Scientific Research Publishing

Journal of Computer and Communications, 2024, 12, 31-39
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

Efficient People Detection with Infrared Images

Maria da Conceição Proença

Marine and Environmental Sciences Centre & ARNET, Aquatic Research Infrastructure Network Associated Laboratory, Department of Physics, Faculty of Sciences, University of Lisbon, Lisbon, Portugal
Email: mcproenca@fc.ul.pt

How to cite this paper: Proença, M.C. (2024) Efficient People Detection with Infrared Images. *Journal of Computer and Communications*, **12**, 31-39.
https://doi.org/10.4236/jcc.2024.124003

Received: January 24, 2024
Accepted: April 6, 2024
Published: April 9, 2024

Abstract

This work focuses on the problem of monitoring the coastline, which in Portugal's case means monitoring 3007 kilometers, including 1793 maritime borders with the Atlantic Ocean to the south and west. The human burden on the coast becomes a problem, both because erosion makes the cliffs unstable and because pollution increases, making the fragile dune ecosystem difficult to preserve. It is becoming necessary to increase the control of access to beaches, even if it is not a popular measure for internal and external tourism. The methodology described can also be used to monitor maritime borders. The use of images acquired in the infrared range guarantees active surveillance both day and night, the main objective being to mimic the infrared cameras already installed in some critical areas along the coastline. Using a series of infrared photographs taken at low angles with a modified camera and appropriate filter, a recent deep learning algorithm with the right training can simultaneously detect and count whole people at close range and people almost completely submerged in the water, including partially visible targets, achieving a performance with F1 score of 0.945, with 97% of targets correctly identified. This implementation is possible with ordinary laptop computers and could contribute to more frequent and more extensive coverage in beach/border surveillance, using infrared cameras at regular intervals. It can be partially automated to send alerts to the authorities and/or the nearest lifeguards, thus increasing monitoring without relying on human resources.

Keywords

Beach Overload, People Counting, Border Control, People Detection, Deep Learning Methods, Remote Surveillance

1. Introduction

In surveillance implementations, it is sometimes enough to have a human/non-

human tag assigned to a target. A task that seems simple in daylight to an operator observing the scene can easily be complicated at night. If the targets also must be counted, even in daylight the problem will increase, there is the possibility of overlap and a whole variety of relative positioning of the targets, which turns a simple problem into a complex one. The limited success of conventional image processing techniques with targets that can take on many different aspects and positions triggers us to deep learning techniques.

The broad application of deep learning techniques [1]-[7] has led to many recent developments in detection in several areas of work where supervised techniques needing trained algorithms are used, mainly due to the possibility of data augmentation. Data augmentation consists of the inclusion of randomized and small changes in the train dataset, in the form of alterations in dimensions, translation, rotation, and shearing as well as small differences allowed in the radiometric range. The main goal is to complement the train images used, getting to train sets including different perspective views and more possible positions and sizes of the targets that could not be assured in the initial dataset.

The software used is one of the latest developments of one-stage algorithms based on convolutional neural networks (CNNs) and the 7th version of the concept You Only Look Once (YOLO) [8]. Although the implementation of Yolo v7 [9] is not related to the initial author of the concept, it has substantial support published [10].

Monitoring people on beaches and water for security reasons is rather more complicated than the usual surveillance because many of the approaches already developed for people detection focus on the head-shoulder binomial [11] [12] [13] [14] [15] and only a few deals with the problem of possible concealment by overlapping [16] [17]. In 2017 [18] developed an approach specific to monitoring large extensions of sand in Brazil, where the main risk for humans comes from the sea (drowning or shark attack) using images acquired in the visible range and working with a collection of image descriptors: Gabor filter, Hu and Zernike moments, Histogram of oriented gradients and Local binary pattern, followed by a classifier. Early work to quantify people in beach scenes [19] to predict trends in tourist activity, describes a two-stage procedure including a moving average algorithm to find potential targets that, once segmented, are submitted to a neural-based type of classification system. Another work in beach surveillance [20] focuses on the interaction between local observations and external data, presenting a system for feature extraction and integration with weather forecasts and wave data, aiming to assess and predict beach safety conditions (mainly for recreational users like surfers).

Infrared imagery can be found in countless areas of current work [21] [22], and has already been mentioned as useful to biometric technology used in border control [23] but we haven't found any specific work related to the one presented here, both in terms of application and simplicity and performance—a highly favorable cost/benefit ratio, considering that it is an algorithm available

online and can run on a laptop with current hardware characteristics.

2. Methodology

Images acquired in infrared have the advantage, among others, of requiring very little ambient light, providing clear images both day and night. To simulate the images that can be acquired by a surveillance infrared camera aimed at the interface sand/water in a coastline, we use infrared images taken at Monte Gordo, a beach on Portugal's southern border that is part of a continuous 20 km stretch of sand. The images were acquired at a low observation angle, from the beach, with a Canon EOS M50 modified to full-spectrum and Canon EF-M 15 - 45 mm lens with an infrared filter Hoya R72. The proprietary RAW digital format was converted to jpg without compression in open-source software [24]. The only preprocessing needed to ensure a homogeneous dataset was implemented in a Matlab environment: location and extraction of the areas of interest followed by dimensions normalization. The large images were cut into tiles of 960 × 960 pixels for reasons of processing speed in the hardware available.

The software used to detect and count people in the images is YOLO v7, released in July 2022 and made publicly available in a GitHub repository [9]. To use YOLO v7 on a new dataset (**Figure 1**) it needs to go through a training stage with a convenient number of images annotated: the objects of interest belonging to the class that is intended to be detected should be identified in a representative subset of images.

This can be done online, for instance with [25] in three simple steps consisting in upload the data set intended for training the algorithm, identify all the occurrences in each image, and download a set of text files with the annotations in YOLO format.

YOLO v7 and its precedent versions have another advantage over conventional algorithms: the possibility of transfer learning, which consists of reusing a

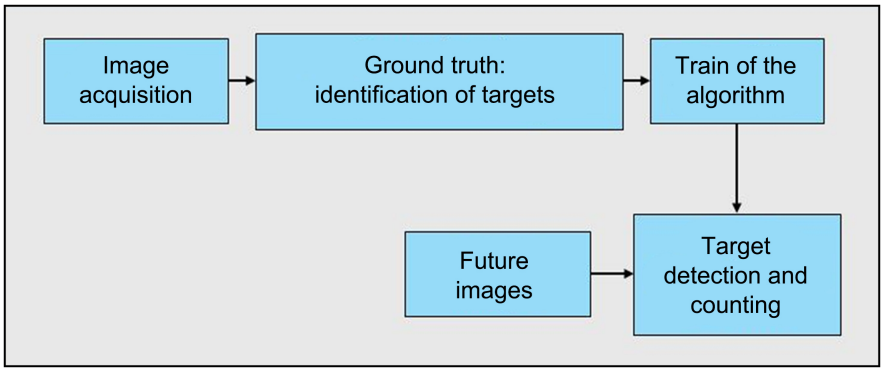


Figure 1. Conceptual diagram describing the data processing steps: 1) in a subset of images all the targets are manually annotated; these annotations are the usual "ground truth" used in all supervised classifiers; 2) the algorithm is trained with the set of images annotated in several iterations (750 in the present example); 3) the resulting algorithm can be applied to any image of the same kind, detecting and counting the targets to which it is trained.

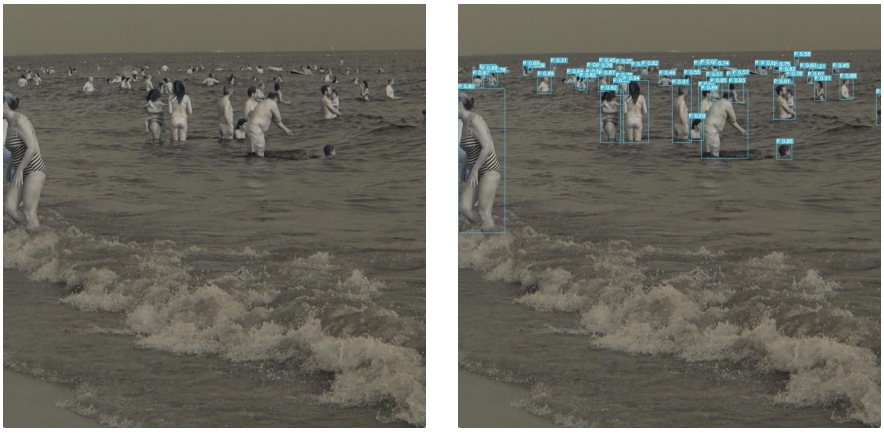
trained network to implement a new problem. Since basic features such as edges, contrasts or shapes are common to many detection problems, an already trained network can be used to implement a new problem, with a set of initial weights from training on very large datasets such as Common Objects in Context (COCO), with 80 classes and over 200,000 annotated images. The discriminators formed from the new training data will define the last layers of the CNN, tunning the detector according to the details of the new application, while the basics defined by the first layers are used as a solid base.

3. Results

We used the e6e model from YOLO v7 and train the new people detection model with augmented data, changing only a few hyperparameters from the default values, namely the amounts of rotation to 0.5 degrees, translation to 0 pixels and the change in scale allowed to 0.5. The annotation of 33 images of a subset of 113 tiles (960 × 960 pixels, 24-bit depth) with one class of objects of interest was done online [25]. Twenty-two images were used for the train and eleven for validation, respecting the recommended ratio of 2:1 between the train and validation images.

The train was done once and took two hours twenty six 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 (**Figure 2**), with a processing time of 0.220 s for each image.

The inference results were fine-tuned: the minimum confidence threshold that reflects the probability associated by the network to each detection of being a true positive was set to 0.20, and the intersection over union threshold, that in the train stage concerns the ratio of intersection to union of the forms predicted

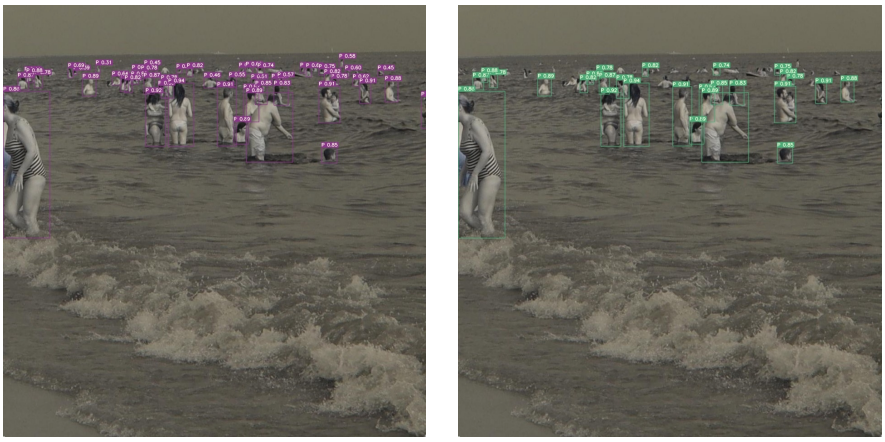


(a) (b)

Figure 2. Example of inference. 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 0.20.

and used, and in the inference phase has the practical result of avoiding double and triple detections of the same subject, was set to 0.30. The confidence for each detection is displayed in the classified image, so it's possible to assess whether any change of the confidence threshold is needed, and in which direction: a lower value will include detections with a smaller probability of being true positives (Figure 3(a)), while a confidence threshold closer to one will limit the count to targets detected with higher confidence (Figure 3(b)).

The recurring problem of concealment is partially solved, but there are still situations in which only a human eye can tell that there is a second target behind the obvious one; Figure 4 shows an image in which partial concealment is solved in three situations with people in water, but we can suspect that the figure in the foreground has someone next to them that the model doesn't detect.



(a) (b)

Figure 3. (a) The image shown in Figure 1(a) with a confidence threshold of 0.3 makes 51 detections; (b) changing the confidence threshold to a more restrictive value of 0.7 reduces the detections to 27 with higher confidence.

(a) (b)

Figure 4. (a) An image where there are several cases of partial concealment; (b) most cases are solved by the model, leaving the case in the foreground where only a human operator can detect the signs of a second target, that can or cannot be there.

The results of the model were evaluated by human curation on the remaining 80 test tiles after detection with the aforementioned parametrization and quantified in terms of Precision and Recall. Precision is defined as the percentage of particles correctly classified among all the particles identified by the algorithm as positives as in Equation (1), including those not identified correctly.

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

Recall refers to the percentage of positives correctly detected in the sum of all occurrences of real positives (Equation (2)).

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

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

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

The resulting F1 score is in the range [0, 1], the highest value indicating perfect precision and recall, and the lowest meaning either precision or recall is zero.

The model gives an overall precision of 0.97 and a recall of 0.92. Considering the 92% of true positives detected (correctly detecting 1647 in 1970 targets identified by human curation) and adding that the counting time was reduced to a few hundred milliseconds, it constitutes an acceptable result, with an F1 score of 0.945.

4. Discussion

The methodology described only requires public domain software, running programs that can be used by anyone with basic informatic skills. A Matlab environment [26] was used in this work to convert the large images to 960 × 960 pixel tiles, but there are open-source software options available online that can do the same, which consists basically of cropping, such as Image J [27]. The deep-learning tool is also open-source software and can be installed and run on an average personal laptop without any special requirements. The train stage preparation needs a human operator to annotate a subset of images containing the objects of interest with an online tool and consumes a few hours of processing. The inference consists of one line of commands and has two outputs: an image with all the occurrences identified by boxes around the targets and the confidence of each detection attached, and a screen output with the total of objects detected and processing time. The parametrization of the inference is the most demanding step, and the most delicate, because the relevance of false positives and false negatives for each situation is a function of the specific priorities of the problem, so several sets of parameters must always be evaluated in the test images.

5. Conclusions

The methodology described is an accurate way to quantify people, specifically in

a beach environment both outside and inside the water. Images are already being acquired automatically in several coastal areas that are either relevant for recreational purposes, such as surfing, or known as entry points often used by smugglers. After an initial training phase that can be carried out outside the implemented workflow, each image will take less than a quarter of a second to process, regardless of the number of targets present, without any subjectivity or delay, day and night, and much less prone to errors than a human operator. The overall procedure could easily become semi-automatic, by sending an alarm to lifeguards in the event of someone entering a dangerous area or alerting the competent authorities in the case of border surveillance; nonetheless, image annotation in the training phase remains the only part where human intervention is the preferable option.

The same procedure can be extended to other areas where the human load has become critical, such as highly touristic historical sites, where it could be used to control access without depriving people of enjoying them. Increasing the likelihood of early detection regardless of observation conditions and enabling a rapid and proportionate response is important for both aspects of people security, ecosystem preservation [28], and border surveillance.

Acknowledgements

I would like to thank José E. Alpedrinha who shared some of the photographs from his private collection to make this work possible and Ricardo N. Mendes for enlightened discussions on the problematic of borders.

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 author declares no conflicts of interest regarding the publication of this paper.

References

[1] Cao, C., Liu, F., Tan, H., Song, D., Shu, W., Li, W., Zhou, Y., Bo, X. and Xie, Z. (2018) Deep Learning and Its Applications in Biomedicine. *Genomics, Proteomics & Bioinformatics*, **16**, 17-32. https://doi.org/10.1016/j.gpb.2017.07.003

[2] Wang, J., Ma, Y., Zhang, L., Gao, R.X. and Wu, D. (2018) Deep Learning for Smart Manufacturing: METHODS and Applications. *Journal of Manufacturing Systems*, **48**, 144-156. https://doi.org/10.1016/j.jmsy.2018.01.003

[3] Perera, P. and Patel, V.M. (2019) Deep Transfer Learning for Multiple Class Novelty Detection. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 11536-11544. https://doi.org/10.1109/CVPR.2019.01181

[4] Ouyang, Y., Wang, K. and Wu, S. (2019) SAR Image Ground Object Recognition Detection Method Based on Optimized and Improved CNN. *IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference*, Chengdu, 20-22 December 2019, 1727-1731. https://doi.org/10.1109/IAEAC47372.2019.8997680

[5] Masita, K., Hasan, A. and Shongwe, T. (2020) Deep Learning in Object Detection: A Review. 2020 *International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems* (*icABCD*), Durban, 6-7 August 2020, 1-11. https://doi.org/10.1109/icABCD49160.2020.9183866

[6] Talaei Khoei, T., Ould Slimane, H. and Kaabouch, N. (2023) Deep Learning: Systematic Review, Models, Challenges, and Research Directions. *Neural Computing and Applications*, **35**, 23103-23124. https://doi.org/10.1007/s00521-023-08957-4

[7] viso.ai (2022) The 100 Most Popular Computer Vision Applications in 2024.

[8] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 779-788. https://doi.org/10.1109/CVPR.2016.91

[9] GitHub (2023) WongKinYiu/yolov7: Implementation of Paper—YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors.

[10] Wang, C., Bochkovskiy, A. and Liao, H. (2022) YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Vancouver, 17-24 June 2023, 7464-7475. https://doi.org/10.1109/CVPR52729.2023.00721

[11] Tu, J., Zhang, C. and Hao, P. (2013) Robust Real-Time Attention-Based Head Shoulder Detection for Video Surveillance. 2013 *IEEE International Conference on Image Processing*, Melbourne, 15-18 September 2013, 3340-3344. https://doi.org/10.1109/ICIP.2013.6738688

[12] Wang, S., Zhang, J. and Miao, Z. (2013) A New Edge Feature for Head-Shoulder Detection. 2013 *IEEE International Conference on Image Processing*, Melbourne, 15-18 September 2013, 2822-2826. https://doi.org/10.1109/ICIP.2013.6738581

[13] Hu, R., Wang, R., Shan, S. and Chen, X. (2014) Robust Head-Shoulder Detection Using a Two-Stage Cascade Framework. 22*nd International Conference on Pattern Recognition*, Stockholm, 24-28 August 2014, 2796-2801. https://doi.org/10.1109/ICPR.2014.482

[14] Guan, Y. and Huang, Y. (2015) Multi-Pose Human Head Detection and Tracking Boosted by Efficient Human Head Validation Using Ellipse Detection. *Engineering Applications of Artificial Intelligence*, **37**, 181-193. https://doi.org/10.1016/j.engappai.2014.08.004

[15] Hsu, F.C., Gubbi, J. and Palaniswami, M. (2015) Head Detection Using Motion Features and Multi-Level Pyramid Architecture. *Computer Vision and Image Understanding*, **137**, 38-49. https://doi.org/10.1016/j.cviu.2015.04.007

[16] Wang, X., Han, T.X. and Yan, S. (2009) An HOG-LBP Human Detector with Partial Occlusion Handling. *IEEE 12th International Conference on Computer Vision*, Kyoto, 29 September-2 October 2009, 32-39. https://doi.org/10.1109/ICCV.2009.5459207

[17] Al-Zaydi, Z.Q.H., Ndzi, D.L., Yang, Y. and Kamarudin, M.L. (2016) An Adaptive People Counting System with Dynamic Features Selection and Occlusion Handling. *Journal of Visual Communication and Image Representation*, **39**, 218-225. https://doi.org/10.1016/j.jvcir.2016.05.018

[18] Silva, R., Chevtchenko, S., Moura, A., Cordeiro, F. and Macario, V. (2017) Detecting People from Beach Images. 2017 *International Conference on Tools with Artificial Intelligence*, Boston, 6-8 November 2017, 636-643. https://ieeexplore.ieee.org/document/8372005

[19] Green, S., Blumenstein, M., Browne, M. and Tomlinson, R. (2005) The Detection and Quantification of Persons in Cluttered Beach Scenes Using Neural Network-Based Classification. *Sixth International Conference on Computational Intelligence and Multimedia Applications* (*ICCIMA*'05), Las Vegas, 16-18 August 2005, 303-308. https://dl.acm.org/doi/10.1109/ICCIMA.2005.57 https://doi.org/10.1109/ICCIMA.2005.57

[20] Browne, M., Blumenstein, M., Tomlinson, R. and Lane, C. (2005) An Intelligent System for Remote Monitoring and Prediction of Beach Conditions. *Proceedings of the International Conference on Artificial Intelligence and Applications*, Innsbruck, Austria, 14-16 February 2005, 533-537.

[21] Rao, P.S., Rani, S.P., Badal, T. and Guptha, S.K. (2020) Object Detection in Infrared Images Using Convolutional Neural Networks. *Journal of Information Assurance and Security*, **15**, 136-143.

[22] Bustos, N., Mashhadi, M., Lai-Yuen, S., Sarkar, S. and Das, T. (2023) A Systematic Literature Review on Object Detection Using Near-Infrared and Thermal Images. *Neurocomputing*, **560**, Article ID: 126804. https://dl.acm.org/doi/10.1016/j.neucom.2023.126804 https://doi.org/10.1016/j.neucom.2023.126804

[23] Kyeremeh, G.K., Abdul-Al, M., Qahwaji, R. and Abd-Alhameed, R.A. (2022) Infrared Imagery and Border Control Systems. 2*nd International Multi-Disciplinary Conference Theme*: *Integrated Sciences and Technologies*, Sakarya, 7-9 September 2021. https://eudl.eu/doi/10.4108/eai.7-9-2021.2314979 https://doi.org/10.4108/eai.7-9-2021.2314979

[24] IrfanView (2023) IrfanView v4.59—Official Homepage. https://www.irfanview.com/

[25] (2022) Make Sense. AI. https://www.makesense.ai/

[26] Matlab (2022) Matlab—MathWorks—MATLAB & Simulink. https://www.mathworks.com

[27] Image J. (2022) Image J. https://imagej.net/ij/index.html

[28] Sardá, R., Valls, J.F., Pintó, J., Ariza, E., Lozoya, J.P., Fraguell, R.M., Martí, C., Rucabado, J., Ramis, J. and Jimenez, J.A. (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