Panoptic Segmentation of Environmental UAV Images : Litter Beach

Ousmane Youme, Jean Marie Dembele, Eugene C. Ezin and Christophe Cambier

July 31, 2024

Panoptic Segmentation of Environmental UAV Images : Litter Beach

Ousmane YOUME[1], Jean Marie Dembele[1,2], Eugene C. Ezin[3], and Christophe Cambier[2]

[1] Gaston Berger University, Saint-Louis, Senegal youme.ousmane1@ugb.edu.sn
[2] UMI 209 UMMISCO, Sorbonne University IRD
[3] IMSP & IFRI, University of Abomey-Calavi, Abomey-Calavi, Benin

Abstract. Convolutional neural networks (CNN) have been used efficiently in several fields, including environmental research. In fact, CNN can help with monitoring of marine litter, which has become a worldwide problem. UAVs have a higher resolution and more adaptable in local area than satellite imagery, which makes it easier to find and count trash. Since the sand is heterogeneous, a simple CNN model encounters plenty of inferences caused by reflections of sand color, human footsteps, shadows, algae present, dunes, holes, and tire tracks. For this type of image, segmentation methods based on CNN can be more appropriate. In this paper, we use an instance-based segmentation method and a panoptic segmentation method that show good accuracy with just a few datasets. The model is more robust and less sensitive to disturbances, especially in panoptic segmentation.

Keywords: Convolutional Neural Network (CNN) · Panoptic Segmentation,· Unmanned aerial Vehicle (UAV) · Environment · Litter Beach.

1 Introduction

Marine pollution is a phenomenon that is growing every year at an unprecedented rate. Each year, more than 10 million tons of waste are thrown onto beaches and oceans, 80% of which come from land and 20% from maritime activities. They are composed of 80% plastic, and their destruction constitutes a problem because the degradation of the polymers leads to micro-particles after fragmentation. [1] Waste on the coast, in addition to being a major ecological have a negative impact on tourism, the economy, and the health of ecosystem species. Therefore, it is essential to find techniques to automatically detect and list the waste on the coast to make treatment easier.[2]

Traditional counting methods have been suggested, like the OSPAR 2010 Guidelines [3], which call for a 100-meter sweep of the beach to collect trash larger than 2.5 cm. Then, it would take 2–5 collectors 3–5 hours to sort the trash and measure the quantity. With the progress of technology, it has recently been shown [4] that aerial images and machine learning methods can be used to monitor and find waste. Some work has been done in this sense: Cecilia et al.

[4] use machine learning to automatically find trash in the ocean and describe it. The oriented gradient histogram, which is a classic way to describe and show features, is used to describe the data first. Then, random forest builds a number of decision trees in the training dataset and uses the results of the whole set of trees to make predictions about the test dataset. Three (3) random forests are built: two for binary true or false predictions and the third for predicting the species of true boxes: plastic, capsule, etc. Gonzalez et al. [5] also use random forest classification to find marine litter. The input images are changed upstream by adding three types of color chains to the RGB: the HSV, the CIE-Lab, and the YCbCr. This creates an input vector with 12 features. Marine litter is detected with an F-Score accuracy of 76% on the beach and a F-Score of 55% on the dunes. At the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), deep learning methods were the best. Since then, these methods have been used in many different fields, such as medicine, the environment, etc., especially in remote sensing to set up monitoring systems where they have been used at all levels. UAV (unmanned aerial vehicle) systems facilitated data acquisition. However, simple CNN models don't work well for task detection in high-resolution images of the environment taken by satellite or UAV [6]. These models show a lot of inferences in a dense area. The task is hard because the color of the sand reflects off of objects, footsteps, shadows, algae, and other things that make it hard to find litter. In this paper, we show that segmentation methods work better for this kind of image of the environment. Two models were chosen for experimentation and have been compared in this work. One method is based on instance segmentation, and an other is based on panoptic segmentation. Both are used on a UAV-collected dataset of trash on the coast.

2 Related works

In the S.O.T.A, few works are noted in the application of CNN for the detection of coastal wastes. Fallati et al. [7] employ deep learning, specifically CNN. They use commercial plastic detection software to generate numerical scores on two different sites (44% and 78% respectively). Papakonstantinou et al [8] proposed a UAS data acquisition and annotation protocol combined with deep learning techniques for automatic detection and mapping of ML concentrations in the coastal zone. Five convolutional neural networks (CNNs) were trained to classify UAS image tiles into two classes: (a) litter and (b) no litter. Testing the generalization ability of convolutional neural networks on an unseen dataset, they found that the VGG19 [9] network in an amount of more than 15.000 samples yielded an overall accuracy of 77.6% and an f-score of 77.42%. Wolf et al. 2021 [10] show a simple convolutional neural network (CNN) that can find and count floating and shoreline-dumped plastic litter. The Aquatic Plastic Waste Detection, Classification, and Quantification System (APLASTIC-Q) was made and trained using very high geospatial resolution images taken during aerial surveys in Cambodia a local environment that is not complex and varied. The machine learning parts of APLASTIC-Q are the plastic waste detector (PLD-CNN) and the plastic waste quantifier (PLQ-CNN). The PLD-CNN successfully classified in the targets into water, sand, vegetation, and plastic waste with an accuracy of 83%. . YH Liao 2022 [11] proposes a marine litter detection system that uses drones. The goal is to use drones instead of people to find trash in the ocean and give government agencies information about pollution in real time. This study used the Internet and computer-to-computer communication. For real-time calculations, a marine litter detection system was put on the onboard computer of a drone. Images of marine litter were provided to train a modified YOLO [12] model (You Look Only Once). It was found that the drone could fly along a path that had already been set up and find trash in coastal areas. The detection results were sent to a data streaming platform for processing and analysis. However, with an accuracy of around of 70% at the most, we cannot confirm the adaptation of model cited in this section to dense areas. Simple CNN architecture can have a high number of false positives caused by sand dunes, shadows, footsteps, algae, and other things that leave traces that mess with the model [7] . The purpose of this paper is to find a more robust and sensitive model adapted for this type of area.

3 Background

In the last few decades, deep learning methods have reached great success in the field of computer vision. They have been used successfully in autonomous cars, robots, disease diagnosis, and other areas. Contemporary methods such as large-scale VGG19 [9], and Xception [13] have been proposed for efficient object detection in videos or images taken on land. But because there are so many of them, it's not always easy to find a good model for a problem when there aren't any ready-made solutions. Also, these methods don't seem to work well with high resolution images from drone or satellite. For this kind of data, segmentation methods are more commonly used. In the next section, we present the segmentation methods.

3.1 Segmentation Methods

In the past few years, segmentation methods have shown how well they work at classifying aerial image data. We can classify them in three parts: semantic segmentation, instance segmentation and panoptic segmentation. Segmentation methods did well in the 2018 deep glob challenge, which regrouped tasks like road extraction, building detection, and land use classification. A variation of Deeplab [14] obtained the best accuracy in the extraction of roads. Of the 22 proposals in this challenge, 14 are segmentation models, 13 of which are U-Net-based models , an U-Sharpened semantic segmentation method with encoder and decoder [15]. On Space-net: A Remote Sensing Dataset and Challenge Series, 2 of the 3 challenge series were won by U-Net-based segmentation methods.

Semantics Segmentation Semantic segmentation is a pixel-level classification of the image. An embedding of the boundary changes between objects of different colors is made. The segmentation method has been used for a long time for image recognition with classical machine learning methods. However, its major development starts with the proposal of the FCN fully convolutional neural network [16]. After a sequence of convolution and reduction, the FCN turns the input image into an output of the same size. The output is then given an up-convolution, which sorts each pixel according to a label. The Unet[15] based on FCN is a basic segmentation method composed of two parallel paths. The first path reduces dimension, including conv(3x3) and max pooling(2x2). With pixel classification, the image dimensions are put back together by following an extended path called up-convolution. HRNET[17], inspired by Unet[15], is made up of four parallel branches, each with a different reduction dimension. ASP[18] uses HRNET[17] and uses two first branches of adaptive spatial pooling to increase the gain in the extraction feature at different levels by using adaptive spatial pooling. There are models based on a spatial pyramid pooling such as PSPNET[19], DeepLab[14] etc.

Instance Segmentation Instance segmentation is a computer vision technique that has become popular in recent years. Differently from semantic segmentation, instance-based segmentation, in addition to classification at the pixel level, adds object detection techniques. This combination allows for more robustness on object detection tasks with a distinction between instances of the same label. A unique instance ID is assigned to each object instance. This enables simple object quantity counting. We distinguish two(2) main methodologies: the bottom-up technique, in which segmentation is used first and then object detection is applied to the proposed output (Deeperlab [20]); the top-down method, in which both techniques are used in parallel and the results of these two prediction branches are merged. Top-Down is divided into two methods: two-stage methods that integrate the region proposal with parallel mask branches i.e MaskRCNN[21]; EfficientNet [22], etc. and one-stage methods that eliminate the region proposal, i.e Polarmask [23]).

Panoptics Segmentation Panoptics segmentation combines segmentation by semantics and instance segmentation techniques. The panoptic identifies the expected output objects and fuses them with mask branch pixel label classification. The rest of the image, including the background, is then separated at the pixel level using semantic segmentation techniques. Most panoptic segmentation models add parallel paths for semantic segmentation to existing per-instance segmentation models. As a result, they include bottom-up instance segmentation models like Panoptic DeepLab [24], two-stage methods like EfficientPS [25], one-stage methods like FPSNet[26] but also some single-path methods.

4 Methodologies

4.1 Area of Study

The study area is located on the coast of Dakar, Senegal, in West Africa. Senegal is a country that registers a lot of tourists and must be favorable for the protection of the environment. On the other hand, some of its places have a lot of waste, especially on the beaches along the coastline. Tavares et al [27] in a study of two locations, Dakar and Mbour, found an average density of debris at least $10cm$ deep of 48.75 debris per m2 and 1.95 debris per m^2 on the surface, 95% of which is plastic. As a result, the establishment of a monitoring system is critical in order to have information on the location and density of waste and thus contain in order to effectively eliminate marine pollution. The place selected is located north of Dakar on the coastline from Dakar to the region of Saint-Louis at an altitude of 14 ° 48' north and 17 ° 18 longitude. The UAV flew at altitudes of 5, 10, and 30 meters over a distance of 300 meters at a 90-degree angle. In view of the images taken, the altitude of 10 meters is more satisfactory in terms of coverage, image resolution, and visibility of debris. For scale variation, we will therefore use the images captured at 10 and $30m$ to train our model.

4.2 Data Acquisition

For the acquisition of necessary data, we use the DJI Mavic PRO drone, which has a high resolution with its L1D-20c RGB color camera, taking images of 5472×3648 pixels. The high resolution is essential for the performance of our model. In addition, UAV is easily manipulated and contains metadata such as shutter speed, aperture, ISO and GPS coordinates (latitude, longitude, and altitude) that can be read with the exiftool software. Given the intensity of the light, which exacerbates the problem with shadows and the whiteness of the sand, each image is subjected to an RGB color attenuation algorithm. The curve of the histogram with the parameters of contrast and brightness on each image is taken, and then they are automatically optimized by defining the optimal alpha (contrast) and beta (brightness), knowing that:

$$f(i,j) = \alpha \times g(i,j) + \beta, \tag{1}$$

i and j are the pixel coordinates, f is the new image, and g is the old. When the frequency of colors is less than a certain threshold value (e.g. 1), the cumulative distribution is computed, the right and left ends of the histogram are cut off. This gives the minimum and maximum ranges. Here is a visualization of the histogram before (orange) and after the clipping (blue). Notice that the most "interesting" sections of the image are more pronounced after cropping.

To calculate the alpha, we take the minimum and maximum gray ranges after clipping and divide them by the desired output range of 255.

$$\alpha = 255/(maximumgray - minimumgray). \tag{2}$$

To calculate the beta, we plug it into the formula where

$$g(i, j) = 0, \tag{3}$$

$$f(i, j) = minimumgray, \tag{4}$$

$$g(i, j) = \alpha \times f(i, j) + \beta. \tag{5}$$

which, after being solved, results in:

$$\beta = -minimumgray \times \alpha \tag{6}$$

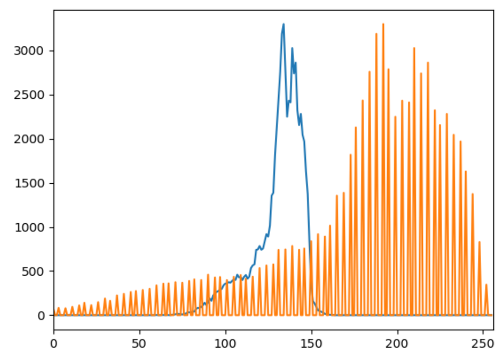

Fig. 1. Difference of variation in images before (orange) and after processing (blue).

This equation gives alpha and beta to apply to each image. Waste on the images obtained after transformation are more visible and clearly differentiate the background that is the sand. The transformed images are divided into a grid of images of size 600×600 adapted to the inputs of the chosen algorithm. This correspond for each image transformed to 72 image-inputs. A final database is obtained for labelling in order to training model.

5 Algorithms Selected

Segmentation architectures can have many different parts, such as anchor bases, bounding boxes, and no-maximum suppression. There are four types of processes that use these mechanisms: top-down (one stage and two stages), bottom-up, single-path, and merged methods. The segmentation by instance and panoptic model for experimentation are chosen based on their specificity. They rank well based on their performance on the coco Val dataset, which is closest to our local contexts. We recall that the demonstration is more related to the type of segmentation regardless of the model than the architecture itself. MaskRCNN[22] is a classical object detection algorithm with instance-based segmentation. Maskrcnn[22] extends FastRCNN[28] with two parallel branches: the first one for classification or regression of bounding boxes and the second one for prediction of masks applied on each ROI. Panoptic-DeepLab[25] simply employs a central keypoint for class-agnostic instance detection, in contrast to DeeperLab [20] . As a result, Panoptic-DeepLab predicts three outputs: (1) semantic segmentation, (2) instance center heatmap, and (3) instance center regression. First, class-agnostic instance segmentation is made by putting the predicted foreground pixels together with the predicted instance centers that are closest to them. When the class-agnostic instance segmentation and the semantic segmentation are put together, the final panoptic segmentation is made.

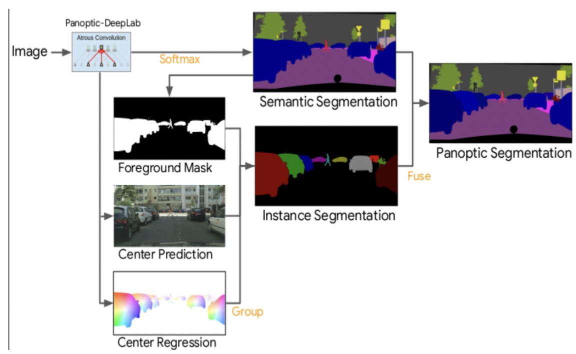

Fig. 2. Panoptic Deeplab architecture images from original paper.[24]

The loss function is calculate as follow:

$$L = \lambda_1(L_c + L_b + L_m) + \lambda_s \times L_s \tag{7}$$

L_c is a softmax cross-entropy classification loss function, L_b is a regression loss smooth, and L_1 is the mask loss function. $lambda_1$ and $lambda_s$ are set to zero or one depending on the training stage, either instance or semantic.

6 Experimentation

6.1 Litter Dataset

The litter dataset is made up of 1500 images of 600 x 600 pixels after data collection by UAV and image preprocessing. For the labels "Litter" and "Algae," a total of 10 worth of objects is calculated per class. Some images do not contain any objects for the model to consider this case. Images are labeled manually with VGG Image Annotator. Finally, a JSON file containing polygons representing the positions of objects in the image is exported. We generate PNG semantic images for the panoptic model. Litter Dataset is split in three (3) dataset: training set for 55% percent, test set 35% and validation set 10%. Metrics of models are calculate in validation set. As metrics, Average Precision, PS (precision in small objects), PL (precision in large objects), and AR (average recall) are retained.

6.2 Parameters Initialization

For parameter initialization, all models share the same optimization parameters. The learning rate is set to 0.0001, and the batch size is set to 2 images per GPU. ResNet 50 trained on the Coco dataset is used for checkpoint initialization. A global step of 1000 is set for the whole training loop.

7 Experimental Result

In this section, we show the results obtained from selected models on the benchmark dataset used. First, we show their quantities and performances, especially in terms of computational resources, average precision, and average recall for MaskRCNN[21] and panoptic quality, precision on small and large targets, and recall quality for sensitivity. Second, we assess the quality and efficiency of selected models. The table 1 is divided into two sections: first, the metrics for the Mask-RCNN model, and then, in the final section, the metrics for the Panoptic DeepLab[24] model.

7.1 Quantitative Comparison

Segmentation methods are preferred in the field of high-resolution recognition due to their mechanisms, including the pixel label classification. MaskRCNN

Table 1. Table of experimental result

	Metrics[1]					
Modele	GPU x Day	NP	AP	APS	APL	AR
MaskRCNN[21]	1.4	20.04 M	35.6	-	-	28.05
Modele	GPU x Day	NP	PQ	PS	PL	RQ
Panoptic Deep Lab[24]	0.5	25.6 M	38.5	37.8	39.1	40, 8

Note: Results are obtained by training in this size of Dataset and does not integrate new data collected and unlabelled

[a] Panoptic Quality PQ is metrics correspond to Average Precision AP for Panoptic Segmentation [24]

trained in more than 20.04 Millions of parameter take more than 1 day for end execution with Average Precision of 35.6% and a little less on Average Recall. Panoptic DeepLab model have +3% more Precision in Panoptic Quality with 25.6 Millions parameters training in few hours. More or less same value in small and large objects. We note an increase of +10% in Average Recall. An analysis of this information is made based on the nature of the models and the environment studied.

7.2 Qualitative Comparison

Instance and panoptic segmentation methods, which include semantic segmentation, have good accuracy on our Litter Beach Dataset with a small amount of data. Inference shows an accurate detection of waste down to a scale of a few centimeters (see Figure 4). The parallel application of the segmentation mask reduces false positives produced by heterogeneity on the coastline. Moreover, Panoptic Deeplab has a higher sensitivity due to the combination of semantic segmentation, which first segments the image by dissociating the background as well as the foreground by recognizing the stuff here as "sand". With this mechanism, object detection is more visible, and this reduces the rate of false negatives. We can see a positive evolution of the semantic loss function as well as a regression with a low rate following center loss on the following curve.

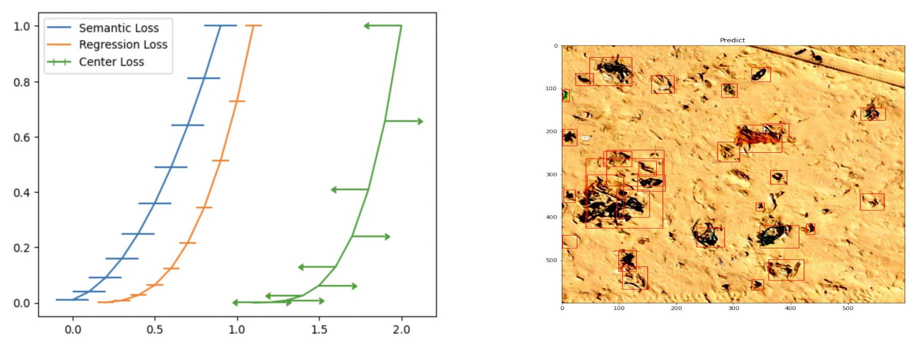

Fig. 3. A) Decreasing of function loss during training step of Panoptic DeepLab, B)Example Prediction of Litter and Algue with MaskRCNN in one images

8 Further Discussions

The use of UAV is an important aspect for saving time and having considerable coverage compared to traditional techniques, which are more tedious. The altitude of 10 meters seems to us more suitable in terms of flight time, coverage, image quality, and waste visibility, as attested by Martins et al.[4] and Fallati et al.[7] but nevertheless, we took this into account for the training of our model images taken at 30 meters. The division of the images into several images of adapted sizes to the inputs of our algorithm allows for better visibility of the waste with a zoom effect. To solve the problem, machine learning techniques have been proposed, such as the random forest in [4] and [5] with F-scores varying between 40% and 78%. It should be noted that in the field of image detection, CNNs are more commonly used than classic methods of machine learning. However, the use of deep learning methods varies depending on the ecosystem studied. Some related work 2 use a customized CNN-based model. We opted for the detection of waste on the coast by segmentation methods, including panoptic, for more reliability. The aims of this paper are to first show that the segmentation method is more adaptable in environmental aerial images such as litter beaches, then demonstrate the robustness of the model when applying panoptic segmentation.

9 Conclusion

In this work, we showed that it is possible to find marine waste to protect the environment. We have demonstrated that Deep Learning, in particular segmentation methods like mask r-cnn, outperforms other classical machine learning methods. This work shows that marine litter detection using Deep learning techniques and UAV drone tool can significantly fight against marine pollution and environmental degradation. So, it is easier for groups that work to protect the ocean to set up an effective system for autonomous monitoring. In other research, it is critical to determine the density of waste elements on the coastline, but it is even more important to have a large scale coverage in accordance with an optimal calculation time. It is also possible to compare the evolution of the states taken through the years.

Acknowledgment

This publication was made possible through the DSTN supported by IRD and AFD. We would like to thank the African Center of Excellence in Mathematical, Informatics, and Tics (CEA-MITIC) and the African Centre of Excellence in Mathematical Science Informatics and their Application (CEA-SMIA) for their support.

Bibliography

[1] Ministères Écologie Énergie Territoires. "Déchets marins." Accessed January 15, 2023. https://www.ecologie.gouv.fr/dechets-marins.

[2] Velander, Kathy, and Marina Mocogni. 1999. "Beach Litter Sampling Strategies: Is There a 'Best'Method?" Marine Pollution Bulletin 38 (12): 1134–40.

[3] Wenneker, Barbara, and Lex Oosterbaan. 2010. "Guideline for Monitoring Marine Litter on the Beaches in the OSPAR Maritime Area. Edition 1.0." Report. OSPAR Commission. https://doi.org/10.25607/OBP-968.

[4] Martin, Cecilia, Stephen Parkes, Qiannan Zhang, Xiangliang Zhang, Matthew F. McCabe, and Carlos M. Duarte. "Use of Unmanned Aerial Vehicles for Efficient Beach Litter Monitoring." Marine Pollution Bulletin 131 (2018): 662–73.

[5] Gonçalves, Gil, Umberto Andriolo, Luís Pinto, and Filipa Bessa. "Mapping Marine Litter Using UAS on a Beach-Dune System: A Multidisciplinary Approach." Science of the Total Environment 706 (2020): 135742.

[6] Youme, Ousmane, Theophile Bayet, Jean Marie Dembele, and Christophe Cambier. "Deep Learning and Remote Sensing: Detection of Dumping Waste Using UAV." Procedia Computer Science, Big Data, IoT, and AI for a Smarter Future, 185 (January 1, 2021): 361–69. https://doi.org/10.1016/j.procs.2021.05.037.

[7] Fallati, Luca, Annalisa Polidori, Christian Salvatore, Luca Saponari, Alessandra Savini, and P. Galli. "Anthropogenic Marine Debris Assessment with Unmanned Aerial Vehicle Imagery and Deep Learning: A Case Study along the Beaches of the Republic of Maldives." Science of The Total Environment 693 (2019): 133581. Geus, Daan de, Panagiotis Meletis, and Gijs Dubbelman. "Fast Panoptic Segmentation Network." arXiv, October 9, 2019. https://doi.org/10.48550/arXiv.1910.03892.

[8] Papakonstantinou, Apostolos, Marios Batsaris, Spyros Spondylidis, and Konstantinos Topouzelis. "A Citizen Science Unmanned Aerial System Data Acquisition Protocol and Deep Learning Techniques for the Automatic Detection and Mapping of Marine Litter Concentrations in the Coastal Zone." Drones 5, no. 1 (2021): 6.

[9] Simonyan, Karen, and Andrew Zisserman. 2015. "Very Deep Convolutional Networks for Large-Scale Image Recognition." arXiv. https://doi.org/10.48550/arXiv.1409.1556.

[10] Wolf, Mattis, Katelijn van den Berg, Shungudzemwoyo P. Garaba, Nina Gnann, Klaus Sattler, Frederic Stahl, and Oliver Zielinski. "Machine Learning for Aquatic Plastic Litter Detection, Classification and Quantification (APLASTIC-Q)." Environmental Research Letters 15, no. 11 (2020): 114042.

[11] Liao, Yu-Hsien, and Jih-Gau Juang. "Real-Time UAV Trash Monitoring System." Applied Sciences 12, no. 4 (January 2022): 1838. https://doi.org/10.3390/app12041838

[12] Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. "You Only Look Once: Unified, Real-Time Object Detection." arXiv. https://doi.org/10.48550/arXiv.1506.02640.

[13] Chollet, François. "Xception: Deep Learning with Depthwise Separable Convolutions." arXiv, April 4, 2017. https://doi.org/10.48550/arXiv.1610.02357.

[14] Chen, Liang-Chieh, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs." arXiv, May 11, 2017.

[15] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. 2015. "U-Net: Convolutional Networks for Biomedical Image Segmentation." arXiv. https://doi.org/10.48550/arXiv.1505.04597.

[16] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully Convolutional Networks for Semantic Segmentation." arXiv, March 8, 2015. https://doi.org/10.48550/arXiv.1411.4038.

[17] Wang, Jingdong, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, et al. "Deep High-Resolution Representation Learning for Visual Recognition." arXiv, March 13, 2020. https://doi.org/10.48550/arXiv.1908.07919.

[18] Liu, Yinglu, Yan-Ming Zhang, Xu-Yao Zhang, and Cheng-Lin Liu. "Adaptive Spatial Pooling for Image Classification." Pattern Recognition 55 (July 1, 2016): 58–67. https://doi.org/10.1016/j.patcog.2016.01.030.

[19] Zhao, Hengshuang, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. 2017. "Pyramid Scene Parsing Network." arXiv. https://doi.org/10.48550/arXiv.1612.01105.

[20] Yang, Tien-Ju, Maxwell D. Collins, Yukun Zhu, Jyh-Jing Hwang, Ting Liu, Xiao Zhang, Vivienne Sze, George Papandreou, and Liang-Chieh Chen. "DeeperLab: Single-Shot Image Parser." arXiv, March 12, 2019. https://doi.org/10.48550/arXiv.1902.05093.

[21] He, Kaiming, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. "Mask R-CNN." arXiv, January 24, 2018. https://doi.org/10.48550/arXiv.1703.06870.

[22] Tan, Mingxing, and Quoc V. Le. "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks." arXiv, September 11, 2020. https://doi.org/10.48550/arXiv.1905.11946.

[23] Xie, Enze, Peize Sun, Xiaoge Song, Wenhai Wang, Ding Liang, Chunhua Shen, and Ping Luo. "PolarMask: Single Shot Instance Segmentation with Polar Representation." arXiv, February 25, 2020. https://doi.org/10.48550/arXiv.1909.13226.

[24] Cheng, Bowen, Maxwell D. Collins, Yukun Zhu, Ting Liu, Thomas S. Huang, Hartwig Adam, and Liang-Chieh Chen. "Panoptic-DeepLab: A Simple, Strong, and Fast Baseline for Bottom-Up Panoptic Segmentation." arXiv, March 11, 2020. https://doi.org/10.48550/arXiv.1911.10194.

[25] Mohan, Rohit, and Abhinav Valada. "EfficientPS: Efficient Panoptic Segmentation." arXiv, February 1, 2021. https://doi.org/10.48550/arXiv.2004.02307.

[26] Geus, Daan de, Panagiotis Meletis, and Gijs Dubbelman. 2019. “Fast Panoptic Segmentation Network.” arXiv. https://doi.org/10.48550/arXiv.1910.03892.

[27] Tavares, Davi Castro, Jailson Fulgêncio Moura, Adam Ceesay, and Agostino Merico. “Density and Composition of Surface and Buried Plastic Debris in Beaches of Senegal.” Science of The Total Environment 737 (October 1, 2020): 139633. https://doi.org/10.1016/j.scitotenv.2020.139633.

[28] Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.” arXiv, January 6, 2016. https://doi.org/10.48550/arXiv.1506.01497.