IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 24, Issue 2, Ser. IV (Mar. –Apr. 2022), PP 24-30
www.iosrjournals.org

Path texture classification

Brian Mendes
Information Technology. Fr. C. Rodrigues Institute of technology. India

Abstract: *Object detection is a computer vision method that locates the position of the object in the image frame. Deep Learning methods have recently made a remarkable breakthrough in many issues such as computer vision, speech recognition and natural language processing. In Deep Learning, Convolution Neural Network is a multi-layer neural network that helps us in the Object Detection process. We trained a model that detects the different surface regions using the Convolution Neural Network with high accuracy. This work is carried out for totally and partially visually impaired people in order to help them classify the path they are walking on. The model used detects the surface regions and notifies the sightless person through text to sound conversion API. This adds vision to the visionless person. In this paper, we have developed a model that detects different path texture for visually impaired people using Faster RCNN, RESNET and Inception Object Detection models from the given input and predicts which model will achieve greater accuracy. The least loss we obtained was 0.02 from Faster RCNN.*

Key Word*: Object detection, Path classification, Visually Impaired, RCNN, RESNET, Computer Vision, Deep Learning, Dataset.*

Date of Submission: 12-04-2022 Date of Acceptance: 28-04-2022

I. Introduction

The spread of visual impairment is a very sensitive issue worldwide. A world statistics by the World Health Organization (WHO) shows that at least 2.2 billion people have a vision impairment or blindness. Among them, approximately 1 billion have vision impairment, including those with moderate / severe vision impairment or blindness due to undetected refractive error (123.7 million), cataract (65.2 million), glaucoma (6.9 million), corneal opacity (4.2 million), diabetic retinopathy (3 million) and trachoma (2 million), as well as near visual impairment due to undetected presbyopia (8 million). Generally, people over 60 years is having more risk of being susceptible to blindness and the number is increasing by 2000 per decade [1].

Visual impairment is a term used to describe any kind of vision loss. These include people who cannot see at all or who has partial vision loss. It is one of the most severe kinds of handicaps that a person must bear and, given numerous technological advances, it remains a great issue until now.

The creation of walking assistants has been a challenging task, and the demand for assistance equipment remains unchanged. Nevertheless, there are other navigation devices and systems available for visually disabled individuals. The most important helpers are dog guides and white cane. As they're very famous, but they are limited to speed, coverage, and capacity that is normally accessible to people with eyes for route planning [2]. Walking assistants have been used to day-to-day navigation and location aid issues since the 1960s. Assistants help people with visual impairments to detect and locate barriers around them by using sensors that are sensitive to external situations [3].

We propose a system to make visually impaired people to identify what kind of path texture is ahead of them while walking from the input images. The Deep Learning method was used to characterize the texture of the road.

II. Objectives

1. To analysis the impact of the pandemic across the globe in fields such as public health, economic effects, climate changes. -
2. Visualize the data by using various visualization tools available with python
3. Cleaning the data to improve the data quality and overall productivity
4. Use machine learning to predict stock market
5. The trained LSTM model will help us visualize how the stock market is affected due to the pandemic and based on past results will also be able to predict where the market is heading

III. Literature Survey

Multiple different assistants are designed for easy walking to lead people with visually impaired vision. Several organizations have long been working to make them economical and well-organized devices. Research in this area is defined briefly as follows

Bai et al. created an intuitive guidance device designed to help people with visual disabilities travel safely [4]. The system consists of a couple of display glasses and some of the cost-efficient sensors that have been developed and tested by several people. The system worked exclusively indoors. A depth camera, an ultrasonic sensor, a Microprogrammed control station for the measurement of obstructive space, a CPU for image processing and sound analysis, etc., is included in the established device. As feedback on the presence of obstacles, audio instructions are provided. In the home, office and supermarket area, the device is being tested. The proposed glass makes indoor eye observation more efficient and strongly supportive.

In order to aid people with visual disability Kaushalya et al. [5] created walking helpers named 'AKSHI'. The Raspberry Pi 2 sensor, Ultrasonic sensor, RFID reader, GSM and tags were used. The machine can identify and locate obstacles at a lower price. The RFID reader is placed on the base of the handle, enabling the identification of obstacles through the RFID tag.

Vera et al. [6] proposed a "blind map" system for people who are visually impaired to access wireless networks indoor and outdoor space. The device is equipped with a number of integrated sensors that can be found in various areas of the body. The sensor can detect the barrier and generate a retroactive audio signal. A peripheral display, an ultrasonic sensor and a WiFi microcontroller, a core device with a camera module, raspberry pi, and a speaker are part of the hardware components used in this design.

Zhou et al. [7] created an adaptive device called "smart eye" that supports people with visual impairments by asking them about their surroundings. It consists of two components, for example, the integrated mobile and wearable sensors. The module includes power (9-V dc battery), the CPU (32-B NXP LPC1768 mbed), the sensor (Ultrasonic Sensor, Motion Sensor) and communications components (Bluetooth or Wi-Fi).

For visually impaired people Bhatlawande et al. [7] proposed an integrated mobility cane for the recognition of obstacles as well as paths. A key abstract map of the environment will be created and maintained. It provides preferential knowledge to those using input signals such as speech, sound or audio.

O'Brien et al. [8] have presented a cost-effective, integrated, standard cane electronic system that produces an disturbing signal by detecting barriers. A custom printed circuit board with a micro-controller controlled the sensor and motor. The battery is included in the weight of around 110 g. When detecting obstacles, the program calculates the distance. If the distance measured is 0.2-0.6 m and an alarm signal is generated and a distance of 0.6-1 m is produced, a new signal is created. When no obstacles exist within 0.2-1 m, the system will locate a new obstacle. Therefore, the device can also detect obstacles within 1 m.

Cardin et al . [9] developed the Wearable Mobility Improvement System for Visually Disabled People. During the journey, users are warned of close obstacles. The machine detects obstacles to the user and sends accurate vibrational feedback through a multi-sonar machine. The programme, which offers new sensing technologies, aims to improve visual disability versatility. The engineered device consists of sensing the environment through sonar sensors and providing the user with vibratory feedback on the location of closest obstacles within the range. The idea is a cyborg interface that reinforces the sensations of the user. That is to say that the user must use it without any conscious effort, after a training period, as an extension of his own functions. The machine contrasts with the standard white cane. It focuses on the identification of the obstacle on the shoulder and enables users to be entirely hand-free.

A new object detection and classification technique, which works in real time and makes the user aware of the texture of the road, is the key contribution to this article. Our cost-effective, portable system. Our model detects surface regions and reports text to the sound conversion API for visually impaired people. In Fast RCNN Inception and Faster RCNN ResNet Object Detection models from the input and predictions, we've built a model that detects different path textures for the visually impaired. The model can be modified in the future by including it in models detecting obstacles such as cars, dogs, cats, human beings, potholes etc.

IV. Implementation Details

A. Research designs and method

Data Description: Data augmentation is a method that enables individuals to massively improve the diversity of available data to training models without actually receiving new data. Data augmentation methods, such as cropping, folding, and horizontal flipping, are commonly used to train massive neural networks.

Image Annotation: By now, we have the train and test images. But the exact location and type of objects in the images has to be explicitly labelled. **The bounding boxes** can be drawn **using Label Box or Label Image** software, the output of which are saved as XML files, corresponding to each image.

For multi-class classification, give different label names for different objects in the image. This information is saved in the generated XML files. Add all the categories to *label_map.pbtxt* in \data folder and modify NUM_CLASSES variable in code, accordingly.

Fig 1: Label image software being used to annotate objects in the image

Generating TF Records: The input to a TensorFlow Object Detection model is a TFRecord file which you can think of as a compressed representation of the image, the bounding box, the mask etc so that at the time of training the model has all the information in one place. The easiest way to create this file is to use a similar script available for TFRecord for Pets dataset and modifying it a bit for our case.

You will also need to create a label.pbtxt file that is used to convert label name to a numeric id. For my case it was as simple as

```
item{
 id:1
 name:‘concrete’
}
item{
id:2
 name:‘grassy’
}
item {
id:1
name:‘muddy’
}
```

TFRecord is TensorFlows binary storage format. It reduces the training time of your model, as binary data takes up less space and disk read more efficient.

Fig 2: Training/Testing dataset.

We used (70-30) % training-testing partition. The training set consists of 538 images and the testing set consists of 165 images. Some of the images are shown in fig 2.

B. Proposed methodology

CNN Model: In this section, we present the proposed solution for object detection and recognition using CNN-based architecture. First of all, we take into account the calculation of the suitability of the object in order to identify the significant parts of the image according to the proposed model. Later on, using convolution layers, a CNN model has been implemented to generate activation maps. Over the next phase, we determine the hierarchical feature maps and create a uniform feature space called a key feature generated by the Max Pooling and De-convolution process. Based on these processes, a ROI pooling model is developed to generate mapped image features. With the support of this feature map image, we generate a proposal using the region proposal generation technique. However, the initial regional proposals generated are noisy due to the region's background response and the edge response. In addition, saliency-based segmentation is used to obtain ROI (region of interest) for the extraction of features. This ROI comprises the respective class and resolution data to maintain the quality of the detection. Thus, we apply the refinement of the region proposal to improve the process of generation of the proposal and, finally, the object detection module is used to perform the detection task. A complete flow of proposed approach is depicted in Fig. 2.

The complete identification process can be divided into two steps as a segmentation and identification that provides the semantic map and bounding boxes on the image that are useful for identification in various stages. Later, those outputs are analysed in order to obtain the initial estimations of bounding boxes and box regions that are likely to be known as segmented regions. In the next step, these bounding boxes are interpreted as original estimates and sent to CNN, which offers images and eventually provides details on the total items present in the shot. Once the objects are identified, the bounding box regression model is introduced to eliminate bounding box errors and, eventually, the recognition procedure is applied.

Faster RCNN: Faster RCNN is the object detection architecture proposed by Shaoqing Ren, Ross Girshick, Jian Sun and Kaiming He in the year 2015 and is one of the popular object detection architectures that uses convolutional neural networks such as YOLO (You Look Only Once) and SSD (Single Shot Detector).

The biggest issue with R-CNN is that it's very slow to run. It can take 47 secs to process a single image on a typical deep learning computer, leaving it useless for real-time image processing situations. The primary issue that slows down R-CNN is the Selective Search process that intends many probable regions and necessitates all of them to be classified. In addition, the process of selecting a region is not "deep" and does not involve learning, restricting its accuracy.

Shoqing Ren et.al. [9] presented an enhanced algorithm named Faster R-CNN, which completely eliminates Selective Search and allows the model to learn directly about the region's proposals. Faster R-CNN takes the original picture and sends it to CNN titled the Region Prediction Network (RPN). This considers a vast number of potential regions, much more than in the original R-CNN model, and uses an effective deep learning approach to determine the regions are very much likely to be of interest.

The region suggestions are then restructured using a pooling layer of region of interest (RoI). This layer on its own is used to classify the images in each region and also to predict offset values for bounding boxes.

Proposed Method: After data annotation, when the data is converted to xml files, we trained on Faster R-CNN model for 100k epochs on NVIDIA GPU using CONFIG file from TENSORFLOW. It took around 12 hours to run 100k epochs.
Here number of classes were kept as 3, l2 regularizer was used, and activation function as relu, and batch size was kept as 24.
We also trained on RESNET for 100k epochs.

Results:
As we have shown a comparison between Faster R-CNN, RESNET and Inception model, we obtained following results

Model	Loss
Faster R-CNN	0.02%
RESNET	0.05%
Inception	0.06%

C. Overview

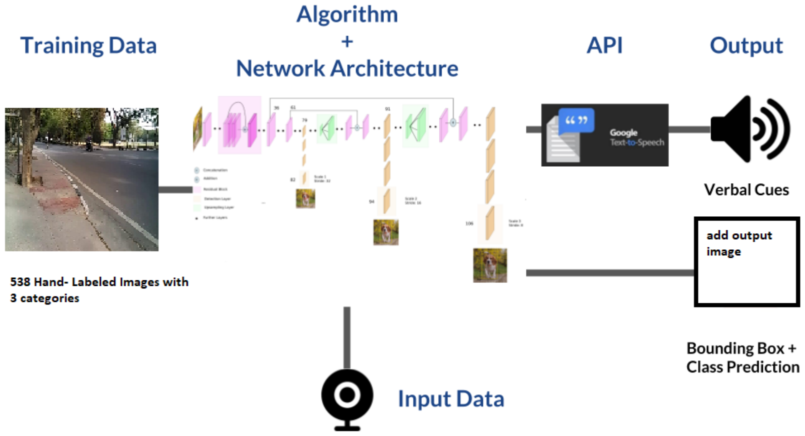

Fig 3. Complete flow of the project

Selecting model hyper parameters: Now you can choose the Mask Model you want to use. The TensorFlow API provides 4 model options. I chose the Mask RCNN Inception V2 which means that Inception V2 is used as the feature extractor. This model is the fastest at inference time though it may not have the highest accuracy. The model parameters are stored in a config file. I used the config file for the coco model of the same type and updated the number of classes and the paths keeping most of the model parameters the same.

Training the model: With the input files and the parameters locked, you can start the training. I was able to train this model on a GPU in a few hours. You can start the training job and the evaluation jobs on two separate terminals at the same time. And initiate tensor board to monitor performance. I stopped training after 100k epochs.

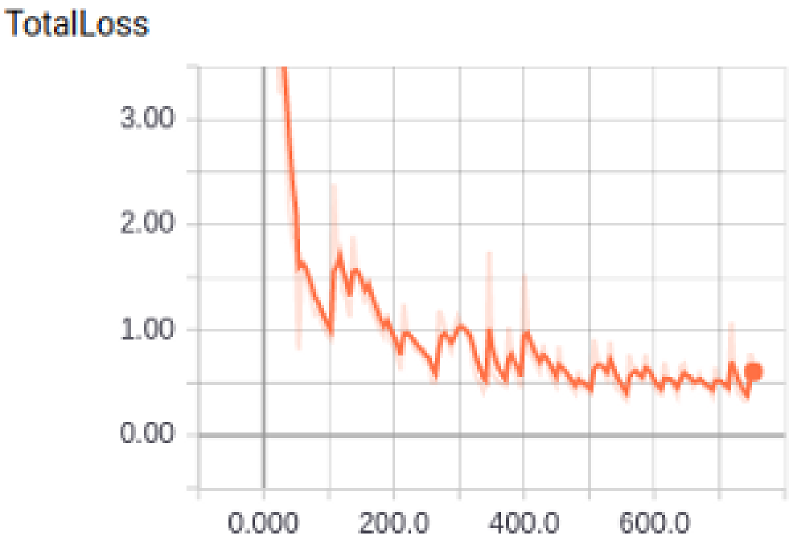

Fig 4: Total loss graph

The coolest thing in Tensor board is that it allows you to visualize the predictions on sample images from test set as training progresses. The gif below shows the model becoming certain of its mask and bounding box predictions as the training progresses.

Test the model on your custom video: To test the model, we first select a model checkpoint (usually the latest) and export it to a frozen inference graph. The script for this is also on my GitHub. I tested the model on a new video from my webcam.

For more information and the source code please check out my GitHub repository: https://github.com/Brianrmendes/Path-texture-classification-for-visually-impaired

V. Conclusion

The paper shows that using Models provided by TENSORFLOW can be used for real time detection and identification of path such as muddy, concrete and grassy. Faster R-CNN proved to be best working among the three models used.

A. Benefits to the society

Machine Learning (ML) is a specialized sub-field of Artificial Intelligence (AI) where algorithms can learn and improve themselves by studying high volumes of available data. In this field, traditional programming rules do not operate; very high volumes of data alone can teach the algorithms to create better computing models.

Given the unique attributes of Machine Learning algorithms, they work best with Big Data because the volumes and complexity of such data are so high. In other words, Machine Learning aspires to mimic the human brain – learning by observing.

With the growing popularity of Artificial Intelligence and AI-enabled smart solutions for businesses, ML is also gaining rapid prominence in the global business world. With the promise of an algorithm economy that is destined to take over the normal business environment in a few years, it is imperative that more AI or ML students, enthusiasts, and practitioners take the time to review the "supplementary literature" available around us to fully understand the scope of Machine Learning applications for the business world. This article takes the readers through a quick review of ML algorithms, use cases, and best practices.

The power of Machine Learning is hidden in the self-teaching algorithms, which when exposed to huge amounts of data, can study and learn for improved results. In Rules of Machine Learning, the technical experts can discover the modus operandi of ML, but for the most effective way to explore ML is to talk to experts on neural networks or Deep Learning, or unsupervised learning. ML algorithms have been categorized as of the following types of learning models: supervised, unsupervised, and reinforcement. A common ML solution known to everyone is Netflix's recommender tool that suggests new movies based on a viewer's past movie selections.

B. SWOC Analysis

STRENGTHS: Our specially trained model has achieved high accuracy and minimal loss which helps the model to detect and differentiate the paths better. This would massively help the visually impaired to know the path they're walking on

WEAKNESS: When the model is presented with a picture other than path, sometimes it would detect different parts of the image as several paths which is something that needs to be worked on, but on a practical standpoint if it is used in a real time camera, there is a very slight chance of such kind of bug occurring

OPPORTUNITIES: With the help of a hardware device which can be attached to a cap or sunglasses, the model can transmit voice for the visually impaired and help them to recognize the path.

CHALLENGES: The AI and ML has a lot of competition and thus it gets difficult to produce the best possible version of the model. The covid pandemic certainly made things difficult for me to get help from my peers or professors which also slowed down the process of building and training the model

C. Future Scope

Features to be added:

1. Extend this model to multiple categories of objects in the same image. The TFRecord creator script would need some modifications so it can properly assign each object the correct label and mask
2. Use of a hardware device to make the model work flawlessly in real-time capture
3. Since the lighter version model was used for this project. We could also see how other models suite this project that are slower, perform in terms of accuracy of detection
4. Can also be launched in industries and tested with volunteers who are willing to test and give feedback so that further improvement on this app can be made

References

[1]. World Health Organization, GLOBAL DATA ON VISUAL IMPAIRMENTS : Fact sheet number 3, 2010. [Online]. https://www.who.int/blindness/GLOBALDATAFINALforweb.pdf?ua=1

[2]. Blasch, B.B., Wiener, W.R., & Welsh, R.L., " Foundations of orientation and mobility," (2nd ed.). NY: AFB Press, New York, 1997.

[3]. Marion, A.H., Michael, A.J., "Assistive technology for Visuallyimpaired and Blind People," Springer, London, UK, 2008.

[4]. J. Bai, S. Lian, Z. Liu, K. Wang and D. Liu, "Smart guiding glasses for visually impaired people in indoor environment," in IEEE Transactions on Consumer Electronics, vol. 63, no. 3, pp. 258-266, August 2017, doi: 10.1109/TCE.2017.014980.

[5]. Kaushalya, V. S. S., K. Premarathne, H. M. Shadir, P. R. Krithika and S. G. S. Fernando. """ AKSHI " : Automated Help aid for Visually Impaired People using Obstacle Detection and GPS Technology." (2016).

[6]. Vera, Daniel & Marcillo, Diego & Pereira, António. (2017). Blind Guide: Anytime, Anywhere Solution for Guiding Blind People. 353-363. 10.1007/978-3-319-56538-5_36.

[7]. Zhou, David & Yang, Yonggao & Yan, Hanbing. (2016). A Smart \"Virtual Eye\" Mobile System for the Visually Impaired. IEEE Potentials. 35. 13-20. 10.1109/MPOT.2015.2501406.

[8]. S. Bhatlawande, M. Mahadevappa, J. Mukherjee, M. Biswas, D. Das and S. Gupta, "Design, Development, and Clinical Evaluation of the Electronic Mobility Cane for Vision Rehabilitation," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 22, no. 6, pp. 1148-1159, Nov. 2014, doi: 10.1109/TNSRE.2014.2324974.

[9]. Faster R-CNN: Towards Real-Time ObjectDetection with Region Proposal NetworksShaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun

Brian Mendes. "Path texture classification." *IOSR Journal of Computer Engineering (IOSR-JCE),* 24(2), 2022, pp. 24-30.