Vehicle detection based on You Only Look Once

Comănac Dragoș-Mihail

***Abstract*—Billions of people that suffer from some form of visual impairment, out of which a significant part is legally blind. Also, a basic human need is mobility, but there aren't enough traditional mobility solutions for all visually impaired persons, such as assistance dogs, thus, for most legally blind people this need can't be easily satisfied. A more scalable solution would be a digital one, which involves computer vision.**

Therefore, the main purpose of this paper is to provide a form of mobile assistive technology, based on object detection for visually impaired persons.

Our object detector is implemented along the lines of You Only Look Once. We train on a subset of Open Images V4 dataset composed of bus, car, and license plate, a single convolutional neural network. Also, we have developed an Android mobile application that uses this object detector in order to visualize the bounding box predictions. The key feature of the application is the accessible live object detection, in which the predictions are converted to sound and played using the mobile device speakers.

I. Introduction

According to the World Health Organization [4], the number of people suffering from some moderate to severe form of distance vision impairment or blindness due to cataract or uncorrected refractive error is around 200 million. Given the number of people that suffer from some form of visual impairment and the fact that computers can substitute visual functionalities, computer vision has the potential to play the main part of assistive technology for visually impaired persons (VIP).

II. Placement in the broader field

Broadly speaking, there are two options of improving public transportation for VIPs: on one hand, classical methods which are based on radio signals, and on the other hand, the more lightweight solutions based on computer vision.

Our solution would fall into this category, but using only an Android phone. We use object detection by developing our own methodology based on YOLO.

III. Object detection methodology

A. Dataset

For training the object detection system we use the Open Images Dataset V4 [2], available at [1]. In total, the dataset contains 9.2 million images, including 14.6 million bounding boxes across 600 classes on 1.74 million images. We use only a subset of classes: bus, car, and license plate or vehicle registration plate as it is called in the original dataset.

B. Model

The object detection model is inspired by YOLOv2 [5]. The neural network is fully convolutional and is composed of three parts: backbone, neck, and head. For the backbone, we use MobileNetv2 [7], the neck is inspired by U-Net [6] and the head is composed of a convolution layer that has 24 filters in our case, so the final output is $13 \times 13 \times 3 \times 8$ after a reshape layer..

C. Loss

During training, we optimize a composed loss function adapted from [3].

D. Data augmentation

We use the following data augmentations: random hue, brightness, contrast, saturation, cutout and mosaic.

E. Training

We train the model for 50 epochs on a GPU using early stop with the patience of 5 epochs and a delta of $1e^{-4}$. This means that if the model does not improve after some epochs, by the given delta, the training stops because we don't want to overtrain, in order to both save time and reduce overfitting also.

F. Inference

The inference represents a pipeline of processing an image and getting the predictions for it. Firstly, there is preprocessing, the second step is passing the image through the actual neural network, and finally there is the postprocessing where the actual boxes are extracted.

IV. Experimental results

In this section we present our final results on the subset of Open Images Dataset V4 [2], comprising three classes: bus, car, and vehicle registration plate and we study the effects of hyperparameter tuning and fine tuning of the model.

V. Conclusions

In conclusion, our solution aims to ease the use of public transport by VIPs. The first step that we have taken in doing this is creating an object detection system that can recognize buses, cars, or vehicle registration plates. This part is implemented using a custom version of YOLOv2 [5] and we obtain, on the test set, a mAP of 70.03%, and for the bus class, we obtain an average precision of 90.01%, for the car class 64.04% and for the vehicle registration plate 56.68%, with a speed of around 5 FPS on a mobile device.

REFERENCES

[1] Ivan Krasin, Tom Duerig, Neil Alldrin, Vittorio Ferrari, Sami Abu-El-Haija, Alina Kuznetsova, Hassan Rom, Jasper Uijlings, Stefan Popov, Shahab Kamali, Matteo Malloci, Jordi Pont-Tuset, Andreas Veit, Serge Belongie, Victor Gomes, Abhinav Gupta, Chen Sun, Gal Chechik, David Cai, Zheyun Feng, Dhyanesh Narayanan, and Kevin Murphy. OpenImages: A public dataset for large-scale multi-label and multi-class image classification. *Dataset available from https://storage.googleapis.com/openimages/web/index.html*, 2017.

[2] Alina Kuznetsova, Hassan Rom, Neil Alldrin, Jasper Uijlings, Ivan Krasin, Jordi Pont-Tuset, Shahab Kamali, Stefan Popov, Matteo Malloci, Alexander Kolesnikov, and et al. The Open Images Dataset V4. *International Journal of Computer Vision*, 128(7):1956–1981, Mar 2020.

[3] Anh Huynh Ngoc. YOLOv2 implementation. Accessed: 29.10.2022, https://github.com/experiencor/keras-yolo2, 2019.

[4] Geneva: World Health Organization. World report on vision. *World Health Organization Publications*, page 77, 2019.

[5] Joseph Redmon and Ali Farhadi. YOLO9000: Better, Faster, Stronger. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6517–6525, 2017.

[6] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *MICCAI 2015. Lecture Notes in Computer Science*, 2015.

[7] Mark Sandler, Andrew G. Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. MobileNetV2: Inverted Residuals and Linear Bottlenecks. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4510–4520, 2018.