Object Identification for Visually Impaired

Michell Thompson Ferreira Santiago[1,2*], Marcus Vinicius Barros Faria[2], Marcos Vinicius Pinto Malvar[2], Luccas Oliveira Nascimento[3]

[1]SENAI CIMATEC University Center; [2]University Center UNIRuy; [3]ÁREA 1 Faculty; Salvador,Bahia, Brazil

In computer vision, several algorithms carry out the process of object recognition through machine learning. Among them, You Only Look Once (YOLO) stands out, a Darknet algorithm that had its most significant relevance in the YoloV4 version for presenting an outstanding performance on both personal computers and mobile devices. This article presents a feasibility study of this tool for the possibility of helping visually impaired people detect objects through their mobile devices and thus help in everyday tasks.

Keywords: Visual Impairment. Android Studio. Convolutional Neural Network. YOLOv4.

Introduction

Computer engineering is an area that covers many aspects related to information technology, including hardware, software, and communication systems. Automation engineering deals with the automation of industrial processes involving the control of automated equipment and systems. Both areas aim to create and improve systems that simplify and streamline complex tasks.

A relevant topic with significant social impact is the development of applications aimed at helping people with visual impairments. According to Souza and Oliveira (2018) [1], the lack of accessibility to applications can limit the independence and participation of these people in society since many services and information are available exclusively through mobile devices. Among these applications, the ones that stand out the most are those that use machine learning and image segmentation techniques to recognize objects and people in the environment.

According to Pires and Lima (2017) [2], object recognition technology has been increasingly used in creating applications that help visually impaired people identify objects around them. There are several computer vision training models, and one presenting high performance in computational cost is the YOLO algorithm (you only look once). Through neighborhood recognition, this algorithm analyzes a group of pixels on the screen and verifies the possible correlated outputs through a Convolutional Neural Network (CNN), thus detecting an object class.

This work aims to develop an application for mobile devices that uses image processing and machine learning techniques for object recognition and description for people with visual impairments. In addition, object recognition technology has been widely used in different areas, such as robotics, security, and electronic commerce [3]. Tools and technologies in mobile application development, such as Android Studio and frameworks like TensorFlow Lite and Yolov4, will provide a practical and accessible solution for this specific audience.

Materials and Methods

The training used Python to create the convolutional neural network through Colab and the YOLOv4 learning algorithm. At the end, a report containing the accuracy metrics obtained during training was generated. To confirm these metrics and ensure the operation and usability of the application in real scenarios, images of practical scenarios were used in different conditions to analyze the results more comprehensively.

Received on 15 September 2023; revised 14 November 2023.
Address for correspondence: Michell Thompson Ferreira Santiago. Avenida Orlando Gomes, 1845, Piatã. Salvador, Bahia, Brazil. Zipcode: 41650-010. E-mail: michell.thompson@gmail.com.

J Bioeng. Tech. Health 2023;6(4):296-300

Following a resolution configuration of 416 x 416 pixels in the JPG format with the most varied characteristics of luminosity, position, and distance, it is possible to create a dataset of images that represents different situations of the objects (Figure 1). The labeling process is carried out from these images, in which each image is carefully labeled using the Labellmg program. After obtaining and preparing the dataset, the next step involved training and testing the Convolutional Neural Network (CNN). The implementation of the CCR was carried out using the Python programming language and the YoloV4 algorithm in conjunction with the Darknet framework. Furthermore, Colab provides a GPU integrated into the virtual environment, which is needed to increase the training speed due to the CUDA feature being available only on NVIDIA graphics cards. A set of 1300 images was used, 1040 for training and 260 for testing the model. These images covered different situations, lighting characteristics, positions, and distances, comprehensively representing the study objects: person, car, and fan. During the training of the RNC, adjustments were made to the parameters to achieve the best possible performance. The training step consisted of feeding the labeled data set. A cross-validation strategy was adopted to assess the model's generalizability, dividing the data set into training and validation subsets. In total, 6,000 epochs were performed, representing cycles in which all images were submitted to training, following the guidelines of AlexeyAB (2021) [4]. The batch size was set to 64 with 16 subdivisions for training efficiency.

Results and Discussion

After completing the training, the Darknet framework provides a report to analyze the precision and accuracy of the model in detecting the objects of interest (Figure 2). Performance metrics, such as average precision (mAP), precision in each class (recall), F1-score, and average IoU (Intersection Index over Average Union), were used to measure how well the Convolutional Neural Network (CNN) was able to correctly identify and locate the study elements, such as person, car, and fan.

However, if more images are introduced into the dataset, there is room for improvement to increase the model's ability to recognize and locate objects of interest more precisely. The analysis of the results, according to the image provided by the framework, revealed that the average precision for the person class is 58.45%, for the car is 73.60%, and for the fan is 100% (Figure 2). The high precision obtained in the fan

Figure 1. Class labeling.

class is due to the number of images available in the database and its geometric nature, which generally presents recurrent characteristics, such as propeller blades. The person class, on the other hand, obtained a lower average precision due to the significant variability of characteristics and appearances present in the data set. After epoch 1400, values are more constant, resulting in insignificant changes in machine learning. However, it is still possible to see a relationship between the number of epochs used and the percentage of success, becoming clear when reaching epoch 4800, reaching 78% accuracy (Figure 3).

Using the GitHub of user hunglc007 (2020) [5], where it is possible to find a specific Android project for neural network models trained by YoloV4, the training performance was verified practically, in which there would be some adversities, such as different levels lighting and distance to objects. Figure 4 presents the results of the practical tests in which it is possible to notice that even with low luminosity, a high precision value was obtained.

The tests revealed a very high recognition capacity in the three classes used, even when the luminosity changes (Figure 4). They demonstrated that the trained model is not biased; it is not limited to identifying only the images with which it was trained. The precision and quality of the model are directly related to the amount of information available, also highlighting that the person class was the one that presented the lowest detection reliability due to its several distinct characteristics. The application showed satisfactory results in detecting and obtaining reasonable reliability in the average detection of the proposed objects, with values above 50%. It is essential to have a comprehensive and diversified dataset to train the model to obtain more consistent and reliable results. These results show the effectiveness of the developed application in object recognition and identification. They confirm the application's potential to improve the independence and quality of life of people with visual impairments, providing an accessible and reliable tool for identifying objects in their surroundings.

Conclusion

Using YOLOv4 with the Darknet framework, it was possible to develop an RNC to detect the objects proposed in this article. It was found that the quality and reliability of the tested images showed more significant variability in classes where the characteristics were adverse. Finally, the results were satisfactory, detecting the required classes and informing the user, through a sound message, informing the class of the objects in front of him. However, it is possible to increase

Figure 2. Performance data.

```
 calculation mAP (mean average precision)...
 Detection layer: 30 - type = 28
 Detection layer: 37 - type = 28
60
 detections_count = 609, unique_truth_count = 223
class_id = 0, name = pessoa, ap = 58.45%        (TP = 54, FP = 34)
class_id = 1, name = carro, ap = 73.60%         (TP = 84, FP = 28)
class_id = 2, name = ventilador, ap = 100.00%           (TP = 16, FP = 0)

 for conf_thresh = 0.25, precision = 0.71, recall = 0.69, F1-score = 0.70
 for conf_thresh = 0.25, TP = 154, FP = 62, FN = 69, average IoU = 52.94 %

 IoU threshold = 50 %, used Area-Under-Curve for each unique Recall
 mean average precision (mAP@0.50) = 0.773498, or 77.35 %
Total Detection Time: 1 Seconds
```

Figure 3. Season training.

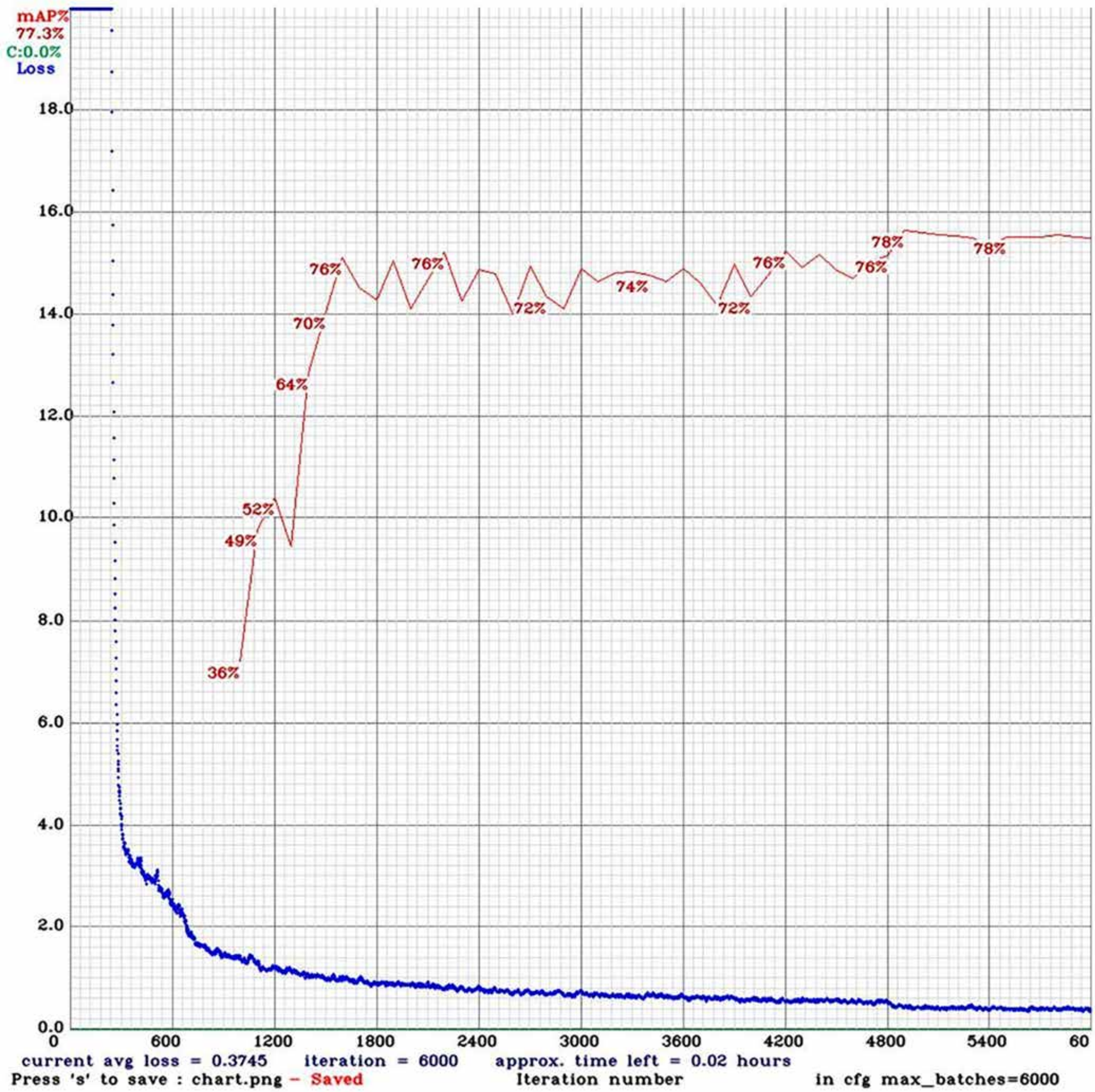
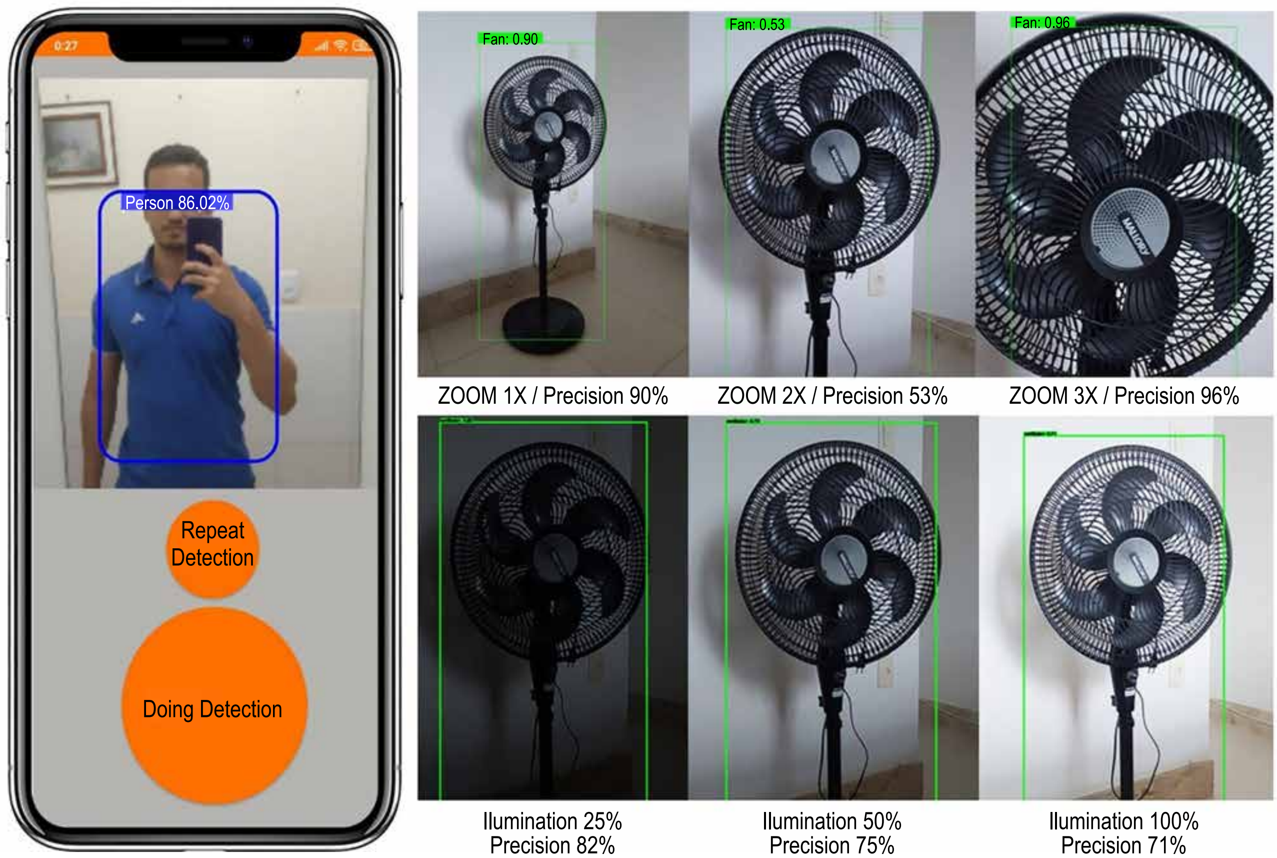


Figure 4. Reliability test.



more classes in the proposed RCC model from a new training, as long as enough images are available to create an image bank for the new desired class and thus perform reliable training.

References

1. Souza JA, Oliveira MB. A falta de acessibilidade em aplicativos pode limitar a independência e a participação dessas pessoas na sociedade. In: Congresso Brasileiro de Informática na Educação, 25., 2018, Fortaleza. Anais [...]. Fortaleza. Sociedade Brasileira de Computação 2018:1-8. Available at: https://sol.sbc.org.br/index.php/educar/article/view/5103. Accessed on:April 30,2023.
2. Pires FLM. Utilização da tecnologia de reconhecimento de objetos para auxiliar pessoas com deficiência visual na identificação de objetos. Revista de Tecnologia e Sociedade 2017;13(2):59-70.
3. Pereira F, Carreira J, Kume G. Aplicações de reconhecimento de objetos em robótica, segurança e comércio eletrônico. Revista Brasileira de Computação Aplicada 2018;10(2):47-59.
4. Alexey AB. YOLOv4: Optimal speed and accuracy of object detection. GitHub Repository 2021. Available at: https://github.com/AlexeyAB/darknet, 2021. Accessed on:April 25, 2023.
5. Việt H. hunglc007. Github. Available at: https://github.com/hunglc007/tensorflow-yolov4-tflite/tree/master Accessed on:April 10,2023.