UNIVERSITY ROVIRA I VIRGILI

MASTER THESIS

Intelligent Assistant System Based on Single Camera and Deep Neural Networks for Aiding Visual Impaired Individuals

Author:
David GEORGE FAROUK MARIE

Supervisor:
Dr. Domenec PUIG
Dr. Hatem A. RASHWAN

Department Of

Intelligent Robotics and Computer Vision Group - IRCV
Department of Computer Engineering and Mathematics (DEIM), School of Engineering (ETSE)

UNIVERSITAT ROVIRA i VIRGILI
Tarragona
September 2021

Declaration of Authorship

I, David GEORGE FAROUK MARIE, declare that this thesis titled, "Intelligent Assistant System Based on Single Camera and Deep Neural Networks for Aiding Visual Impaired Individuals " and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. Except for such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date: 1/9/2021

Acknowledgements

First, I would like to thank my master coordinator, Dr. Susana Alvarez, for her fast responses and offering a helping hand whenever I needed. Secondly, I want to thank my thesis advisors Dr. Domenec Puig and Dr. Hatem Rashwan of the department of computer engineering and mathematics at the University of Rovira i Virgili. As they were always guiding and showing a helpful hand whenever I had questions or doubts and steered me to the right direction. I am also thankful for my friends (UC) who supported me in different ways as they showed me that family is not only relation between blood. I would also like to thank Armin Masoumian, who guided and taught me about the depth estimation models. Finally, I would have never managed to accomplish this task without the blessings and the loving support of my Brother Domadious George, my father George Farouk and my mother Amal Naeim.

UNIVERSITY ROVIRA I VIRGILI

Abstract

Department of Computer Engineering and Mathematics (DEIM), School of Engineering (ETSE)

Master in computer security and artificial intelligent

Intelligent Assistant System Based on Single Camera and Deep Neural Networks for Aiding Visual Impaired Individuals

by David GEORGE FAROUK MARIE

Object Detection and Depth Estimation methods can help visual impaired individuals to understand the scene in front of them. there are multiple applications that provide help to those individuals by connecting them (by a video call) to others who can help in describing the scene to them. However, We believe that I can give an alternative to those applications using light and fast machine learning models and avoid the interaction of the human support. The project is divided into Two parts. The First part is Object Detection in a scene. We used YOLOV5S model that was trained on the COCO date-set with 80 different objects. The second part in the depth estimation model. We used Midas from pytorch after trying multiple depth models. the depth estimation model will help me to estimate the distance of each object extracted from the depth model. The final goal is to take an image or a video from the user and extract the objects with the distance of each of them and send this data to the user in a sound note format. ***Keywords:*** Scene Description, Object Detection, Depth Estimation, Midas, YOLOV5, deep learning.

Contents

List of Figures

Chapter 1

Introduction

1.1 Overview

1.1.1 Visually impaired

Visually impaired individuals suffer dramatically in their daily life because of the lack of visually impaired aided systems. According to (*Blindness and vision impairment*) there are 2.2 billion people all over the globe have a distance vision impairment. half of those cases could be prevented if it has been addressed sooner. there is a lot of causes for visual impairment. According to (Steinmetz et al., 2021 the causes of visual impairment are uncorrected refractive errors, cataract, age-related macular degeneration, glaucoma diabetic retinopathy, corneal opacity and trachoma

Therefore, this project will provide a solution or system to aid the visually impaired individuals to understand the scene in front of them using AI and machine learning technology.

1.1.2 Scene Description

the goal of this project is to analyze the scene from a picture. By analyzing I mean to detect each object in this picture, and for each object an estimate of distance between it and the camera will be given. The aim is to use a single camera instead of stereo or Lidar cameras for estimating the depth. Depth estimation/prediction from a single image is challenging, and is an ill-posed problem, however it plays a key role to understand a given scene

1.1.3 Machine learning

Machine learning (ML) is the knowledge of a computer-based algorithm that take action through knowledge and acts as humans do (**wu2013skull**). Furthermore, it improves their learning over time in an independent way, by providing the data-set where the machine learning algorithms can learn from (**mikolov2010recurrent**). We have three machine learning techniques such as supervised, unsupervised and semi-supervised learning. In a supervised learning procedure, the data-set must be labeled as we know what output should be expected, like linear regression algorithm or support vector regression. On the other hand, in the unsupervised learning. It creates its own target value like clustering or Knns. semi-supervised learning method are mixed between both supervised and unsupervised algorithms.

1.1.4 Deep learning

The concept of deep learning (DL) is under the umbrella of ML field and term deep learning when the neural network have a lot of hidden layers and every layer had

its own function. This technique try to simulate how the human brain processes data. DL is used in the field of computer vision as it produce a very good result, for example there are DL models that are used for object detection, object classification and even depth estimation. its is also used in other fields than Computer Vision.

1.1.5 Problem and limitation

Scene description oppose a big challenge right now as it is composed of a multiple models working together, and it is hard to make robust system to face this challenge. three models will be used to face this challenge. object detection model, depth estimation model and SVR model that will predict the distances using the values from the depth model.

1.2 Objectives

Our aim in this thesis is to create an intelligent assistant system based on neural network models that will have the ability to understanding a scene and guiding VII to independently walk indoor or outdoor in order to aid visually impaired individual in their everyday life. The proposed model will be accurate, and robust to illumination and lighting changes in a scene. The proposed will a light model to be deployed on a device with limit memory, e.g., Arduino chips or smartphones. we will follow the next steps:

- Choose a reliable deep learning model for object detection and classification
- Choose a model for depth estimation to estimate the depth of each object classified from the previous model.
- Collect our own data-set to check the depth estimation model accuracy and get absolute distance.
- Train a machine learning model to convert the mean intensity pixel from the depth estimation model with the true distance.

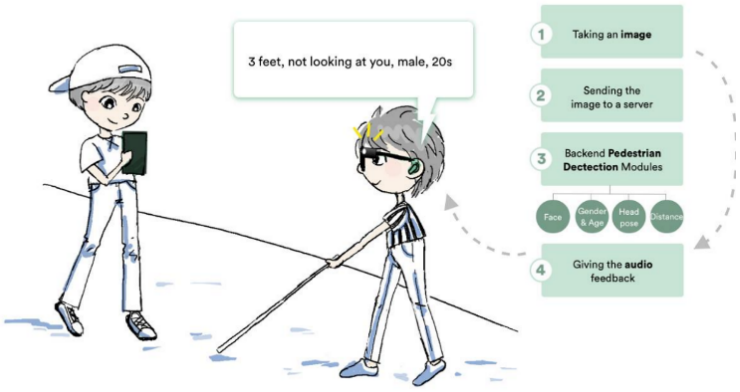

FIGURE 1.1: Simple example For our objective

1.3 Publications

Conference

1. Absolute distance prediction based on deep learning object detection and monocular depth estimation models Armin MASOUMIANa,1, David G.F. MAREI, Saddam ABDULWAHAB, Julian CRISTIANO, Domenec PUIG and Hatem A. RASHWAN. 23rd International Conference of the Catalan Association for Artificial Intelligence

1.4 Thesis organization

This thesis is outlined as follows:

In Chapter 1, provided an overview of the technologies that will be used.

In Chapter 2, discuss the related work.

In Chapter 3, we discuss the methodology

In Chapter 4,Experimental result of our system.

in Chapter 5, Conclusion and future work

Chapter 2

Related works

2.1 Object Detection Techniques

2.1.1 YOLO

In computer vision, the challenge of object detection need to be considered as it is very common due to the dependency of the shapes and textures of the objects where it needs to be detected. The detection techniques also help to get more information about the object, like the probability if the object detect is right or wrong (e.g., human detected and 70 percent it's a female). object detection is seen in various application like Optical Character Recognition, Tracking vehicles or even identifying iris code. Nowadays, deep learning have been applied in the entire area of computer vision and right now one of the state of the art for object detection is You Only Look Once (YOLO) Redmon et al., 2015. the Yolo models is always updating as it started with yolov1 in 2016, and now we have yolov5 in 2021. We are using YOLOv5 small as our approach as it is a light version. in addition, it produces the needed result. in the next figure, we will show the yolov4 performance with other state of the arts object detection models.

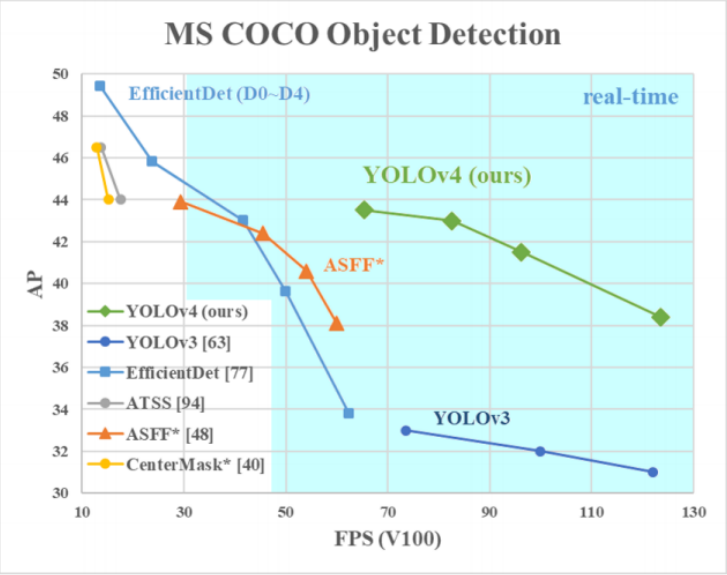

FIGURE 2.1: YOLOv4 and other state-of-the-art object detectors. YOLOv4 runs twice faster than EfficientDet with comparable performance. Improves YOLOv3's AP and FPS by 10 percent and 12 percent, respectively (Bochkovskiy, Wang, and Liao, 2020).

next this graph show the yoloV5 performance and all its versions

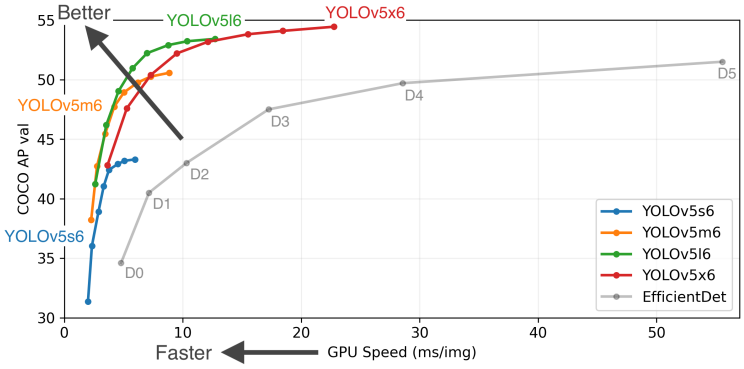

FIGURE 2.2: YOLOv5 different versions performance (Jocher et al., 2021a).

the yolo architecture have three main parts. The first part is Backbone CSPDarknet. The second part is Neck:PANet and finally Head which is the yolo later.(Xu et al., 2021)

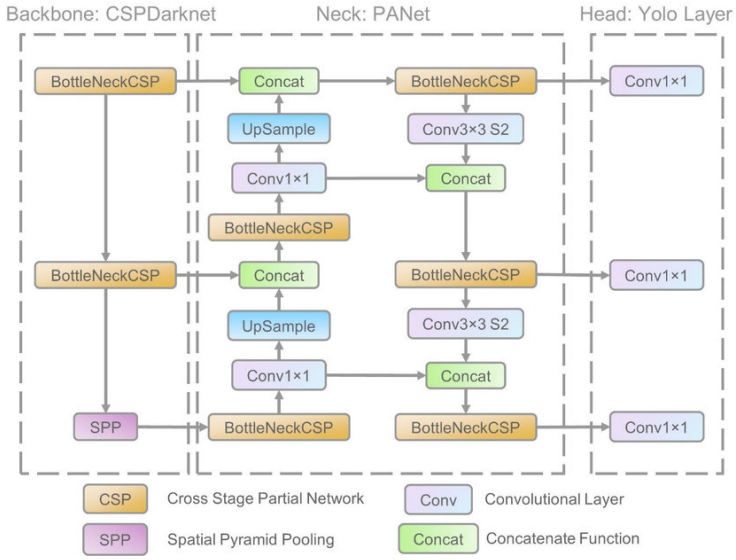

FIGURE 2.3: YOLO Architecture (Xu et al., 2021).

Backbone Networks

A backbone network works as feature extractor for object detection, which uses images as input and outputs feature maps of the identical input image. The CSP models are based on the densenet which is designed to connect layer in the cnns. the CSP also address duplicate gradient problems in large convnets problems which is extremely important to the yolo family as it reduces the parameters and the flops which result lighter and faster model. in the next figure, we can see and example of backbone model,

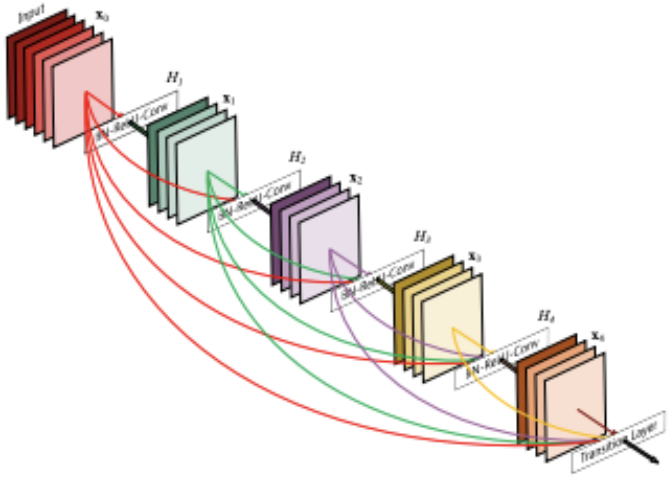

FIGURE 2.4: A 5-layer dense block with a growth rate of k = 4. Each layer takes all preceding feature-maps as input. (Huang et al., 2016).

PA-Net Neck

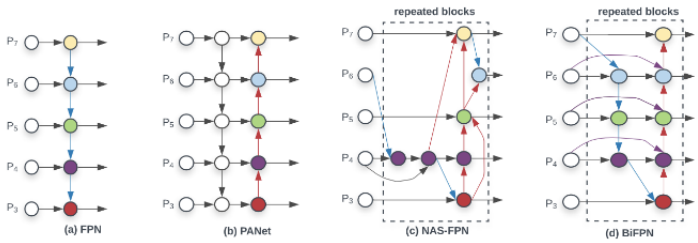

FIGURE 2.5: Pa$_n$et(*Tan*, *Pang*, *AndLe*, 2019).

Each one of the Pi above represents a feature layer in the CSP backbone. The above figure is a product from research done by Google Brain on the EfficientDet object detection architecture. so the neck is a series of layer mix to aggregate the image features and pass it to them for prediction.

Model Head

The model head is mainly used for the final detection part. it applied the boxes of each object and generate the probability of the object score as we will see in the next figure

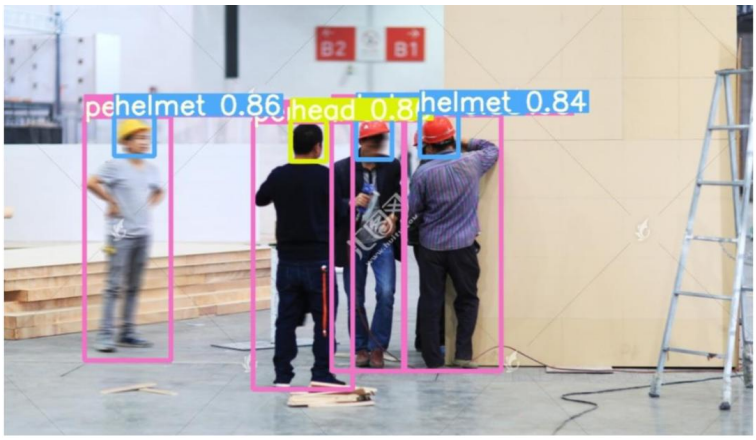

FIGURE 2.6: yolo tested image

2.1.2 R-CNN

Before Yolo, Region Based Convolutional Neural Networks was once the state of the art. it consists of three different modules. The first module generates independent regional proposal which define the group of candidates for the next detection step.The next module is a large convolutional neural network (CNN) that extracts a fixed-length feature vector from each part.The last module is a set of class specific linear SVM that gives the prediction of the object (Girshick et al., 2013

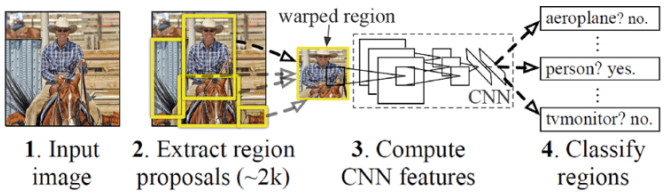

FIGURE 2.7: R-CNN Modules (Girshick et al., 2013)

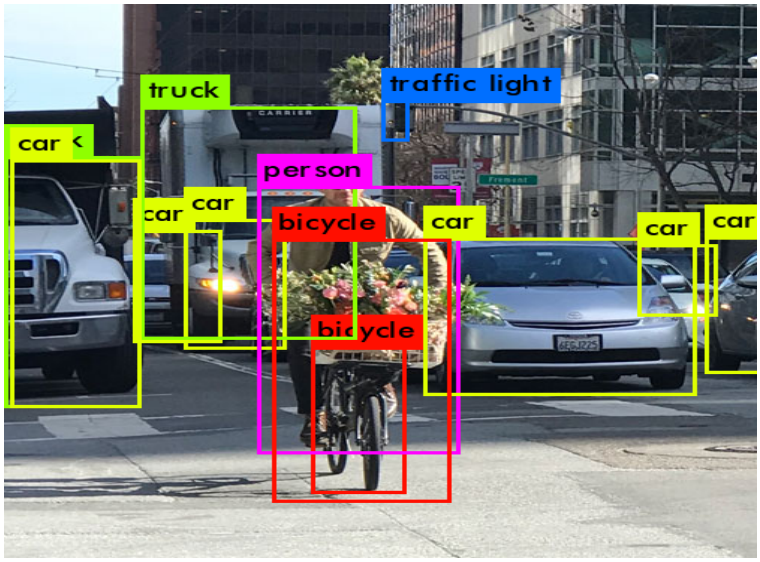

FIGURE 2.8: R-CNN Example(Ren et al., 2015)

2.2 Depth Estimation

Depth estimation is one of the most challenging computer vision tasks. as its goal, it to replace the sensors by trying to estimate the depth of the scene from 2-d images. the input is normal, an RGB image and the output is a depth image.

FIGURE 2.9: Depth disparity from kitti data-set (Geiger et al., 2013).

depth estimation also has multiple applications like smoothing blurred parts of an image, self-driving cars or grasping in robotics. there is a lot of depth estimation models that works for different scenarios. there is monocular depth estimation which we are using as approach which estimate the depth from one image. Also, there is stero depth estimation models that's generates a disparity map which is more accurate but requires supervised learning. We will be using Midas depth estimation model as this model was trained in multiple datasets, and it doesn't need high gpu.

2.2.1 Supervised Depth Estimation

Supervised depth estimation help decrease the innate difficulty as we can see the relationship between the colored images. some various approaches of supervised depth estimation is end to end sampling, optical flow, transfer learning and combining local prediction (Howard et al., 2017)

2.2.2 Unsupervised Depth estimation

Unsupervised methods are used to avoid aforementioned problems for training models. Only original images and pre-trained models such as DenseNet is needed. Regarding unsupervised methods, various approaches for depth estimation have been proposed, such as generative adversarial networks (Pilzer et al., 2018), temporal information and separate pose networks (Zhou et al., 2017).

2.2.3 Midas Model for depth estimation

Midas compute relative inverse depth from an image (Ranftl, Bochkovskiy, and Koltun, 2021). There are also multiple Midas models for each use case, from small to big. the small models are light and fast, and big models have better accuracy. all the

midas model have been trained in six different data sets to insure multi-objective optimization to ensure high quality of input ranges for the duration of 6 months with multiple GPUS.

Model	DIW, WHDR	Eth3d, AbsRel	Sintel, AbsRel	Kitti, δ>1.25	NyuDepthV2, δ>1.25	TUM, δ>1.25	Speed, FPS
Small models:							iPhone 11
MiDaS v2 small	**0.1248**	0.1550	**0.3300**	**21.81**	15.73	17.00	0.6
MiDaS v2.1 small URL	0.1344	**0.1344**	0.3370	29.27	**13.43**	**14.53**	30
Big models:							GPU RTX 3090
MiDaS v2 large URL	0.1246	0.1290	0.3270	23.90	9.55	14.29	51
MiDaS v2.1 large URL	0.1295	0.1155	0.3285	16.08	8.71	12.51	51
MiDaS v3.0 DPT-Hybrid URL	0.1106	0.0934	0.2741	11.56	8.69	10.89	46
MiDaS v3.0 DPT-Large URL	**0.1082**	**0.0888**	**0.2697**	**8.46**	**8.32**	**9.97**	47

FIGURE 2.10: Zero-shot error (the lower — the better) and speed (FPS) (Ranftl, Bochkovskiy, and Koltun, 2021).

2.2.4 zero-shot cross-dataset transfer

To test the midas performance, a zero to shot cross protocol was used. Which means that the model was trained on a certain dataset and then the testing was on a different dataset. this initiation was used because it is believed that this is a more faithful proxy to the real world.

2.3 Linear Regression

linear regression is a machine learning algorithm that depends on the supervised learning. It is mainly used to find relationship between variable and forecasting.Different regression models differ on the relation between the inputs and the outputs. In most linear regression models, the objective is to minimize the sum of squared errors.

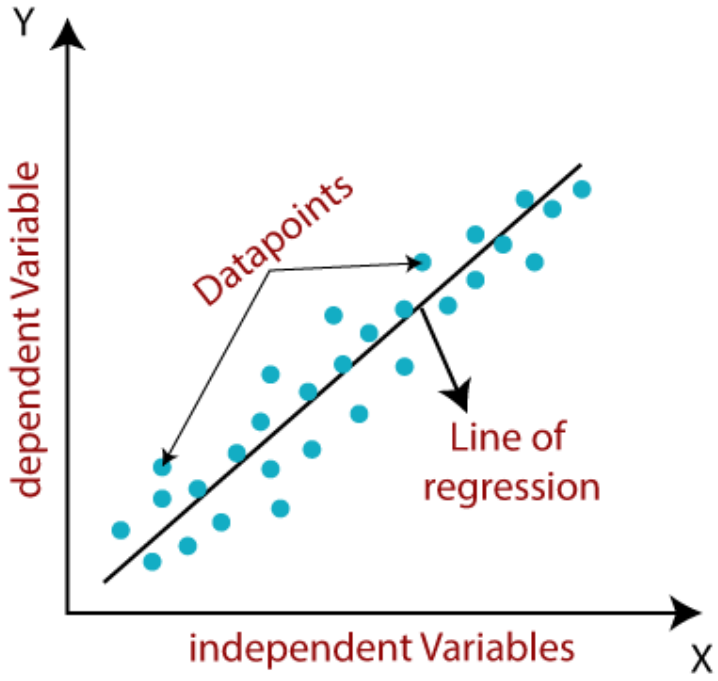

FIGURE 2.11: Linear Regression example for dependent and independent Variables(Bonaccorso, 2017)

to have an accurate linear regression model we must have big data-set to train from and this is not possible for our scenario. That's why, for our approach, we will use a support vector regression model to predict the distance of the object from the pixels value resulted from the depth estimation model.SVR advantage is that it doesn't have the best fit line. it has a hyperplane that has a maximum and minimum number of points. Support vector regression was chosen as we don't have a big data collected for the training period. Its biggest advantage is that it let us define the range of acceptable error to find the probate line.

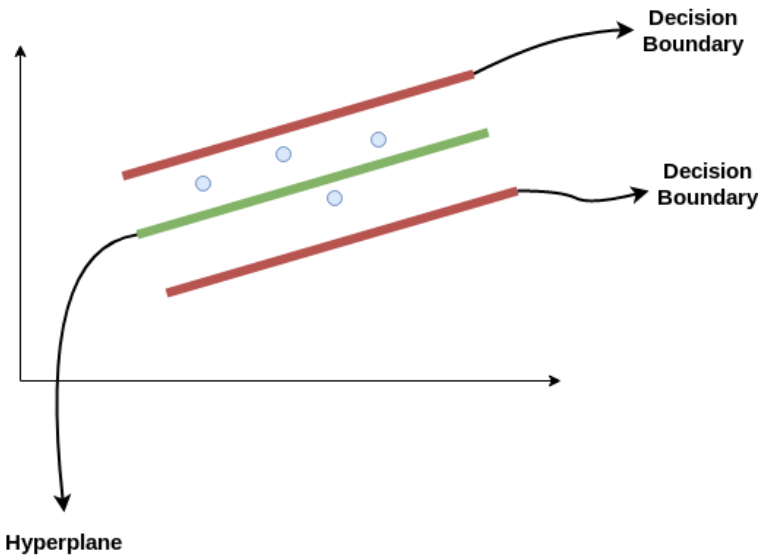

FIGURE 2.12: SVR Example (Cortes and Vapnik, 1995)

SVR tries to fit the best line within a threshold value that we use. This threshold value is the distance between the hyperplane and boundary line. and the advantage of SVR are the following

- Easily updated
- Robust to outliers
- Fast With high prediction accuracy
- Easy Implementation

Chapter 3

Methodology

As shown in Chapter 2, we have reviewed the deep learning method for object detection: YOLOv5 in order to detect the objects in the image. We choose the light version yolov5s. Next, we talked about depth estimation, and we choose midas model from pytorch because of its flexibility as it was trained on 6 data set, and it also has a light version. Our goal is to retrieve the mean pixels of the objects detected in yolo from the output of the depth estimation model. Then for each object we know the ground truth distance.We have two diff rent approaches to calculate the distance. the first mathematical approach is through a mathematical equation.In the second approach, We will train the SVR with mean pixel feature as an input and the ground truth distance as an output. So the SVR model can predict the distance from the mean pixels of each object.

3.1 Object detection based on YOLO Model

We used YoloV5S which is the lighter version of yolo family for the object detection. it's already trained for months on multiple GPUS, so we won't add any training, as the coco data-set that was used for training serve the same purpose we want.

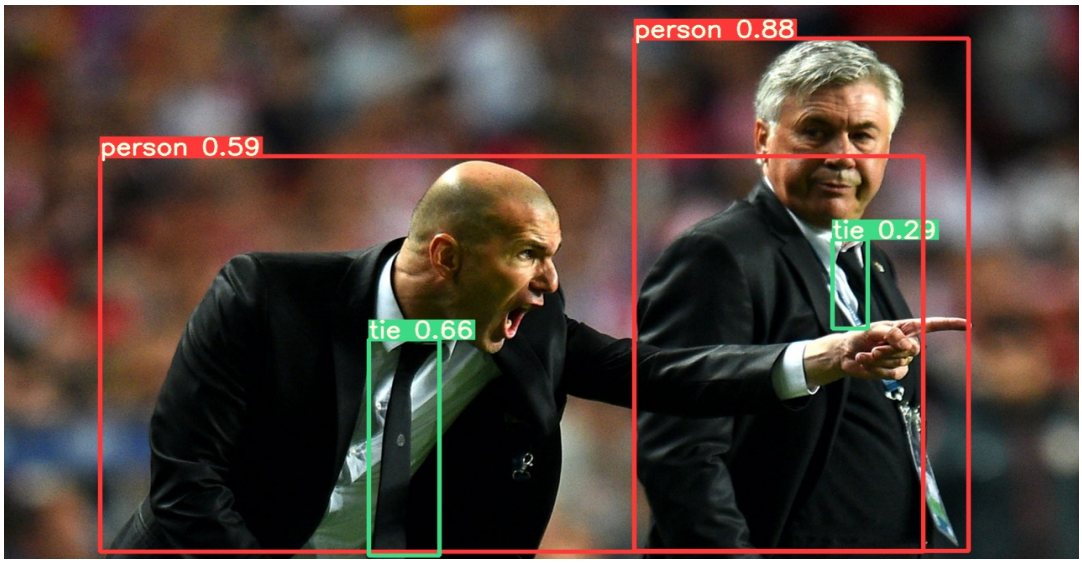

FIGURE 3.1: YOLO Example(Jocher et al., 2021b)

In summary, the model architecture of YOLOv5 is very close to YOLOv4. Besides, it derives most of performance improvement compared to object detection methods. the YOLOv5 model is a fast compact object detection model that is very performant relative to its size and it has been steadily improving as shown in figure**??**.

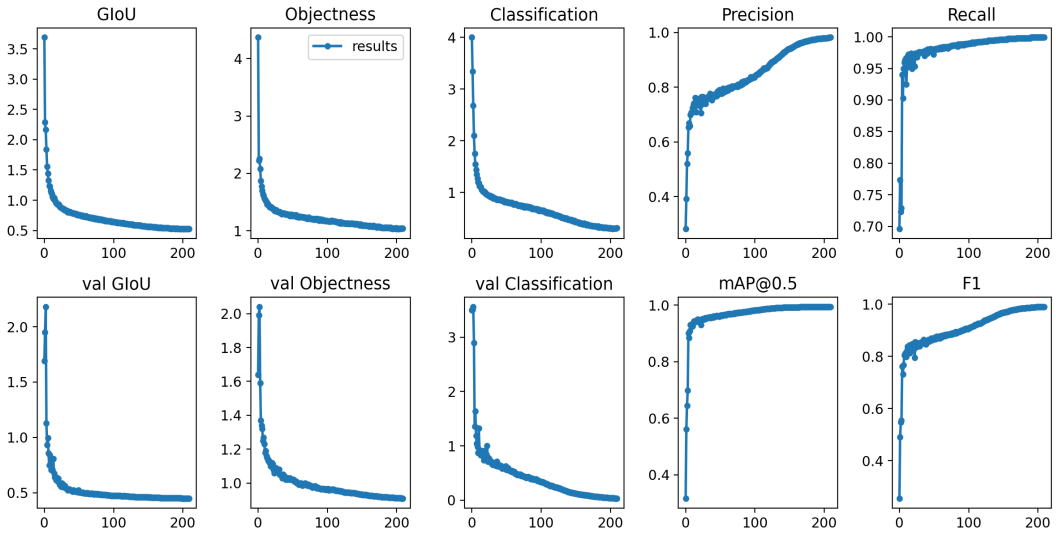

FIGURE 3.2: YOLO Performance table (Jocher et al., 2021b)

3.2 Depth Estimation based on the MiDas Model

MiDaS calculates the relative inverse depth from one image.We Chose This model because it was trained on 6 distinct datasets using multi-objective optimization to ensure high quality on a wide range of inputs.CRanftl, Bochkovskiy, and Koltun, 2021. MiDas introduces a new loss function that absorbs the diversities between different datasets used in the training stage, thereby eliminating compatibility issues and allowing multiple data sets to be used for training simultaneously. Midas advanced the state of the art in monocular depth estimation. MiDas was been evaluated the robustness and generality of models via zero-shot cross-dataset transfer that can help in the deployment of monocular depth models in practical applications.

FIGURE 3.3: Midas Example (Ranftl, Bochkovskiy, and Koltun, 2021

3.3 Absolute distance estimation using Support Vector Regression

Our approach to get the distance from a single image. we used two parallel deep networks: one for object detection (YOLO) and the other for depth estimation(Midas). The predicted depth will be extracted from Midas. In turn, with YOLOv5, the objects inside the image will be localized and classified. In addition, the localization of each object defined by bounded boxes will be detected on the estimated depth image. Finally, the relevant distance of an object will be calculated by the median estimated distance of all pixels inside the defined bounded box by passing it through the trained support vector machine model. it was straight forward. As we imported the svr model for sic-learn library. We have one input feature (mean intensity of pixel of an object) and one output (the hand measured distance of this object from the camera). As for the model parameter, we only change the kernel type to linear instead of the default rbf. we tried all the possible kernels, but the linear produced the best results.

3.4 Absolute distance estimation using Mathematical approach

here we also used two parallel deep networks: one for object detection (YOLO) and the other for depth estimation(Midas). The predicted depth will be extracted from Midas. In turn, with YOLOv5, the objects inside the image will be localized and classified. In addition, the localization of each object defined by bounded boxes will be detected on the estimated depth image. Finally, the relevant distance of an object will be calculated by the median estimated distance of all pixels inside the defined bounded box by passing it through the following quadratic function equation (Taha

and Jizat, 2012)

$$Y = (c.0 + c.1X + c.2X^2)H$$

where, c0, c1, c2 coefficients can be obtained using the least square equations, h is the camera height, and X is the relevant distance from the object to the beginning of the camera's field of view (Taha and Jizat, 2012)

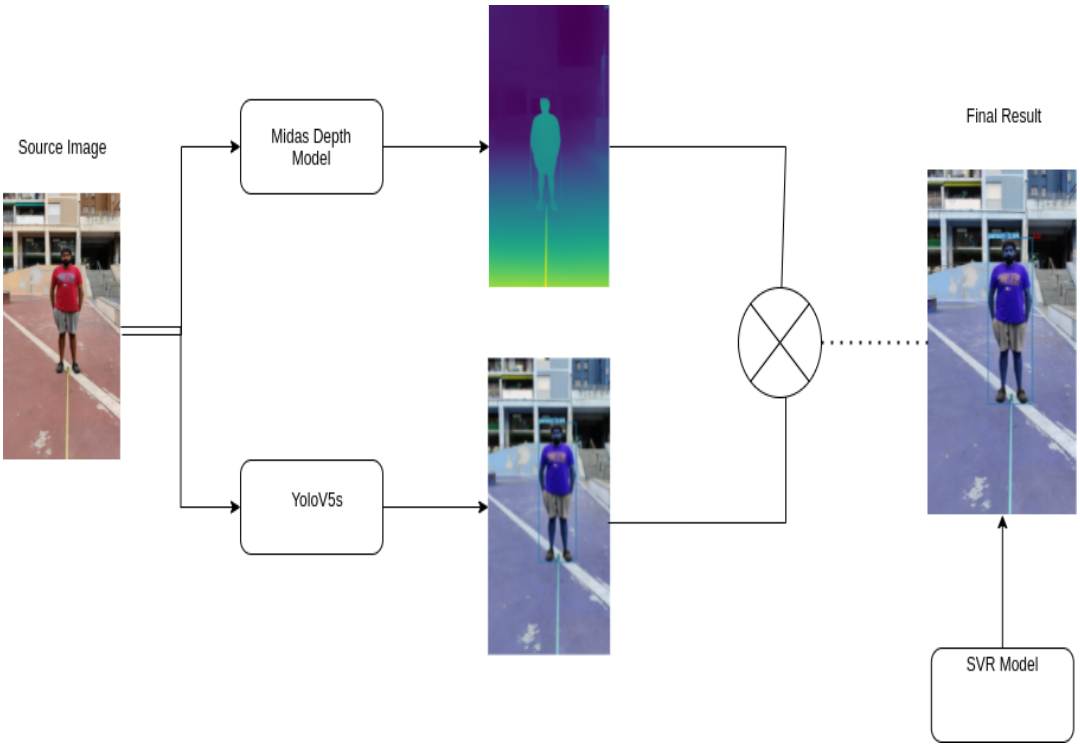

FIGURE 3.4: The proposed workflow for SVR model approach.

Chapter 4

Experimental Result

4.1 Data-set Preparation

For our data-set, we didn't need to train the yolo or the depth model. However, when we tried to calculate the distance of the object, we prepared our own private data-set that contains 100 images 777 x 1350 with a hand-held camera. Monocular RGB camera was mounted on a static stand and the absolute distance of each object was manually measured from the camera. The absolute distance from the camera and objects have manually been defined of all objects in scenes.During collecting the data-set, we imitated potential static obstacles on the front of the camera. These obstacles located in different distances from the camera test-stand.

FIGURE 4.1: Collected data-set. left image is 4.4 meters and right image is 5 meters.(Our own data-set)

4.2 Experimental Results

We experimented using the two approaches mentioned before (the support Vector Regression and the mathematical approach) and we will show the results of each approach in the next part. but first lets see Midas model depth estimation on our data-set.

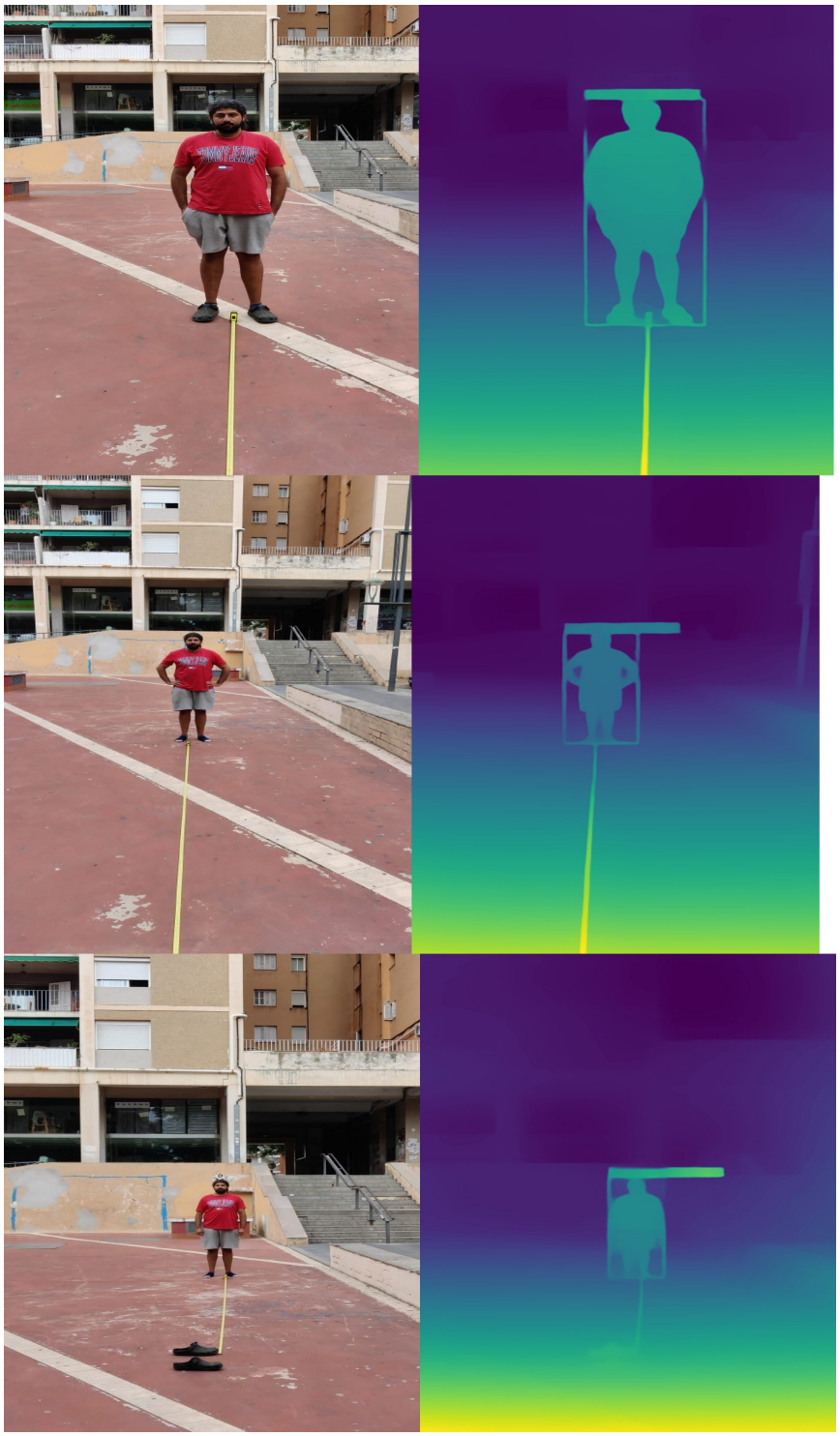

FIGURE 4.2: Midas depth test (our own data-set)

4.2.1 Support Vector Regression result

Next we will show the output result images from the Svr approach, the first row image's real value was 3.4 meters. the model prediction on the right is three meters.The Second image is 6 meters far, and the model predicted 6 meters. The last row was 11 meters far and the model predicted 10.9 meter.

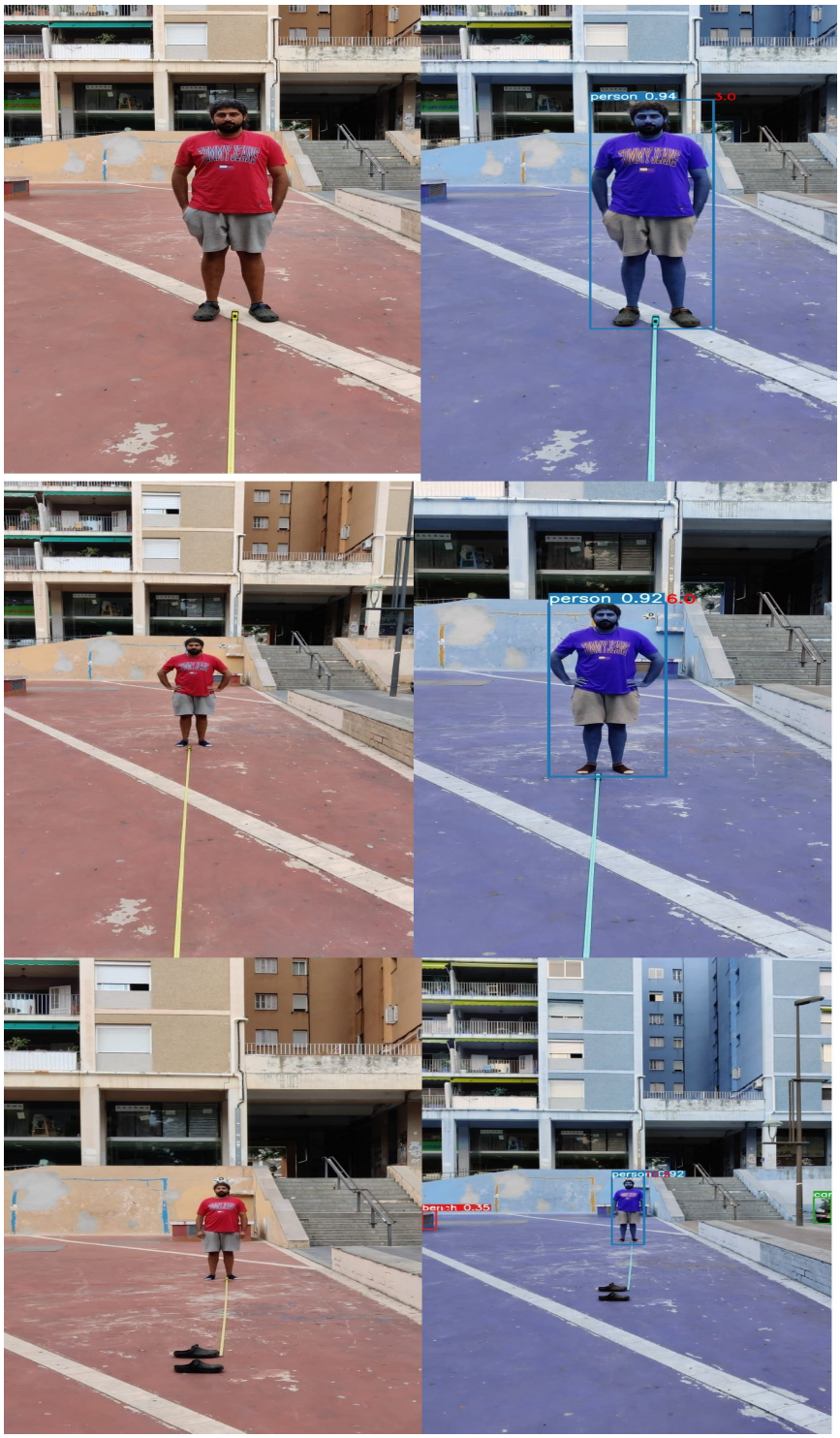

FIGURE 4.3: SVR model prediction images(Our own data-set

Absolute distance (m)	Predicted distance (m)	Error (M)
3.4	3	0.4
6	6	0
11	10.9	0.1
9	8	1
7	5.5	1.1

FIGURE 4.4: SVR Error Table

From this table we can see that the error increase in the middle values. However, the accuracy is big when the object is too far or too close

4.2.2 Mathematical approach

Next we will show the result from the mathematical approach and demonstrates the qualitative results of the proposed framework, and it shows some examples of our own private data-set including the original images, estimated depth images

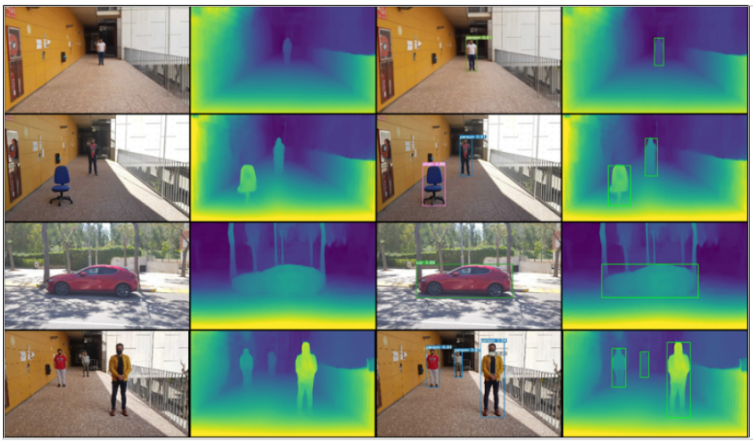

FIGURE 4.5: Mathematical approach (our own data-set)

Absolute distance (m)	Predicted distance (m)	Error (m)
11.2	10.91	0.29
3.5	3.45	0.05
8.0	8.09	0.09
10.1	9.83	0.27
8.0	8.13	0.13
12.0	11.69	0.31
4.0	3.88	0.12

FIGURE 4.6: Mathematical approach Error Table

4.3 Experiment Limitation

The biggest limitation we faced is the lack of data-set to train our SVR as we need to create our own data-set, and it was not big enough. Another limitation was my lack of programming skills in python and the idea of creating a virtual environment, as it took me sometime to understand it. If we had a stronger GPU, maybe we could've chosen heavier models for the yolo and midas.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

Scene description is a challenging task, and it will always remain a challenging task. But we hoped to have a different approach to tackle this problem with the approach mentioned before. I believe there is a lot to improve if we had better data-set and more time. even the pretrained models we used as yolo and midas are still getting updates regularly on their GitHub repos. in the end, we reached a satisfactory result, and we hope this result may help other researchers in the future to tackle this problem in a more efficient way.

5.2 Future work

In the future, we hope to deploy this project in the market to aid the visually impaired individuals. to convert the result obtained for each object as it class and distance to a voice track so the visually impaired can hear it and aid him to understand the scene in front of him. We also hope that a better data-set source to be collected for the object distance from the camera, as we believe this will help to improve the result a lot (e.g.. data-set that have image and the info about every object in the image as the class type and distance from the camera.)

Bibliography

Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao (2020). *YOLOv4: Optimal Speed and Accuracy of Object Detection*. eprint: arXiv:2004.10934.

Bonaccorso, Giuseppe (2017). *Machine Learning Algorithms: A Reference Guide to Popular Algorithms for Data Science and Machine Learning*. Packt Publishing. ISBN: 1785889621.

Cortes, Corinna and Vladimir Vapnik (1995). "Support-vector networks". In: *Machine learning* 20.3, pp. 273–297.

Geiger, Andreas et al. (2013). "Vision meets Robotics: The KITTI Dataset". In: *International Journal of Robotics Research (IJRR)*.

Girshick, Ross et al. (2013). *Rich feature hierarchies for accurate object detection and semantic segmentation*. eprint: arXiv:1311.2524.

Howard, Andrew G. et al. (2017). *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*. eprint: arXiv:1704.04861.

Huang, Gao et al. (2016). *Densely Connected Convolutional Networks*. eprint: arXiv:1608.06993.

Jocher, Glenn et al. (Apr. 2021a). *ultralytics/yolov5: v5.0 - YOLOv5-P6 1280 models, AWS, Supervise.ly and YouTube integrations*. Version v5.0. DOI: 10.5281/zenodo.4679653. URL: https://doi.org/10.5281/zenodo.4679653.

Jocher, Glenn et al. (2021b). *ultralytics/yolov5: v5.0 - YOLOv5-P6 1280 models, AWS, Supervise.ly and YouTube integrations*. DOI: 10.5281/ZENODO.4679653. URL: https://zenodo.org/record/4679653.

Pilzer, Andrea et al. (2018). *Unsupervised Adversarial Depth Estimation using Cycled Generative Networks*. eprint: arXiv:1807.10915.

Ranftl, René, Alexey Bochkovskiy, and Vladlen Koltun (2021). "Vision Transformers for Dense Prediction". In: *ArXiv preprint*.

Redmon, Joseph et al. (2015). *You Only Look Once: Unified, Real-Time Object Detection*. eprint: arXiv:1506.02640.

Ren, Shaoqing et al. (2015). "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." In: *NIPS*. Ed. by Corinna Cortes et al., pp. 91–99. URL: http://dblp.uni-trier.de/db/conf/nips/nips2015.html#RenHGS15.

Steinmetz, Jaimie D et al. (Feb. 2021). "Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study". In: *The Lancet Global Health* 9.2, e144–e160. DOI: 10.1016/s2214-109x(20)30489-7. URL: https://doi.org/10.1016/s2214-109x(20)30489-7.

Taha, Zahari and Jessnor Arif Mat Jizat (Feb. 2012). "A Comparison of Two Approaches for Collision Avoidance of an Automated Guided Vehicle Using Monocular Vision". In: *Innovation in Materials Science and Emerging Technology*. Vol. 145. Applied Mechanics and Materials. Trans Tech Publications Ltd, pp. 547–551. DOI: 10.4028/www.scientific.net/AMM.145.547.

Tan, Mingxing, Ruoming Pang, and Quoc V. Le (2019). "EfficientDet: Scalable and Efficient Object Detection". In: eprint: arXiv:1911.09070.

WHO. *Blindness and vision impairment*. https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment. (Accessed on 09/02/2021).

Xu, Renjie et al. (Feb. 2021). "A Forest Fire Detection System Based on Ensemble Learning". In: *Forests* 12, p. 217. DOI: 10.3390/f12020217.

Zhou, Tinghui et al. (2017). *Unsupervised Learning of Depth and Ego-Motion from Video*. eprint: arXiv:1704.07813.