This item was submitted to [Loughborough's Research Repository](#) by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

Real-time and embedded compact deep neural networks for seagrass monitoring

PLEASE CITE THE PUBLISHED VERSION

<https://doi.org/10.1109/SMC42975.2020.9283372>

PUBLISHER

IEEE

VERSION

AM (Accepted Manuscript)

PUBLISHER STATEMENT

Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

LICENCE

All Rights Reserved

REPOSITORY RECORD

Wang, Jiangtao, Baihua Li, Yang Zhou, Qinggang Meng, Sante Rende, and Emanuele Rocco. 2020. "Real-time and Embedded Compact Deep Neural Networks for Seagrass Monitoring". Loughborough University. <https://hdl.handle.net/2134/12964874.v1>.

Real-time and Embedded Compact Deep Neural Networks for Seagrass Monitoring

Jiangtao Wang
Computer Science
Loughborough University
Loughborough, UK
J.WANG5@lboro.ac.uk

Baihua Li*
Computer Science
Loughborough University
Loughborough, UK
B.LI @lboro.ac.uk

Yang Zhou
Computer Science
Loughborough University
Loughborough, UK
Y.ZHOU5@lboro.ac.uk

Qinggang Meng
Computer Science
Loughborough University
Loughborough, UK
Q.MENG@lboro.ac.uk

Sante Francesco Rende
ISPRA
National Institute for Environmental
Protection and Research
Rome, Italy
Francesco.rende@isprambiente.it

Emanuele Rocco
Witted Srl
Rovereto, Italy
Emanuele@witted.it

Abstract— We propose an efficient and robust segmentation network for automated seagrass region detection. The proposed network has a simple architecture to save computational demands as well as inference energy cost. More importantly, the scale of network can be feasibly adjusted, to balance the network computational demands and segmentation accuracy. Experimental results show that our proposed network is robust to segment the various seagrass patterns with 90.66% mIoU (mean Intersection over Union) accuracy. It had achieved 200 frames per second (FPS, 1.42 times faster than the second-best network GCN) on desktop GPU, and 18 FPS on NVIDIA Jetson TX2. It also has 3.45M parameters and 0.587 GMACs FLOPs (FLoating Point Operations), only 14.6% and 10.8% of those in GCN respectively. To segment a single image on the Jetson TX2, our architecture requires an average energy of 0.26 Joule. This energy cost is only 46% of DeepLab, which shows the proposed network to be an energy efficient one. The proposed network demonstrates accurate and real-time segmentation capability, and it can be deployed to low-energy embedded AUVs for sea habitat protection.

Keywords— Deep learning, semantic segmentation, seagrass monitoring, marine science, embedded systems.

I. INTRODUCTION

Seagrass are considered one of the most ecological important coastal habitats, supporting many marine species, regulating water quality, and providing a long-term carbon sequestration. Effective management of seagrass meadows is also fundamental for capturing pools of high value and climate change mitigation. For this AUVs (Autonomous Underwater Vehicles) can be used to monitor the health status of seagrass meadows on a regular basis without extra human assistance.

However, on board analysis poses many challenges. On the vision side, the water visibility conditions vary over time, geographical locations and even depths. The seagrass meadows themselves change color and texture over the year because of its natural life cycle. During the mission, the camera poses and distance with respect to the benthic habitats will vary as well. To accurately recognize and precisely segment them, the AUVs vision methods need to be robustly designed.

On the hardware side, the AUVs they rely on embedded electronics and batteries. They have so limited capability in terms of computing energy as memory and FLOPs and energy autonomy. The AUVs intelligence needs so also to be optimized to satisfy such strict hardware constraints. To address the above challenges, we propose a novel segmentation network, which is both accurate, compact and energy efficient. Its encoder-decoder structure is shown in Fig. 1. We considered that the network encoder usually requires the most FLOPs and parameters (see Table II). So, to reduce the network total FLOPs, we adopted the light-weight backbone MobileNet [16] as encoder. To further reduce the computational load, we then pursued a light encoder with as few convolution layers as possible.

Our main contributions are then as follows:

1) We proposed a segmentation network with a compact structure with only 3.45M parameters and 0.587G FLOPs (with 256x256 input resolution). Comparing with state of art networks such as SegNet, it allows a reduction of 99% on FLOPs. The size of proposed network is feasible to control by simply setting hyper-parameters, to enable compress the network in further.

2) We showed that our network still achieves state-of-the-art accuracy with 90.66% mIoU on a seagrass public dataset. The achieved accuracy is significantly better than the baseline seagrasses segmentation result of 80.9% reported in [1].

3) The proposed network trained on seagrass images taken at 0~2m seafloor distance, is able to infer images taken from 2~6m with 77.73% mIoU. The proposed network allows for changes of seafloor distance and visibility conditions.

4) We demonstrated, that, given input image resolution of 256x256 pixels, the proposed segmentation network achieves real-time seagrass segmentation on embedded platforms (Jetson TX2) at 18FPS with just 0.26 Joule energy consumption per inference. We showed so that such a network could be deployed on AUVs embedded devices with reduced energy needs.

II. RELATED WORK

With global coverage of around 3.45x10⁵ km² [4, 12] and growing along most coastlines of every continent except

*Corresponding author

Fig. 1. Overall architecture of the proposed segmentation neural network. It consists of the encoder, decoder and the skip connections. The encoder adopts the same architecture of MobileNet [28]. The decoder comprises five groups of up-sampling layers where the red arrows refer to the skip connections. The numbers over the blocks refer to the features size of output, and below ones indicate the number of output channels. We set α and β to 1 defaultly as default.

Antarctica [6], seagrass meadows are one of the most effective carbon sinks on Earth. However human direct or indirect disturbances, including those from global anthropogenic changes, impact the seagrass meadows ecosystems [6, 12] and threaten their capacity as key carbon capture agent. There is the clear need of frequently monitoring the seagrass ecology to assess possible meadow damage and assess its development over time to better plan, execute and validate conservation efforts. This can be achieved by taking underwater images of the benthic habitat and by measuring the seagrass meadow coverage and distribution to estimate key seascape multi-metric indices, such as the conservation index CI Index [8] and the seagrass habitat structure index [9].

Traditional machine learning methods for seagrass detection, with reduced or without human interaction, as proposed by [1, 2, 10-13], can achieve accurate results. To start with, these methods separate a full resolution seagrass image into small patches or super-pixel regions. Then, from those, they extract the feature maps with various feature extraction and image processing methods: Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP) or recently deep Convolutional Neural Networks (CNN). Finally, by using different kinds of machine learning classifiers, such as Support Vector Machine (SVM) or Logistic Model Tree (LMT), those patches and regions are classified and then mapped to pixel-wise classified (or segmentation).

However, it is still difficult to deploy these methods on embedded electronics for real-time analysis on AUVs. To execute the multiple and complex processing steps for image processing, they in fact need many resources, as memory, and FLOPs. On the embedded boards, if they can at all, these traditional methods take so a long time to extract images features and inference the result. There is so the need for light, efficient and accurate algorithms for seagrass segmentation to be used on embedded platforms.

End2End deep learning (DL) networks can address the above segmentation problems. By easily leveraging GPU resources, they can perform faster end-to-end analysis than traditional methods. They also have demonstrated on AUVs [7, 11, 13-14] as well. Yet, such proposed DL methods need a very complex architecture with five DL models to be inferred at the same time

and then averaged. Such complexity poses a computational burden on the real-time inference performances. So even if such DL networks work well on desktop servers with powerful GPUs, it is challenging to run them on the limited computational resource of AUVs embedded systems.

III. APPROACH

As shown in Fig. 1, the overall architecture of our proposed segmentation network consists of an encoder network followed up by a decoder network. They are linked together through skip-connections. The encoder compresses the input image into multi-scale feature maps by down-sampling; the decoder enables to reconstruct the same resolution segmentation by up-sampling. Skip-connections forward further spatial information and low-level features from the feature maps at different encoder layers to corresponding decoder layers. Below we look into these three components in details considering the computational constraints posed by the embedded systems.

A. Efficient Encoder

The objective is to determine a backbone network, which is structured as the encoder. It must keep the network parameter number and overall FLOPs as small as possible, while still be effective on extracting the feature maps. To flexibly address the trade-off between computation cost and network accuracy performance on embedded devices further, the encoder scale should be adjustable. This allows the network to be flexible and meet the specific energy and computation cost requirement on a chosen embedded platform.

Inspired by [19 - 23] and as shown in Fig. 1, we employ the basic MobileNet [16] without classification layers as encoder. The encoder has nine convolution layers and blocks and down-samples the input image to 1/32 resolution feature maps. According to the theoretical analysis of [16], it can significantly reduce the convolution computing demands, by utilizing Inversed Residual Convolution in down-sampling blocks. MobileNet is capable of maintaining a high segmentation accuracy at a significantly reduced network parameters and FLOPs. In fact, assuming 256x256 input resolution, it has only 9.4% parameters of ResNet-50, and 15.1% of VGG-16, 7.5% FLOPs of ResNet-50, and 2% of VGG-16.

Fig. 2. The structure of up-sampling block. H and W refer to the resolution of feature maps after each layer, C and C' refers to channels of each feature maps, where $C' < C$.

In our experiments, as shown in Table II, the encoder of our architecture could take most proportion of parameters (64.3%) and FLOPs (69.5%). Therefore, utilizing an encoder with a small number of parameters and less computational demands poses a significant overall saving on RAM and CPU/GPU resources. Moreover, the encoder scale can be flexibly reconfigured by simply modifying the hyper-parameter α . We set the parameter $\alpha = 1.0$ through the experiments.

B. Compact Decoder

We propose a new compact decoder structure for use on embedded devices. The decoder learns to recover the resolution from the down-sampled feature maps, and eventually predicts the pixel-wise seagrass segmentation of the input image with same input resolution. As represented by the brown layers in Fig. 1, the decoder comprises five up-sampling blocks. More specifically as shown in Fig. 2, each block has the following structure: at first, one Convolution layer to reduce feature channels, followed Batch Norm and ReLU6 layers, then 2x bi-linear interpolation, and finally a Residual Convolution layer.

The motivation of using this simple symmetrical encoder-decoder architecture is to avoid the complex structure requiring extra computing resources [19, 21]. Comparing to Pyramid Scene Parsing (PSP) [22] and Atrous Spatial Pyramid Pooling (ASPP) [23], we developed a compact structure which has only few channels to reduce FLOPs and parameter number. On the other hand, employing 1x1 Convolution layers at the beginning of up-sampling to reduce the feature map channels (from both last forward and skip-connections) compresses the decoder and further saves FLOPs. The decoder FLOPs account for only 30.5% of the whole network, while the decoder parameters take

35.7%, as shown in Table II in section IV. In comparison, the FLOPs proportion of the decoder in DeepLab [23] with ASPP is 82%. Utilizing this simple structure we can compress the network further without losing accuracy. It helps also to deploy the architecture on embedded devices and achieve the goal of accurate and real-time analysis. Same as the encoder, the network allows to reset the number of channels in each decoder layer to set up the network size. Thus, we use a hyper-parameter β to scale the number of channels in each decoder layer; β is experimentally set to 1.0. The basic channel number in each decoder blocks is shown in Fig. 1.

C. Skipconnection

Skip connections, as shown by red arrows in Fig. 1, enable to directly transmit low-level features extracted from encoder to their corresponding decoder blocks. They help to reconstruct higher resolution prediction from feature maps and mitigate spatial information loss caused by up-sampling. We have also attempted and tested the refinement module in skip connection, such as [24]. However, experimental results showed that the simplest forward structure overperforms such refinements with still largely reduced computational demands.

IV. EXPERIMENT AND RESULT

In this section, we compare the proposed network performance with the four state-of-art segmentation architectures of U-Net [20], SegNet [19], GCN [24] and DeepLab [23]. All the networks were trained on an NVIDIA RTX 2080Ti GPU. Section A presents the large public seagrass database used for the experiments. Section B explains the parameters and details of the network training. Section C compares the accuracy, speed and storage cost of the investigated models. Moreover, we show that the proposed model has good generalization when performing on the new data with a larger distance to the seabed. Section D investigates the impacts of different network parameters (α and β) on the accuracy and computational demands of the proposed network. Section E shows the performance and energy consumption of investigated networks on the embedded device.

A. The Seagrass Dataset

The public seagrass dataset [1] comprises 12682 seagrass images with 1920 x 1080 pixel resolution and is taken from different depths to the seabed. 6037 of them were manually labelled by experts with binary masks of seagrass and non-seagrass classes. Fig. 3 shows the seagrass image samples from

	0m ~ 1m	1m ~ 2m	2m ~ 3m	3m ~ 4m	4m ~ 5m	5m ~ 6m
SAMPLE						
Ground Truth						
Total Images	89	2522	2274	471	332	349

Fig. 3. Samples of seagrass images and their ground truth, taken from different depth to the seabed. Black colour in ground truth indicates the seagrass mask, and white indicates the non-seagrass mask. The first and second row present seagrass samples and their corresponding ground truth from different depth condition, and the last row indicates the total number of images from corresponding depth conditions. Obviously, the light conditions change with increasing depths. The changes of seagrass shapes and outline increase the difficulty of seagrass detection, even for human experts.

the dataset and their corresponding ground truths in different depth groups from the dataset

As in the benchmark work by Reus et al. [1], we split 6037 labelled images into three groups: the training set of 4223 images, the test set of 1204 images and the validation set of 610 images. We used the two-evaluation metrics proposed in [15]: F-Score, which measures the double overlapping proportion of total pixels of image; mean Intersection over Union (mIoU), which evaluates the overlapping proportion of predictions and benchmark masks.

B. Network Training and Implication Details

We applied data augmentation methods to increase data diversity, since it makes the network more robust on varying seagrass data patterns. In particular, we performed random image colour jitter, data random rotation, random horizontal and flip for each training image. That leads to doubling the number of training images. Such data augmentation led to doubling the number of training images. The network takes input images with 256×256 resolution.

We use the ADAM optimizer with initial learning rate 0.001 and we set the batch size to 64. The number of training epochs is set to 100 and other training parameters are left as default. The cross-entropy loss function used to train the network is defined as the equation (3):

$$Loss = -\frac{1}{N} \sum_{i=1}^N y_i \cdot \log(p(\hat{y}_i)) + (1 - y_i) \cdot \log(1 - p(\hat{y}_i)) \quad (3)$$

where y_i is the ground truth class label (1 for seagrass and 0 for non-seagrass) of pixel i , and $p(\hat{y}_i)$ is the predicted class probability of the pixel i , N is the total number of pixels.

C. Comparisons with State of the Art Deep-Learning Models

We compare the performance of two configurations of our proposed segmentation network, default network ($\alpha = 1$ and $\beta = 1$) notated as Ours, and fine-tuned network ($\alpha = 1$ and $\beta = 1.2$) notated as Ours* with U-Net [20], SegNet [19], GCN [24], DeepLab [23] and the benchmark method reported in [1].

As in [1], we separately trained such networks and tested them on the 0m ~ 2m and 0m ~ 6m datasets separately. Table I shows their accuracy comparison in mean IoU and F-Score metrics, while Fig. 4 presents the predictions of seagrass segmentation result. As previously discussed, the 0m ~ 2m trained models were also tested on the 2m ~ 6m dataset, to test if they were robust enough to give the right predictions as well at seafloor distances larger than 2 meters.

We found that all four models achieved a high accuracy level around 91%~93% mIoU and 95%~97% at F-Score for 0~2m images, also 89%~91% mIoU and 94%~96% at F-Score for 0~6m images. This shows that seagrass segmentation is becoming increasingly difficult with larger seabed distances, since at large distances the seagrass images are significantly different and less clear. Our fine-tuned network slightly overperformed other four models, when the model was trained and applied to the same depth group images in 0m ~ 6m depth, with 89.71 mIoU. At the same time, our model achieved the best robustness at different imaging depth. The model trained on 0m - 2m depth images achieved the highest 77% mean IoU, and 87% F-Score on 2m ~ 6m images captured at low visibility and resolution conditions.

Fig. 4. Segmentation results on seagrass public dataset generated by different deep neural networks. All the networks are trained using 0m ~ 6m training images, results presented are from 0m ~ 6m images in the test set.

TABLE I. COMPARISON OF SEGMENTATION NEURAL NETWORK ACCURACY, NUMBER OF PARAMETERS, FLOPS AND SPEED.

Neural Networks	0m ~ 2m training data				0m ~ 6m training data	
	0m ~ 2m test data		2m ~ 6m test data		0m ~ 6m test data	
	mIoU	F-Score	mIoU	F-Score	mIoU	F-Score
Ours	90.70	95.12	77.39	87.25	89.71	94.56
Ours*	91.34	95.46	77.73	87.85	90.66	95.08
U-Net [20]	92.59	96.14	74.72	85.49	90.00	94.74
SegNet [19]	92.03	95.84	76.18	86.60	89.75	94.60
GCN [24]	92.58	96.13	75.10	85.61	90.10	94.80
DeepLab [23]	91.54	95.58	62.47	77.45	89.15	94.30
Benchmark	81.88				80.90	

TABLE II. COMPARISON OF NETWORK COMPUTATIONAL DEMANDS AND SPEED

Networks	Encoder	Parameters of Encoder (M)	Parameters of Network (M)	FLOPs of Encoder (G)	FLOPs of Network (G)	FPS
Ours	MobileNet	2.22 (64.3%)	3.45	0.408 (69.5%)	0.587	200
Ours*	MobileNet	2.22 (59.8%)	3.71	0.408 (64.3%)	0.634	196
SegNet [19]	VGG-16	14.7 (43.2%)	34	20 (30.3%)	66	108
GCN [24]	ResNet-50	23.5 (99.5%)	23.6	5.36 (98.8%)	5.43	140
DeepLab [23]	MobileNet	1.81 (31.1%)	5.8	1.3 (18.1%)	7.28	180
U-Net [20]	U-Net	9.41 (70.2%)	13.4	14.6 (47.3%)	31	171

Anyhow, among all the networks, Ours has the smallest number of parameters (14.6% of GCN) and has the lowest computational demand (just 10.8% FLOPs of the GCN). Also by testing them all on desktop PC with a GeForce RTX 2080Ti GPU (test with image resolution 256x256), the proposed network Ours showed the fastest inference speed at 200FPS (1.42 times faster than the GCN).

Hence, while still achieving competitive accuracy with the state-of-the-art, the proposed segmentation network showed significant advantages on speed, on-board storage and computational cost.

D. Network Configuration Parameters and Their Impact on Network Performance

For a real-time embedded system, the trade-off between network performance (accuracy and inference speed) and network computation demands (memory, energy cost and FLOPs on embedded devices) should be balanced carefully. In this section so, we discuss the effects of the hyper-parameters, α and β , on network accuracy and computational demand, as FLOPs. Such hyper-parameters, α and β are respectively linked with the encoder and decoder sections of the network. By making them smaller, we decrease the number of channels in each layer, and as a result, compress the network, reducing its FLOPs. By adjusting α and β , we can so address the trade-off between network accuracy, performance and FLOPs. In fact, on one side, such compression can negatively affect network accuracy. Yet reducing the required FLOPs will help the network deployment on embedded devices.

To study how network accuracy and FLOPs depend on α and β , we build various networks by modifying one parameter, while keeping the other one fixed, and measure their mIoU accuracy. As from Fig 5, we find that when we expand either the encoder, by varying α from 0.5 to 1.2 with $\beta = 1$, or the decoder, by changing β from 0.5 to 1.2 with $\alpha = 1$, the mIoU increases. Yet by decreasing such hyper-parameters, we can greatly reduce the required FLOPs without losing much on accuracy. For example, by choosing α to 0.6 and β to 1.0, the required FLOPs will be reduced by 64.3% with only 1.68% accuracy loss with respect to higher hyperparameter values. For better accuracy instead, we can choose α to 1 and β to 1.2, reaching so 90.66% mIOU with 0.634 GMACs FLOPs.

E. Speed and Energy Consumption on TX2 Embedded Device

We tested the speed and energy cost of these networks on the AUV embedded platforms. We considered the NVIDIA Jetson TX2 with 256-core NVIDIA Pascal™ GPU and 8GB RAM on the Orbitty carrier board. Either Bluetooth or Wifi were disabled while the monitor was disconnected, only the ethernet link was

Fig. 5. Proposed networks with different configurations. Top figure shows the models with different weighted encoders (when $\beta = 1.0$), and the one below shows the models with different weighted decoders (when $\alpha = 1.0$).

active to access the TX2. We ran all networks on onboard TX2 GPU with a 256x256 image as input. We metered the TX2 energy consumption with two Fluke-115 multimeters, measuring the current and voltage in series and parallel with the TX2 energy input. While each network was running, we measured the TX2 energy consumption very 250 milliseconds for 100 samples and averaged the results. From such average, we subtracted the idle energy of the TX2 which was measured to be 1.98 W. The result was then divided by the measured FPS on TX2 GPU and we obtained E_{INF} , which is the energy necessary to inference one input image.

TABLE III. SPEED OF SEGMENTATION NETWORKS TESTED ON TX2

Network	Ours	SegNet	GCN	DeepLab	U-Net
Times(ms)	56	189	71	60	113
Relative Time Cost	1x	3.4x	1.29x	1.1x	2.0x
FPS	18.0	5.3	13.9	16.7	8.80

TABLE IV. ENERGY CONSUMPTION COMPARISON TESTED ON TX2

Network	Ours	SegNet	GCN	DeepLab	U-Net
$\bar{P}$ (W)	4.72	14.94	9.08	9.40	14.62
σ_P (W)	0.06	0.39	0.18	0.25	0.19
C(J)	0.26	2.82	0.65	0.56	1.66

Table IV presents the comparisons of average energy needs, their standard deviations and EINF for the various networks. Among all, the proposed network has the lowest energy consumption per image, EINF is 0.26 Joule, and it is only 46% of the second-best DeepLab. As indicated by its low standard deviation, it has also the lowest energy fluctuations. For the energy efficiency point of view, the proposed network is so the most adapted for deployment on embedded boards for AUVs use.

V. CONCLUSION

In this study, we proposed a deep learning network for seagrass segmentation, which is accurate, fast, tunable, compact and energy-efficient on embedded electronics. It has mIoU of 90.66% when training on images with 0-6m camera-to-seafloor distance, requires only 0.634 GMACs FLOPs, which is the smallest value among all the networks. For inference speed performance, on a desktop Nvidia 2080ti and embedded Jetson TX2 card, it inferences respectively at 196 and 18 FPS. At the same time, this network is very energy efficient requiring only 0.26J per inference frame on a Jetson TX2. Such energy requirement was just 46% of the ones of the second-best, DeepLab. Hence the demonstrated accuracy, speed and efficiency of the proposed segmentation architecture allow it to be deployed for realtime use on embedded electronics. It will enable the deployment of End2End segmentation tasks on embedded AUVs for live measurements at sea. Hence it will be a step for the autonomous and effective of survey and management of seagrass meadow and for the preservation of their carbon capture and ecological value.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support of the Engineering and Physical Sciences Research Council (EPSRC) Centre for Doctoral Training in Embedded Intelligence under grant reference EP/L014998/1 and Witted Srl for their support and input to this research. The first author would also thank Chinese Scholarship Council (CSC) and Loughborough University for supporting his study.

REFERENCES

- [1] G. Reus, T. Möller, J. Jager, S. T. Schultz, C. Kruschel, J. Hasenauer, K. Neuderth, "Looking for Seagrass: Deep Learning for Visual Coverage Estimation," in *2018 OCEANS - MTS/IEEE Kobe Techno-Oceans (OTO)*, Kobe, 2018, pp. 1-6.
- [2] M. Massot-Campos, G. Oliver-Codina, L. Ruano Amengual and M. Miró-Juliá, "Texture analysis of seabed images: Quantifying the presence of Posidonia oceanica at Palma Bay," in *2013 MTS/IEEE OCEANS - Bergen*, Bergen, 2013, pp. 1-6.
- [3] C.M. Duarte, I.J. Losada, I.E. Hendriks, I. Mazarrasa, and N. Marbà, "The role of coastal plant communities for climate change mitigation and adaptation," *Nature Climate Change*, vol. 3 no.11, 2013, pp. 961-968.
- [4] Short, F.T., "Global Distribution of Seagrasses (Version 4.0). Fourth Update to the Data Layer Used in Green and Short (2003)", 2016
- [5] R.J. Orth, T.J. Carruthers, W.C. Dennison, C.M. Duarte, J.W. Fourqurean, K.L. Heck, A.R. Hughes, G.A. Kendrick, W.J. Kenworthy, S. Olyarnik, and F.T. Short, "A global crisis for seagrass ecosystems," *Bioscience*, vol. 56 no. 12, 2006, pp.987-996.
- [6] C.F. Boudouresque, G. Bernard, G. Pergent, A. Shili, A. and M. Verlaque, "Regression of Mediterranean seagrasses caused by natural processes and anthropogenic disturbances and stress: a critical review," *Botanica Marina*, vol. 52 no. 5, 2009, pp. 395-418.
- [7] Weidmann, F., Jäger, J., Reus, G., Schultz, S.T., Kruschel, C., Wolff, V. and Fricke-Neuderth, K., "A Closer Look at Seagrass Meadows: Semantic Segmentation for Visual Coverage Estimation," in *OCEANS 2019 - Marseille, Marseille, France*, 2019, pp. 1-6.
- [8] D. Moreno, P.A. Aguilera, and H. Castro, "Assessment of the conservation status of seagrass (Posidonia oceanica) meadows: implications for monitoring strategy and the decision-making process," *Biological Conservation* vol. 102 no. 3, 2001, pp. 325-332.
- [9] A.D. Irving, J. E. Tanner, and S. G. Gaylard, "An integrative method for the evaluation, monitoring, and comparison of seagrass habitat structure", *Marine Pollution Bulletin*, vol. 66 no. 1-2, 2013, pp. 176-184.
- [10] A. Burguera, F. Bonin-Font, J. L. Lisani, A. B. Petro and G. Oliver, "Towards automatic visual sea grass detection in underwater areas of ecological interest," in *IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA)*, Berlin, 2016, pp. 1-4.
- [11] F. Bonin-Font, A. Burguera and J. Lisani, "Visual Discrimination and Large Area Mapping of Posidonia Oceanica Using a Lightweight AUV," *IEEE Access*, vol. 5, 2017, pp. 24479-24494
- [12] Y. Gonzalez-Cid, A. Burguera, F. Bonin-Font and A. Matamoros, "Machine learning and deep learning strategies to identify Posidonia meadows in underwater images," in *OCEANS 2017 - Aberdeen, Aberdeen*, 2017, pp. 1-5.
- [13] F. Bonin-Font, M.M. Campos, and G.O. Codina, "Towards Visual Detection, Mapping and Towards Visual Detection, Mapping and Quantification of Posidonia Oceanica using a Quantification of Posidonia Oceanica using a Quantification of Posidonia Oceanica using a Lightweight AUV," *IFAC-PapersOnLine*, vol. 49, no. 23, 2016, pp. 500-505.
- [14] M. Martin-Abadal, E. Guerrero-Font, F. Bonin-Font and Y. Gonzalez-Cid. (2018). "Deep Semantic Segmentation in an AUV for Online Posidonia Oceanica Meadows Identification," *IEEE Access*, vol. 6, pp. 60956-60967, 2018.
- [15] J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, 2015, pp. 3431-3440.
- [16] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, 2018, pp. 4510-4520.
- [17] J. Deng, W. Dong, R. Socher, L. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, 2009, pp. 248-255.
- [18] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 2016, pp. 770-778.
- [19] V. Badrinarayanan, A. Kendall and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 39, vol. 12, 2017, pp. 2481-2495.
- [20] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-assisted Intervention*, 2015, pp. 234-241.
- [21] H. Noh, S. Hong and B. Han, "Learning Deconvolution Network for Semantic Segmentation," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Santiago, 2015, pp. 1520-1528.
- [22] H. Zhao, J. Shi, X. Qi, X. Wang and J. Jia, "Pyramid Scene Parsing Network," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 2017, pp. 6230-6239.
- [23] L. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 801-818.
- [24] C. Peng, X. Zhang, G. Yu, G. Luo and J. Sun, "Large Kernel Matters — Improve Semantic Segmentation by Global Convolutional Network" in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 2017, pp. 1743-1751.