Smart Stick: An Aid To Visually Impaired

Dr Seema Kolkur
Computer Department
Thadomal Shahani Engineering College
Mumbai,India
seema.kolkur@thadomal.org

Remston Dsa
Computer Department
Thadomal Shahani Engineering College
Mumbai,India
remstondsa7@gmail.com

Shubham Dwivedi
Computer Department
Thadomal Shahani Engineering College
Mumbai,India
shubhamdwivedi1210@gmail.com

Medha Gavkar
Computer Department
Thadomal Shahani Engineering College
Mumbai,India
medhagavkar55@gmail.com

Abstract*— Vision, auditory, olfactory, taste, and touch are the five senses through which human beings collect information of the surrounding environment. Out of the five senses, vision accounts for almost 84% of the total information. Hence, an individual with visual impairment suffers ample challenges and it becomes difficult to deal with it in the world outside. Visual impairment describes any kind of vision loss, either total vision loss or partial vision loss. Visual Impairment can be curtailed through an optic surgery or with the use of technology. A surgery incurs huge medical expenses and involves health risks. This research paper proposes 'Smart Stick', a cost-effective solution that aims to solve the common problems faced by visually challenged individuals. A 'Smart Stick' is an integration of Convolutional Neural Network, IoT, Artificial Intelligence, Computer Vision, and Image Processing. Some of the basic functionalities that are achieved by 'Smart Stick' are object detection, face recognition, weather updates, current location, alert systems, speech to text, and navigation. Thus, 'Smart Stick' can be a cost-effective solution and may serve as the potential aid for visually challenged individuals.

Keywords*— Smart Stick, Visual Impairment, Convolutional Neural Network, IoT, Artificial Intelligence, Computer Vision, Image Processing, Technology, Cost-Effective Solution (key words***)**

1. INTRODUCTION (*HEADING 1*)

According to a report by the World Health Organization (WHO) and the International Agency for Prevention of Blindness (IAPB), nearly 285 million people around the world are visually impaired out of which 39 million people are completely blind [1]. According to the experts, global blindness is projected to triple in numbers by 2050 due to the rapidly aging global population, widespread chronic eye diseases, pollution increase, and genetic disorders [2].

It is not a herculean task to comprehend the number of challenges that are faced and inflicted by visually challenged individuals. Vision impairment impacts decision-making in one's daily life and the consequences are unfortunately dealt with. Thus they have a constant need for assistance in their daily life. Since they are distracted by the obstacles, they tend to rely on Physical travel assistive aids like guide dogs and white canes but they have their shortcomings. These limitations can be undertaken by the use of technology. This research paper proposes a cost-effective solution that aims to solve the basic problems faced by visually impaired individuals called 'Smart Stick'.

'Smart Stick' achieves 6 basic functionalities namely

1. Object detection
2. Face recognition
3. Alert systems
4. Location updates
5. Weather updates
6. Speech to text

All the data is provided by the user in the speech format through the microphone. The speech is then converted into text format using the Speech-to-text module. According to the data provided by the user, appropriate computations will take place, and the relevant result is provided as the output in the speech format through the earphones using the text-to-speech module. All the computations are enumerated on the server, hence the huge computational cost is saved and latency is minimized. 'Smart Stick' tries to use all the available resources more efficiently to achieve the proposed functionalities. For real-time object detection and face recognition, the system accesses the user's mobile camera to collect image frames, and the data collected is computed on google colab using the YOLOv4 algorithm. For emergency alerts, the system sends SMS, and Mails using Twilio API and SMTP protocol respectively. Current location and Weather updates are provided through the integration of the GPS module. Speech-to-text is enabled using the DeepFace algorithm via the user's microphone and the output is given in speech format using text-to-speech module via the user's earphones. Finally, for temperature sensing the system uses the LM35 sensor module that provides input in Kelvins(^{o}K). Thus, 'Smart Stick' can be a cost-effective aid for the blind.

2. LITERATURE REVIEW

Obstacle detection system for visually impaired people is a sought-out project which comes in a lot of variations. The system proposed in [3] consists of a sensor unit, vibration, and voice unit, artificial vision unit, and GPS unit. It uses ultrasonic sensors and cameras to detect obstacles. When it comes across an obstacle, the voice circuit will get activated and the user gets alerted by providing a certain type of voice. In the model proposed in [4], a photo is taken through the camera in the system and is sent to the server for processing. This image is then processed using the Mask-RCNN algorithm. The output of what objects are present is read

aloud to the user. Smart rehabilitative shoes and spectacles proposed in [5] facilitate safe navigation and mobility of blind individuals. Each shoe is mounted with ultrasonic transducers to detect objects at different heights. The spectacles are instrumented with a pair of ultrasonic transducers mounted centrally above the bridge, and with a buzzer at one of the temples [5,6].

R-CNN, SSD, and YOLO are some of the algorithms that are most popularly used for object detection and recognition. According to studies conducted in [7], Yolov3 is faster than both SSD and R-CNN when it comes to speed. It also provides more accurate results than SSD and R-CNN. For object detection and recognition, a YOLOv3 based detector was proposed as the most effective [7]. The system proposed in this paper uses the Yolov4 algorithm. It applies a single neural network to the complete image. This network divides the image into regions and predicts bounding boxes and probabilities for every region [8]. These bounding boxes are weighted by the predicted probabilities [9].

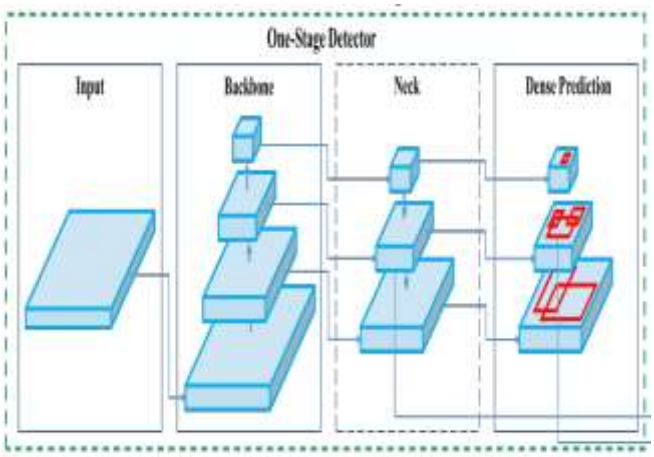
A lot of processing is required in such systems and this may cause a delay in getting the output if the processing power of the device is not very high. This is not at all a favorable situation as the delay in output may lead to some mishap. Hence, there is a requirement for a GPU that will support the high processing need for the system. But using a GPU will make this system cost-intensive. Our system uses a virtual GPU for all the processing in the system. This makes our system cost-effective.

Unlike other proposed systems which offer only the object detection and recognition feature, our system comes integrated with many other features. The additional features are getting the current location, receiving weather information, and sending emergency alerts to the user's emergency contact. The emergency alert will be sent in three forms. The first one will be through a call to the emergency contact of the user. The second way will be via email to the emergency contact. The email will contain the user's location. The last way will be through an SMS.

3. METHODS

3.1 YOLO V4 algorithm

YOLO (You Only Look Once) is a one-stage object detection algorithm. YOLO version 4 is faster and more accurate compared to YOLO version 3 as YOLO version 4 can be trained on a single 1080 Ti or 2080 Ti GPU with a mini-batch size[10]. YOLOv4 gives accurate results in real-time but it can be costly since it requires huge computation power. In order to avoid latency in the system, at least one GPU needs to be integrated into the system. In order to achieve a cost-effective solution, we use virtual GPUs in order to process huge computations. Fig 1 clearly shows the internal bifurcation of one stage detector algorithm. Here the one-stage detector consists of 4 subparts namely input, backbone, neck, and dense prediction(head). The input data passes through the three stages sequentially starting from Backbone to Neck and finally to the Head where the YOLOv4 algorithm is being applied.



- ***Backbone***- They are feature extractors like ResNet, DenseNet, VGG which are pre-trained with image classification datasets like ImageNet and tuned on the detection dataset. YOLO V4 uses CSPDarknet53 as a feature extractor

- ***Neck***- They are used to extract different feature maps of multiple stages with different scales of the backbone. The YOLO algorithm uses FPN, PANet, Bi-FPN as the neck.

- ***Head***- They involve the classification and regression of the bounding box(x,y,h,w) and compute the probability of the k class + 1 background class.

3.2 DARKNET -YOLO

Darknet is an open-source neural network framework designed in C and CUDA which supports CPU and GPU computations [11]. **Fig 2** shows the graph of FPS (Frames Per Second) vs AP (Accuracy Precision) and compares object detection algorithms like Center Mask*, ASFP*, ATTS, EfficientDet, YOLOv3, and YOLOv4 over MS COCO Dataset. This system uses Google Collab in order to access virtual GPUs for smooth processing and minimum latency.

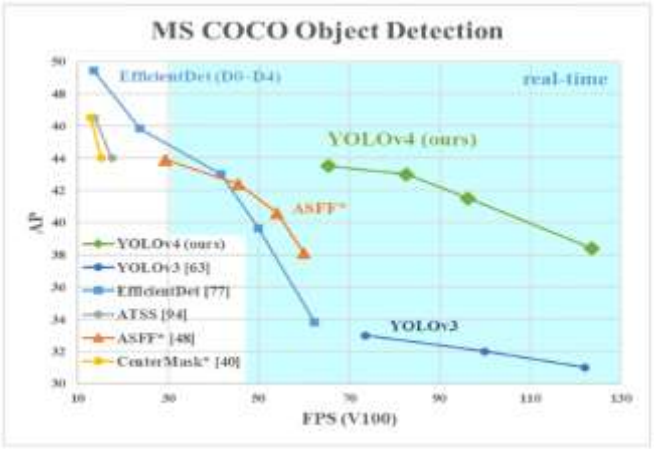


Fig 2 MS COCO Object Detection

3.3 ARCHITECTURE OF THE PROPOSED SYSTEM

Fig 3 depicts the architecture diagram of the complete proposed system. The system is governed by the user's speech and integrates both hardware and software. The system includes a physical body that connects to the cloud. The cloud is required to compute and process the inputs given to the system via API server calls. The Mobile camera is connected to the server via IP Webcam API that helps in object detection and recognition. For inbuilt SMS, voice calls, and text messages, the server is connected to the Twilio API which is imported from the python module.

A Simple Mail Transfer Protocol (SMTP) which is a built-in library in python is used for an automatic generation of an email. The user can request the current location and weather updates by simply making a GPS/ GNSS API request and Weather API request respectively. The output is then generated in the form of voice through the Text-to-Speech module. Thus, the voice proctoring is carried by the earphones connected to the system. For smooth processing, a single 1080Ti GPU is required.

Thus, the system is the combination of API and IoT integration. The huge computations are enumerated on the server using the virtual GPU using Google Colab. This avoids the latency in the system. The voice is thus proctered using the deep Face algorithm.

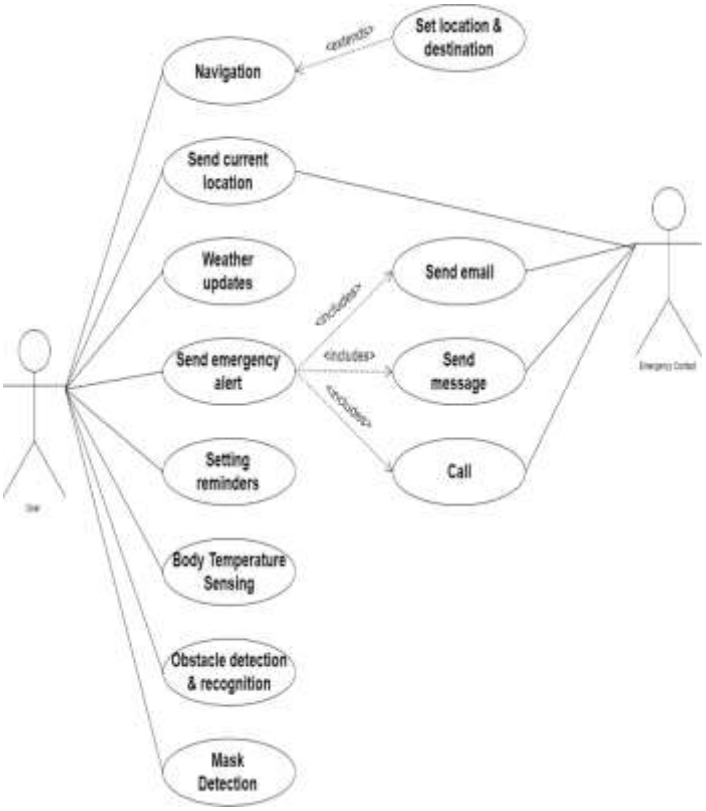


Fig 4 Use Case Diagram

Fig 4 describes the working of the system in the form of a Use Case Diagram. The user can simply give a voice command to request functions like Current location, Weather updates, and Object detection. The user can also give a voice command to set the destination and trigger an emergency alert either through call, text, or email to the registered emergency contact.

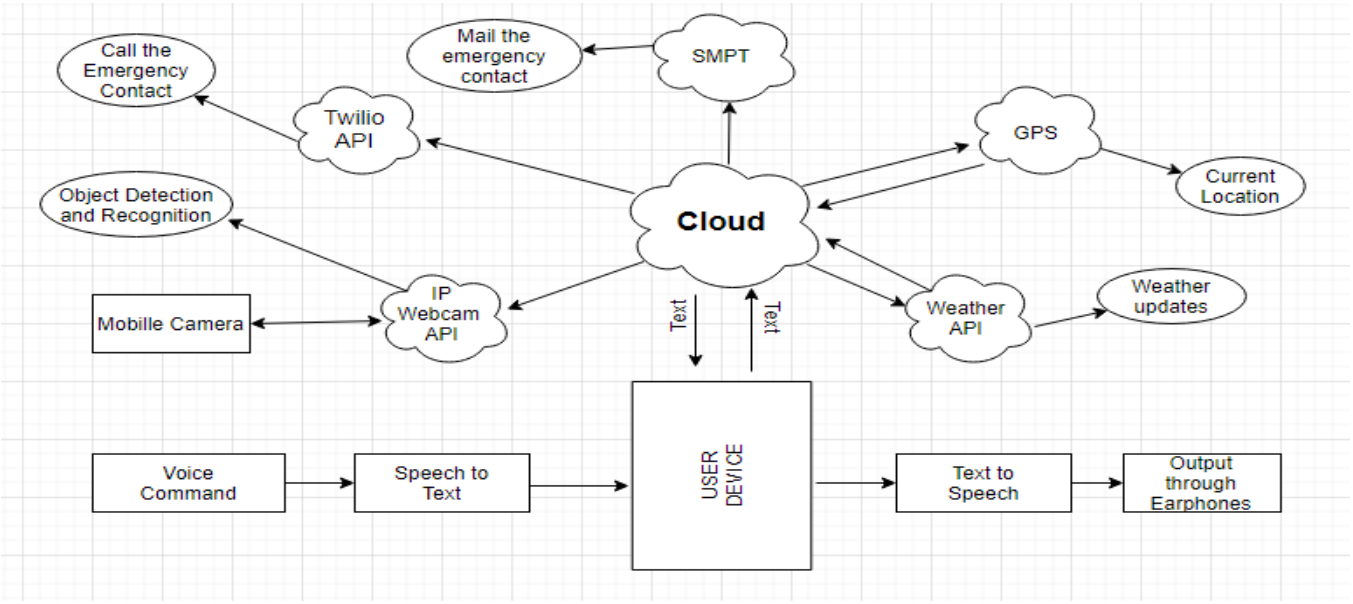


Fig 3 Architecture Diagram

4. RESULTS

4.1 OBJECT DETECTION AND RECOGNITION

Object Detection and Recognition is the primary and most essential module of this system. This module is responsible for the obstacle detection and recognition of the objects present in the path of our user using the YOLO V4 algorithm. The YOLO v4 algorithm consists of a multi-layer architecture and uses independent logistic classifiers and binary cross-entropy loss for class predictions and multilabel approach instead of softmax approach which allows classes to be more multiple and specific for individual bounding boxes.

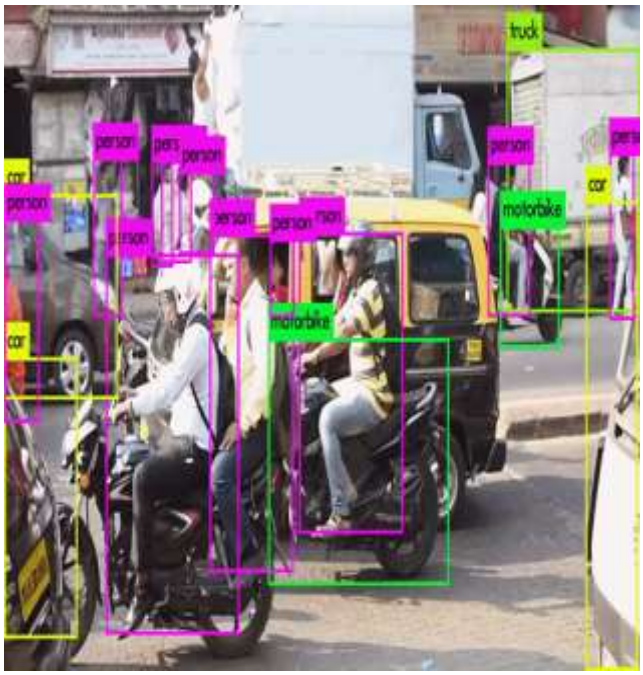


Fig 5 Object Detection and Recognition

The smartphone of the blind user acts as our source for the video input for this module. The data in the video format is transmitted from the user device to the server for the next step. After getting the input video, now we process this data for object detection and recognition using the YOLO V4 algorithm. The YOLOv4 algorithm generates bounding boxes which have confidence scores associated with it. In the first stage of processing the threshold parameter is set and all the boxes lower than the predefined threshold limit are excluded from further processing. The remaining boxes undergo non-maximum suppression which eliminates unnecessary unwanted bounding boxes. After this the default values for width (inpWidth) and height (inpHeight) for the input video feed are set. We set each of them to 416 for balanced results. You can also change both of them to 320 to get faster results or to 608 to get more accurate results. All of this helps our system in accurately detecting and recognizing objects in the camera frame of the live video feed. As seen in **Fig 5** the module successfully identifies a majority of the objects present in the frame. All objects of the same class or category are represented with a same colour bounding box whereas objects of different class or category are distinguished with a different coloured bounding box.

Table 1 : Comparison of detection speeds for Yolo v3 and Yolo v4

Detection	320x320	416x416	512x512
YoloV3 FPS	24.38	20.94	18.57
YoloV4 FPS	22.15	18.69	16.50

4.2 WEATHER UPDATE

```
The weather at this location is haze
The temperature is 29.0 degree celsius
The humidity is 65%
The wind speed is 12.96 km/hr
```

Fig 6 The output for Weather update

This module provides the user with classified and prioritized weather information with the intention of not overburdening the user with too much information. The weather data obtained in real-time and accurately. This module accesses the current user location with the help of the location module on the basis of which it fetches a detailed weather report JSON file from the weather API.
This JSON file consists of a lot of weather measuring parameters out of which a few are selected as seen in **Fig 6** and is reported to the user in audio format.

4.3 MASK DETECTION

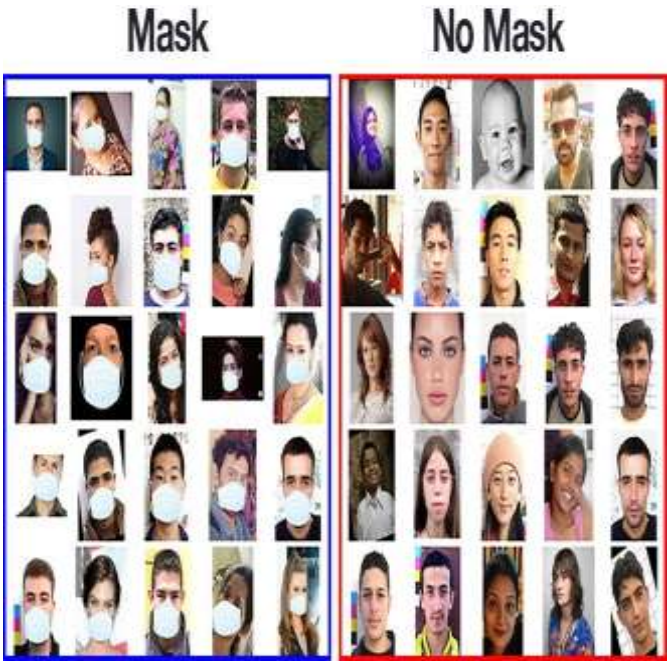


Fig 7 Mask Detection training dataset

This module is a face mask detection module implemented using Keras, OpenCV, and TensorFlow libraries. This module accesses the webcam of the user's device and checks whether the user has worn a mask or not. It has 2 convolutional layers. There are two convolution layers for data pre-processing and training purposes. **Fig 7** gives us an insight on the data pre-processing and training end. Then, we

flatten the convolutions and connect them to a dense layer of neurons. The final output layer consists of 2 neurons, with a mask, and without a mask. Each image is first passed through the classifier which gives us the region of interest (ROI) i.e. the face and then converted into a 100x100 image. This image is then passed to the trained CNN which will predict if the face has a mask on it or not.

4.4 LOCATION TRACKING

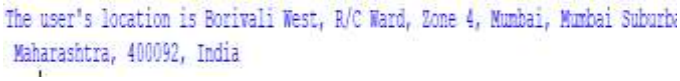


Fig 8 The output for Current Location

This is an essential module especially from the point of view of a visually impaired user. This module lets the user know its current live location at any instant of time on the basis of just a single voice command. This module initially with the assistance of the location API tracks the user's location in terms of latitudinal and longitudinal coordinates and then for user convenience converts the coordinates in address form and reports it to the user. This module also assists other modules in their functionalities as this module acts as an input source for the weather module and assists in the functioning of the emergency module.

4.5 EMERGENCY ALERT

This module as the name suggests is designed for unwanted and fatal emergency situations if the user encounters one our system ensures that the user has the assistance. As soon as this system is activated via voice command, this module puts multiple mechanisms like the SMTP protocol and Twilio messaging API into action. The SMTP protocol sends an emergency alert mail along with the user coordinates to the registered email id. The Twilio messaging API does the same task but sends the information to the registered mobile number

5. CONCLUSIONS

The proposed system has the potential to completely transform the lives of the visually impaired. This system provides visually impaired people with great features and functionalities like object detection and recognition, location tracking, alert system, and multiple voice-controlled functionalities. The most important distinguishing factor for this system is its requirements, this system only requires a smartphone to provide the user with the above functionality which makes this system not only efficient but also cost-effective. This system also eliminates the need of carrying a blind stick for our visually impaired users hence reducing the baggage.

6. REFERENCES

[1] World Health Organization. 2021. *Global data on visual impairment*. [online] Available at: <https://www.who.int/blindness/publications/globaldata/en/#:~:text=Globally%20the%20number%20of%20people,are%2082%25%20of%20all%20blind.> [Accessed 22 March 2021].

[2] SEE International. 2021. *Global Blindness Projected to Triple by 2050 - SEE International*. [online] Available at: <https://www.seeintl.org/blog/global-blindness-2050/#:~:text=SEEING%20WHAT%20THE%20FUTURE%20HOLDS,to%20115%20million%20by%202050.> [Accessed 22 March 2021].

[3] Shruti Dambgare , Prof.A.Sakhare,"Smart Stick for Blind:Obstacle Detection,Artificial Vision and Real- Time assistance via GPS" Conference (NCICT)2011, Proceedings 2nd National Conference on Information and Communication Technology 2011.

[4] P. Rohit, M. S. Vinay Prasad, S. J. Ranganatha Gowda, D. R. Krishna Raju and I. Quadri, "Image Recognition based SMART AID FOR VISUALLY CHALLENGED PEOPLE," 2019 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2019, pp. 1058-1063,doi: 10.1109/ICCES45898.2019.9002091.

[5] Ziad O. Abu-Faraj, Paul Ibrahim, Elie Jabbour and Anthony Ghaoui, "Design and Development of a Prototype Rehabilitative Shoes and Spectacles for the Blind", IEEE Int. Conf. BioMedical Engineering and Informatics, 2012, pp. 795-799.

[6] B. Deepthi Jain, S. M. Thakur and K. V. Suresh, "Visual Assistance for Blind Using Image Processing," 2018 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2018, pp. 0499-0503, doi: 10.1109/ICCSP.2018.8524251.

[7] Deep Learning for Real-Time 3D Multi-Object Detection, Localisation, and Tracking: Application to Smart Mobility. Antoine Mauri, Redouane Khemmar *, Benoit Decoux, Nicolas Ragot, Romain Rossi, Rim Trabelsi, Rémi Boutteau, Jean-Yves Ertaud and Xavier Savatier.

[8] Redmon, Joseph and Farhadi, Ali. "YOLOv3: An Incremental Improvement" *https://pjreddie.com/darknet/yolo/*. Accessed March 2021.

[9] YOLOv3: An Incremental Improvement. Joseph Redmon, Ali Farhadi University of Washington.

[10] Chen, Q. et al. "Smartphone-Based Outdoor Navigation and Obstacle Avoidance System for the Visually Impaired." MIWAI (2019).

[11] Redmon, J., 2021. *Darknet: Open Source Neural Networks in C*. [online] Pjreddie.com. Available at: <https://pjreddie.com/darknet/> [Accessed 25 March 2021].