set. The use of simulated images is due to the requirement of depth maps that can be only properly obtained in indoor environments, as described in the “[Datasets](#)” section. However, models trained with simulated data perform better than the ones trained with real data in images where the background is the



sea floor rather than pure water, such as the sample image shown in Fig. 6. We believe the simulated dataset present a diverse structure in relation to the background that is similar to presented in this testing image. Thus, the networks trained with this dataset is able to segment in a better way than the networks using only real data in this type of test image.

The results of pre-trained DeepLabv3+ are better than the pre-trained SegNet. However, the results without pre-training in SegNet are better. We believe the main reason for this is the larger size and complexity of the DeepLab architecture. Larger architectures are generally more prone to overfitting that imply inferior generalization performance, specially when the amount of training data is limited. Despite this, the SegNet adopted in this work lacks in size and complexity to achieve a competitive performance, even given the



relatively small amount of training samples. In addition to that, it is normal that a more complex network has a higher computational cost, what is also true in this case. When running the validation process, using a computer with an Intel Core i7-7700K, 16GB RAM and a NVIDIA GeForce GTX Titan X 12GB, DeepLabv3+ achieve a mean of 10.06 FPS, already SegNet achieve 16.54 FPS.

Surprisingly, the use of pre-processing with UDCP [39] harmed the segmentation performance of both architectures. However, UGAN [40] improved the performance in DeepLab segmentation. We believe this happened because UDCP tend to produce some artifacts that may confuse the network, leading to inaccurate segmentation. Already, UGAN as a deep learning technique reaches better results in restoration process without producing to much artifacts that can confuse the network.

## Conclusion

In this paper, we presented a set of datasets to train deep CNN architectures to the task of underwater image segmentation in the wild. We evaluated the impact of pre-training and simulated training data on the network performance. We also present a working solution based on DeepLabv3+ image segmentation architecture achieving a mIoU accuracy of  $\approx 91.9\%$  on a random test set of 300 real underwater images. We prove that this network architecture is able to properly segment with a small number of training images. Qualitative evaluation leads us to believe that our results are superior to those of traditional underwater segmentation methods. Another important contribution is our publicly available dataset of 700 manually segmented underwater images in the wild and their respective ground truths. To the best of our knowledge, we are the first work to present a CNN approach to underwater image segmentation in the wild.

Future work includes the evaluation of other network architectures and generative adversarial networks [41], which could help in removing small artifacts that are not correctly penalized by simple loss functions. We also plan to increase the number of images of our real dataset.

## Acknowledgements

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research. This paper is also a contribution of the Brazilian National Institute of Science and Technology - INCT-Mar COI funded by CNPq Grant Number 400551/2014-4. We also would like to thank the colleagues from NAUTEC-FURG.

## Authors' contributions

Conceptualization, Paulo Drews-Jr; data curation, Isadora Souza; funding acquisition, Paulo Drews-Jr and Silvia Botelho; methodology, Paulo Drews-Jr and Isadora Souza; project administration, Paulo Drews-Jr and Silvia Botelho; resources, Paulo Drews-Jr and Silvia Botelho; software, Isadora Souza, Igor Maurell, and Eglon Protas; supervision, Paulo Drews-Jr and Silvia Botelho; validation, Paulo Drews-Jr, Isadora Souza, Igor Maurell and Eglon Protas; writing—original draft, Paulo Drews-Jr, Isadora Souza, Igor Maurell and Eglon Protas; writing—review & editing, Paulo Drews-Jr, Eglon Protas and Silvia Botelho. The authors read and approved the final manuscript.

## Funding

This research was partly funded by CNPQ, FAPERGS, and ANP-PRH 27.

## Availability of data and materials

The datasets generated and used in this research are referenced in the text using footnotes.

## Declarations

### Competing interests

The authors declare that they have no competing interests.

Received: 29 March 2020 Accepted: 30 August 2021

Published online: 04 October 2021

## References

1. Fabric JN, Turla IE, Capacillo JA, David LT, Naval PC (2013) Fish population estimation and species classification from underwater video sequences using blob counting and shape analysis. In: 2013 IEEE International Underwater Technology Symposium (UT). pp 1–6. <https://doi.org/10.1109/UT.2013.6519876>
2. Donaldson JA, Drews-Jr P, Bradley M, Morgan DL, Baker R, Ebner BC (2019) Countering low visibility in video survey of an estuarine fish assemblage. *Pac Conserv Biol* 26:190–200
3. Drews-Jr P, Hernández E, Elfes A, Nascimento ER, Campos M (2016) Real-time monocular obstacle avoidance using underwater dark channel prior. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp 4672–4677. <https://doi.org/10.1109/IROS.2016.7759687>
4. Gaya JO, Gonçalves LT, Duarte AC, Zanchetta B, Drews-Jr P, Botelho SSC (2016) Vision-based obstacle avoidance using deep learning. In: 2016 XIII Latin American Robotics Symposium and IV Brazilian Robotics Symposium (LARS/SBR). pp 7–12. <https://doi.org/10.1109/LARS-SBR.2016.9>
5. Schechner YY, Karpel N (2004) Clear underwater vision. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) Vol. 1. pp 1–1. <https://doi.org/10.1109/CVPR.2004.1315078>
6. Drews-Jr P, do Nascimento E, Moraes F, Botelho S, Campos M (2013) Transmission estimation in underwater single images. In: 2013 IEEE International Conference on Computer Vision Workshops. pp 825–830. <https://doi.org/10.1109/ICCVW.2013.113>

7. Badrinarayanan V, Kendall A, Cipolla R (2017) Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39(12):2481–2495. <https://doi.org/10.1109/TPAMI.2016.2644615>
8. Weiss K, Khoshgoftaar TM, Wang D (2016) A survey of transfer learning. *J Big Data* 3(1):9. <https://doi.org/10.1186/s40537-016-0043-6>
9. Cimpoi M, Maji S, Vedaldi A (2015) Deep filter banks for texture recognition and segmentation. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp 3828–3836. <https://doi.org/10.1109/CVPR.2015.7299007>
10. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) Imagenet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. pp 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>
11. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp 3431–3440. <https://doi.org/10.1109/CVPR.2015.7298965>
12. He K, Gkioxari G, Dollár P, Girshick RB (2017) Mask R-CNN. *CoRR abs/1703.06870*. [1703.06870](https://arxiv.org/abs/1703.06870)
13. Ren S, He K, Girshick R, Sun J (2015) Faster r-cnn: Towards real-time object detection with region proposal networks. In: Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1, NIPS'15. MIT Press, Cambridge. pp 91–99. <http://dl.acm.org/citation.cfm?id=2969239.2969250>
14. Chen L, Papandreou G, Schroff F, Adam H (2017) Rethinking atrous convolution for semantic image segmentation. *CoRR abs/1706.05587*. [1706.05587](https://arxiv.org/abs/1706.05587)
15. Chen L-C, Zhu Y, Papandreou G, Schroff F, Adam H (2018) Encoder-decoder with atrous separable convolution for semantic image segmentation:801–818. [https://doi.org/10.1007/978-3-030-01234-2\\_49](https://doi.org/10.1007/978-3-030-01234-2_49)
16. Rai RK, Gour P, Singh B (2012) Underwater image segmentation using clahe enhancement and thresholding. *Int J Emerging Technol Adv Eng* 2:118–123
17. Kim E, Lee S (2013) Comparative studies of remove background algorithms for objects extraction of underwater images. *Int J Softw Eng Appl* 7:459–468
18. Zhang R, Liu J (2006) Underwater image segmentation with maximum entropy based on particle swarm optimization (pso). In: First International Multi-Symposiums on Computer and Computational Sciences (IMSCCS'06) Vol. 2. pp 360–636. <https://doi.org/10.1109/IMSCCS.2006.280>
19. Wang S, Xu Y, Pang Y (2011) A fast underwater optical image segmentation algorithm based on a histogram weighted fuzzy c-means improved by pso. *J Mar Sci Appl* 10(1):70–75. <https://doi.org/10.1007/s11804-011-1043-8>
20. Li X, Song J, Zhang F, Ouyang X, Khan SU (2016) Mapreduce-based fast fuzzy c-means algorithm for large-scale underwater image segmentation. *Futur Gener Comput Syst* 65:90–101. <https://doi.org/10.1016/j.future.2016.03.004>. Special Issue on Big Data in the Cloud
21. Rajasekar M, Aruldoss CK, Anto Bennet M (2015) Underwater k-means clustering segmentation using svm classification. *Middle-East J Sci Res* 23:2166–2172
22. Chen W, He C, Ji C, Zhang M, Chen S (2021) An improved k-means algorithm for underwater image background segmentation. *Multimedia Tools Appl* 80:1–25
23. Liu Y, Li H (2020) Design of refined segmentation model for underwater images. In: 2020 5th International Conference on Communication, Image and Signal Processing (CCISP). pp 282–287. <https://doi.org/10.1109/CCISP51026.2020.9273503>
24. Labao AB, Naval PC (2017) Weakly-labelled semantic segmentation of fish objects in underwater videos using a deep residual network. In: Nguyen NT, Tojo S, Nguyen LM, Trawiński B (eds). *Intelligent Information and Database Systems*. Springer, Cham. pp 255–265
25. Zivkovic Z (2004) Improved adaptive gaussian mixture model for background subtraction. In: Proceedings of the 17th International Conference on Pattern Recognition, 2004. *ICPR 2004 Vol. 2*. pp 28–312. <https://doi.org/10.1109/ICPR.2004.1333992>
26. Chen Z, Zhang Z, Bu Y, Dai F, Fan T, Wang H (2018) Underwater object segmentation based on optical features. *Sensors* 18(1). <https://doi.org/10.3390/s18010196>
27. Salah MB, Mitiche A, Ayed IB (2011) Multiregion image segmentation by parametric kernel graph cuts. *IEEE Trans Image Process* 20(2):545–557. <https://doi.org/10.1109/TIP.2010.2066982>
28. Ancuti C, Ancuti CO, Haber T, Bekaert P (2012) Enhancing underwater images and videos by fusion. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. pp 81–88. <https://doi.org/10.1109/CVPR.2012.6247661>
29. McGlamery BL (1980) A computer model for underwater camera systems. In: Proceedings of SPIE Vol. 0208. pp 221–231. <https://doi.org/10.1117/12.958279>
30. Jaffe JS (1990) Computer modeling and the design of optimal underwater imaging systems. *IEEE J Ocean Eng* 15(2):101–111. <https://doi.org/10.1109/48.50695>
31. Gonçalves L, Gaya J, Drews-Jr P, Botelho S (2017) Deepdive: an end-to-end dehazing method using deep learning. In: 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). pp 436–441. <https://doi.org/10.1109/SIBGRAPI.2017.64>
32. Schechner YY, Karpel N (2005) Recovery of underwater visibility and structure by polarization analysis. *IEEE J Ocean Eng* 30(3):570–587
33. Duarte A, Codevilla F, Gaya JDO, Botelho SSC (2016) A dataset to evaluate underwater image restoration methods. In: OCEANS 2016 - Shanghai. pp 1–6. <https://doi.org/10.1109/OCEANSAP.2016.7485524>
34. Silberman N, Hoiem D, Kohli P, Fergus R (2012) Indoor segmentation and support inference from rgb-d images. In: Fitzgibbon A, Lazebnik S, Perona P, Sato Y, Schmid C (eds). *Computer Vision – ECCV 2012*. Springer, Berlin. pp 746–760
35. Simonyan K, Zisserman A (2015) Very Deep Convolutional Networks for Large-Scale Image Recognition. In: Bengio Y, LeCun Y (eds). 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings. <http://arxiv.org/abs/1409.1556>. Accessed 17 July 2019
36. Kingma DP, Ba J (2015) Adam: A Method for Stochastic Optimization. In: Bengio Y, LeCun Y (eds). 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings. <http://arxiv.org/abs/1412.6980>. Accessed 25 July 2019

37. Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A (2010) The Pascal Visual Object Classes (VOC) Challenge. *Int J Comput Vis* 88(2):303–338. <https://doi.org/10.1007/s11263-009-0275-4>
38. Chollet F (2017) Xception: deep learning with depthwise separable convolutions. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). <https://doi.org/10.1109/cvpr.2017.195>
39. Drews-Jr P, Nascimento E, Botelho S, Campos M (2016) Underwater depth estimation and image restoration based on single images. *IEEE Comput Graphics Appl* 36(2):24–35. <https://doi.org/10.1109/MCG.2016.26>
40. Fabbri C, Islam MJ, Sattar J (2018) Enhancing underwater imagery using generative adversarial networks. In: 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE. pp 7159–7165. <https://doi.org/10.1109/icra.2018.8460552>
41. Ledig C, Theis L, Huszar F, Caballero J, Aitken A, Tejani A, Totz J, Wang Z, Shi WE (2016) Photo-realistic single image super-resolution using a generative adversarial network. *CoRR abs/1609.04802*. <https://doi.org/10.1109/cvpr.2017.19>

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)

---