

# Beach Surveillance: A Contribution to Automation

Maria da Conceição Proença<sup>1,2</sup>, Ricardo Nogueira Mendes<sup>3</sup>

<sup>1</sup>Marine and Environmental Sciences Centre & ARNET, Aquatic Research Infrastructure Network Associated Laboratory, Lisbon, Portugal

<sup>2</sup>Department of Physics, Faculty of Sciences, University of Lisbon, Lisbon, Portugal

<sup>3</sup>Institute for Nature Conservation and Forests, Algés, Portugal

Email: mcproenca@fc.ul.pt

**How to cite this paper:** Proença, M. C., & Mendes, R. N. (2024). Beach Surveillance: A Contribution to Automation. *Journal of Geoscience and Environment Protection*, 12, 155-163.

<https://doi.org/10.4236/gep.2024.1212010>

**Received:** November 18, 2024

**Accepted:** December 22, 2024

**Published:** December 25, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

The problem of human overload in many habitats is becoming increasingly urgent, as it is the driving force that destroys ecosystems beyond repair. This paper describes a possible workflow for beach surveillance, using a deep learning solution available online that runs on a standard laptop with RGB images acquired with a standard camera. The software is YOLO v7, a state-of-the-art real-time object detection model presently used for autonomous driving, surveillance, and robotics. The workflow and parametrization needed for building a model are described, along with examples of the results over 180 test images that ensures an overall precision of 0.98 and recall of 0.94 (F1 = 0.96). The model was parametrized to focus on a minimum number of false positives; from the 5672 possible detections identified by human curation, 5285 were correctly identified and located, 387 missed and there are 116 mistakes. A minimum of computational skills is needed to reproduce this implementation in any user data of the same kind.

## Keywords

People Counting, Beach Surcharge, Human Detectors, Deep Learning Methodologies

## 1. Introduction

Controlling the anthropogenic pressure in beach areas near the shore is important for several reasons, including preserving natural habitats: coastal ecosystems are delicate and complex, and they are often threatened by human activity such as beach development and recreational activities (Luna da Silva et al., 2017).

Controlling human presence can help preserve the natural habitats and biodiversity of these areas, as well as protecting wildlife: many species of marine animals, such as shorebirds, use beaches as nesting and breeding grounds. Human presence can disturb these animals and interfere with their natural behaviors. By controlling human presence, we can minimize the impact on vulnerable species.

Human activity on beaches can lead to pollution, including litter (Corraini et al., 2018) and chemical pollution from sunscreen and other products, like plastics (Thushari and Senevirathna, 2020) used and abandoned. Education and less anthropogenic pressure can reduce the amount of pollution entering the marine environment near the shoreline.

Beaches can be dangerous places in at least two ways: in the water, waves and currents can be strong enough to overcome a good swimmer and, on the opposite side, on land, unstable cliffs can cause fatal landslides. Controlling human presence can help ensure public safety and prevent irreversible situations by keeping human pressure at a level that the locally available rescue services can respond to.

It can be challenging to accurately estimate the number of people visiting public beaches, especially during peak times or busy seasons (Green et al., 2005). This can be due to several factors, such as the extent of beach areas, the variability in visitor patterns, and the difficulty in counting people in such a dynamic environment. However, there are several methods (Morgan, 2018) that can be used to estimate the number of visitors to public beaches, such as:

1) Manual counts: This involves physically counting the number of people entering and leaving the beach area at specific times or intervals. While this method can be time-consuming and labor-intensive, it can provide an accurate estimate of the number of visitors.

2) Automated counters: These are devices that use sensors or cameras to count the number of people entering and leaving the beach area. This method can be more efficient than manual counts, but it may not be as accurate in crowded or congested areas.

3) Surveys: Visitors can be asked to fill out surveys that collect information about their visit, such as the date and time they arrived, the duration of their stay, and the activities they engaged in. This method can provide valuable insights into visitor behavior and preferences, but it may not capture the entire population of visitors.

4) Aerial imagery, like UAV or drone acquired images (Turner et al., 2016; Adade et al., 2021), can provide a bird's-eye view of the beach extend and can be used to estimate the number of people based on the density of crowds, if the ground resolution is not enough to discriminate individuals.

While estimating the number of visitors to public beaches can be challenging, using a combination of these methods can help provide a more accurate estimate of human presence and inform management decisions to protect the environment, ensure public safety, and enhance visitor experiences, evolving to an ecosystem-based management system (Sardá et al., 2015), crucial for both cases—

beaches that are still semi-wild and beaches in urban areas (Cabioch and Robert, 2022).

In the case of beach management (Domingo, 2021), deep learning methods such as You Only Look Once (YOLO) could be used to detect and track people on the beach using images or video footage from surveillance cameras or drones. This would enable authorities to monitor beach activity, estimate visitor numbers, and identify potential safety risks or environmental concerns. The YOLO series (Redmon et al., 2016) are popular object detection algorithms available online that uses convolutional neural networks (CNNs) to identify and track objects in real or near real-time, with a minimum of requirements in terms of operator skills and hardware—any version can work in a regular laptop. YOLO can also be trained to recognize specific behaviors or activities, such as swimming or sunbathing, which can provide additional insights into visitor behavior patterns. The algorithm can be customized to detect and track specific objects or people of interest, such as lifeguards or emergency responders, to ensure fast communication and more effective coverage in terms of assistance.

If beach surveillance data is shared and used in conjunction with road authorities, it can help avoid human surcharge in more critical areas; for example, if beach surveillance data shows that visitor numbers are exceeding the capacity of a particular beach area, road authorities could implement traffic control measures or divert visitors to less crowded areas. Authorities can use the data to identify potential safety risks or environmental concerns and take proactive measures to address them.

By sharing data on visitor numbers, traffic flows, and visitor behavior patterns, authorities can work together to manage visitor flows and prevent the impact of excessive human activity on sensitive ecosystems.

Overall, sharing beach surveillance data with road authorities can help promote sustainable tourism and protect delicate ecosystems from the negative impacts of human activity. It requires collaboration and coordination between multiple stakeholders, but the benefits of such efforts can be significant for both the environment and visitors.

This paper intends to show the possibility of controlling the beach population, detecting and counting people with RGB images acquired from a high point of view with a regular surveillance camera, using a YOLO v7 model to process the images. What's new is the use of high-performance online deep learning software, which is not fooled by differences in the background (water, sand, rocks, vegetation) because it focuses on the objects used to train it.

## 2. Materials and Methods

A collection of images of a surveillance camera acquired in August, 2013 at Portinho da Arrábida zone, an area with several beaches in southwest Portugal, were used. The camera is a Nikon D80, programmed to aperture priority, and automatic white balance, with a lens of 300 mm focal length. These are large RGB

images with 3 bands, 2896 by 1944 pixels, that were pre-processed in a Matlab environment (Matlab, n.d.) to tiles of a convenient size ( $960 \times 960$ ) considering the processing capabilities of the hardware available.

The methodology followed is based on the use of a recent version of YOLO deep-learning algorithms, YOLO v7 (July 2022), available online in a GitHub repository (GitHub, n.d.).

The conceptual framework of this methodology follows the steps of supervised classification: the objects of interest (OoI) must be previously identified in a representative subset of images; these annotated images are divided into training and validation datasets and will be the basis for building the model using CNNs to define discriminative features. It takes several iterations before the training data produces a model that, when tested on the validation data, gives satisfactory results (Figure 1).

To manually identify the targets, we used a tool available online, MakeSenseAI (n.d.), which works in three distinct stages: 1) upload the set of 33 images for annotation, 2) identify the OoI in each image with the tools available in the graphical interface and 3) download the annotations in the form of text files in a format compatible with YOLO.

The set of annotated images and corresponding labels is divided into two subsets, one (22 images) for training and one (11 images) for validation. The first contains 781 annotated examples of the objects of interest and the second contains 465, together constituting a representative sample of the targets, both in terms of position and location. The remaining 180 image tiles were used to evaluate the performance of the model.

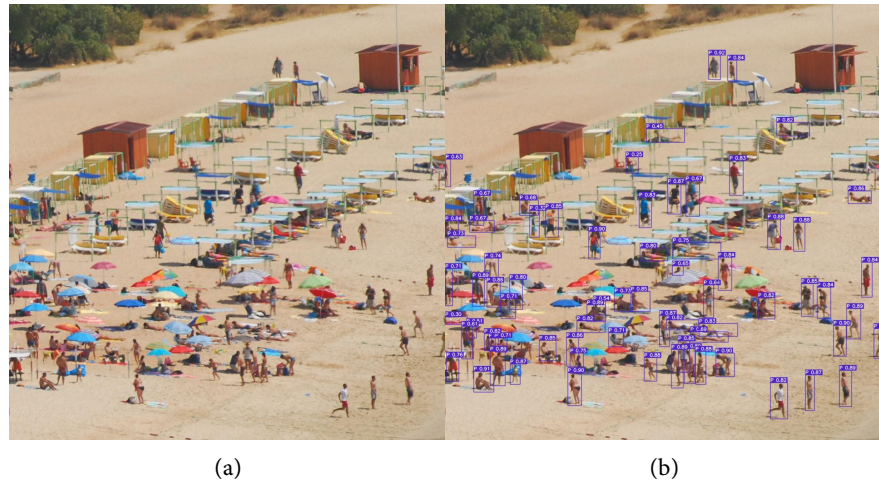
One of the strengths of YOLO v7 and its previous versions is the possibility of transfer learning (Wang et al., 2022), which consists of reusing an already trained network as the basis for a new problem. Since basic features such as edges, forms or shapes are common to many different objects, a trained network can be used to implement a new problem, with a well-established set of initial weights from training on very large datasets, such as Common Objects in Context (COCO), which was trained with over 320,000 annotated images (Lin et al., 2015). The new discriminators coming from the images we have annotated are added to this solid base, defining the last layers of the CNN, and adjusting the detector to the new problem.

The train was done once and took two hours nineteen minutes for 750 iterations in a laptop equipped with dual Core Intel i7-10750H processor, 16 GB SDRAM, and a graphic unity NVIDIA GeForce RTX 2060 Max-Q 6GB. The resulting weights defining the new model can be used to detect the same objects of interest on any similar image, with a processing time around 0.2 s for each image.

The results are quantified as usual in terms of Precision and Recall. Precision is defined as the percentage of objects correctly classified in the set of all the targets identified by the model as positives (Equation (1)).

$$\text{Precision} = \text{True Positives} / (\text{True Positives} + \text{False Positives}) \quad (1)$$

Recall concerns the percentage of positives correctly detected in the sum of all occurrences really positives (Equation (2)), that include false negatives—undetected occurrences.



**Figure 1.** Example of an image before and after inference processing. The objects of interest detected by YOLO v7 in image (a) are marked in blue in image (b), with the respective confidence. The confidence threshold is 0.24.

$$\text{Recall} = \text{True Positives} / (\text{True Positives} + \text{False Negatives}) \quad (2)$$

The F1 score is a cross measure of the accuracy of a model in a data set, based in the harmonic mean of the accuracy and recall of the results (Equation (3)).

$$\text{F1} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (3)$$

A high value of F1 (near 1, as F1 range is [0, 1]) indicates perfect results of both precision and recall, and a low value means that the model has a poor performance either in precision, recall or both.

### 3. Results and Discussion

The largest model of YOLO v7 was used (e6e) and trained with a few custom alterations in the default hyperparameters, to be fully applicable to the nature of the “human” data, namely the possible change in scale was constrained to a factor of 1:2, rotation to 0.5 degrees, and translation to zero pixels. Shear, perspective and flip upside down are kept to zero; the admissible alterations for hue, saturation and intensity in the color space are the defaults, 0.015 for hue, 0.7 for saturation and 0.4 for intensity.

The resulting model, presenting a precision of 0.83, recall of 0.78 and mAP@50 of 0.82, was fine-tuned to find the best values for the threshold intersection over union (IoU) and the minimum confidence to be applied in the inference. The first one concerns the ratio of intersection to union of the forms predicted and used in the train stage, contributing to the inference phase to avoid double and triple detections centered in the same target.

The confidence threshold determines the minimum confidence we demand

from the model about each target really being of the class trained. As the confidence for each detection may be displayed in the classified image along with the bounding box around the object detected, we can easily assess whether any changes are required. A low value of confidence will include detections with a small probability of being true positives, while a higher confidence threshold will keep only detections with a high probability of such. We use an IoU value of 0.35 and a confidence threshold of 0.24.

The model was applied to the remaining test dataset of 180 images, and the results were evaluated by human curation, resulting in three numbers associated to each test image—number of targets detected by the model, and numbers of false positives and false negatives amongst the detections (**Table 1**).

**Table 1.** Detections, false positives and false negatives in 180 images.

Detected (conf = 0.24, IoU = 0.35)	5401
True Positives	5672
False Positives	116
False Negatives	387

The model achieved gives an overall precision of 0.980 and a recall of 0.936 on the test data set. Considering the 93.2% of true positives detected (correctly detecting 5285 in 5672 targets identified by human curation in the 180 images) and the processing times around 0.2 s per image, it constitutes an interesting result for application in surveillance, with an F1 score of 0.958, considering that this deep-learning tool is open-source software and can be installed on an average laptop without any special top-end hardware requirements. The train stage needs a human operator to identify the objects of interest in a subset of images, with an online tool, which can consume a few hours of processing, or not that much, if automatic identification can be used.

The inference has two outputs: an image with all the occurrences identified by the model surrounded by boxes with the confidence of each detection attached, and a numeric output including the total of objects detected and processing time.

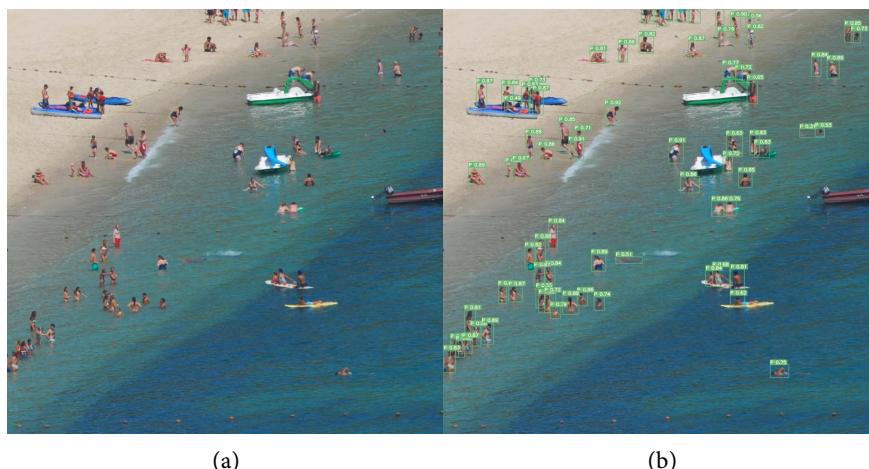
The parametrization of the inference, meaning the values for the ideal confidence threshold and IoU for a particular dataset, is the most demanding step in the procedure, and strongly depends on the weights given to false positives and false negatives in the problem, so it depends on the goal involved in each application.

We focus on minimizing false positives, so we worked with a parametrization that produces 98.3% of the 180 test images with a maximum of 3 false positives and 76.1% with a maximum of 3 false negatives, being able to detect many partial targets (**Figure 2**).

In **Figure 2**, we can see several objects partially hidden by the water that are detected with high confidence, as well as many targets in different positions that do not represent a challenge to this kind of algorithm. Targets on different backgrounds—water, light-colored dry sand and dark-colored wet sand (**Figure 2**)—



have the same probability of correct detection, since the training of this type of model focuses on the characteristics of the target, regardless of the background.



**Figure 2.** Example of an image before and after inference with  $\text{IoU} = 0.35$  and confidence threshold 0.24. The objects of interest detected by YOLO v7 in image (a) are marked in green in image (b), with the respective confidence.

## 4. Conclusion

The acquisition of images with a high line of sight to a beach area allows the use of the methodology described to quantify the human surcharge with good precision, without any estimates or unnecessary interactions, in any time lapse desired.

To implement this type of deep learning solution an initial annotation stage must be carried out outside the workflow, preferably by a human operator who annotates the train and validation images, due to the wide variety of poses on the beach that would not be easily identified in automatic labeling. The variety of situations observed is beyond the data augmentation capabilities offered by hyper parameterization, which is mostly limited to affine transformations.

Once the model is ready, each image with any number of targets present will take less than a second to process, and the final figures will be more consistent, since they are not subject to the subjectivity of a human interpreter.

There are situations in which this type of remote surveillance is highly desirable, such as certain beaches where the lack of consolidation of the cliffs is a known risk, but which are frequented anyway, often from access by sea; these are situations in which surveillance has two relevant aspects at the same time, the safety of people and the preservation of ecosystems (Sardá et al., 2015).

The procedure described can be applied to any guarded area where the human load has become critical, such as historic sites or fragile dune ecosystems, where it could be used to keep the flow of people within the observed area of interest to reasonable numbers, for instance with data exchange with the entry points.

## Funding

This study had the support of national funds through Fundação para a Ciência e

Tecnologia, under the project LA/P/0069/2020, (<https://doi.org/10.54499/LA/P/0069/2020>) granted to the ARNET (Aquatic Research Network Associated Laboratory), UIDB/04292/2020, (<https://doi.org/10.54499/UIDB/04292/2020>) and UIDP/04292/2020 (<https://doi.org/10.54499/UIDP/04292/2020>), granted to MARE (Marine and Environmental Sciences Centre).

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- Adade, R., Aibinu, A. M., Ekumah, B., & Asaana, J. (2021). Unmanned Aerial Vehicle (UAV) Applications in Coastal Zone Management—A Review. *Environmental Monitoring and Assessment*, 193, Article No. 154. <https://doi.org/10.1007/s10661-021-08949-8>
- Cabioch, B., & Robert, S. (2022). Integrated Beach Management in Large Coastal Cities. A Review. *Ocean & Coastal Management*, 217, Article ID: 106019. <https://doi.org/10.1016/j.ocecoaman.2021.106019>
- Corraini, N. R., de Souza de Lima, A., Bonetti, J., & Rangel-Buitrago, N. (2018). Troubles in the Paradise: Litter and Its Scenic Impact on the North Santa Catarina Island Beaches, Brazil. *Marine Pollution Bulletin*, 131, 572-579. <https://doi.org/10.1016/j.marpolbul.2018.04.061>
- Domingo, M. C. (2021). Deep Learning and Internet of Things for Beach Monitoring: An Experimental Study of Beach Attendance Prediction at Castelldefels Beach. *Applied Sciences*, 11, Article 10735. <https://doi.org/10.3390/app112210735>
- GitHub (n.d.). *WongKinYiu/yolov7: Implementation of Paper—YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors*. <https://github.com/WongKinYiu/yolov7>
- Green, S., Blumenstein, M., Browne, M., & Tomlinson, R. (2005). The Detection and Quantification of Persons in Cluttered Beach Scenes Using Neural Network-Based Classification. In *Proceedings of the Sixth International Conference on Computational Intelligence and Multimedia Applications (ICCIMA '05)* (pp. v-x). IEEE. <https://ieeexplore.ieee.org/document/1540692>
- Lin, T., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D. et al. (2015). Microsoft COCO: Common Objects in Context. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer Vision—ECCV 2014* (pp. 740-755). Springer International Publishing. [https://doi.org/10.1007/978-3-319-10602-1\\_48](https://doi.org/10.1007/978-3-319-10602-1_48)
- Luna da Silva, R., Chevtchenko, S., Alves de Moura, A., Rolim Cordeiro, F., & Macario, V. (2017). Detecting People from Beach Images. In *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)* (pp. 636-643). IEEE. <https://doi.org/10.1109/ictai.2017.00102>
- MakesenseAI (n.d.). <https://www.makesense.ai/>
- Matlab (n.d.). *MathWorks. Products & Services*. [https://www.mathworks.com/products.html?s\\_tid=gn\\_ps](https://www.mathworks.com/products.html?s_tid=gn_ps)
- Morgan, D. (2018). Counting Beach Visitors: Tools, Methods and Management Applications. *Coastal Research Library*, 24, 561-577. [https://doi.org/10.1007/978-3-319-58304-4\\_27](https://doi.org/10.1007/978-3-319-58304-4_27)



- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 779-788). IEEE. <https://doi.org/10.1109/cvpr.2016.91>
- Sardá, R., Valls, J. F., Pintó, J., Ariza, E., Lozoya, J. P., Fraguell, R. M. et al. (2015). Towards a New Integrated Beach Management System: The Ecosystem-Based Management System for Beaches. *Ocean & Coastal Management*, 118, 167-177. <https://doi.org/10.1016/j.ocecoaman.2015.07.020>
- Thushari, G. G. N., & Senevirathna, J. D. M. (2020). Plastic Pollution in the Marine Environment. *Heliyon*, 6, e04709. <https://doi.org/10.1016/j.heliyon.2020.e04709>
- Turner, I. L., Harley, M. D., & Drummond, C. D. (2016). UAVs for Coastal Surveying. *Coastal Engineering*, 114, 19-24. <https://doi.org/10.1016/j.coastaleng.2016.03.011>
- Wang, C., Bochkovskiy, A., & Liao, H. M. (2022). YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 7464-7475). IEEE. <https://doi.org/10.1109/cvpr52729.2023.00721>