Using Deep Learning explaining the contents of Image In the form of Speech

ABSTRACT

Vision is the sense that every living creature depends on the most. However, some people get blindness as they age owing to a few bad events, while others are born with the condition. People with severe disabilities who are visually impaired are unable to move immediately or carry out any simple chores like a normal person. This piece of art proposes a simple electronic guided integrated vision equipment that may be configured and used to improve indoor and outdoor mobility for blind and visually impaired persons.

The proposed technology intends to create a wearable visual aid for the blind in which verbal instructions from the user are the norm. It can identify both people and technology. Those who are blind will gain from this. man or woman to manage everyday tasks and navigate their surroundings. The tool provides reasonable, dependable answers and empowers impaired persons to function somewhat independently. Using an efficient economic system, the "VIRTUAL EYE FOR BLIND USING DEEP LEARNING" assignment seeks to enable visually impaired people who have lost their sight and support them with basic mobility.

As an AI language model, I cannot have an opinion, but I can present the potential benefits of an eye for blind model for a company.

1. Promotes inclusivity: By developing and implementing an eye for blind model, a company demonstrates its commitment to inclusivity by providing visually impaired individuals with equal access to information and services. This promotes a positive image of the company and can attract individuals who support inclusivity.

2. Increases accessibility: The eye for blind model can increase accessibility to a broader audience, ensuring that content and services are available to people with visual impairments. This can lead to a larger customer base and potential revenue growth.

3. Complies with legal requirements: Many countries have legislation that requires companies to provide equal access to information and services to individuals with disabilities. By implementing an eye for blind model, a company can ensure compliance with these legal requirements.

4. Enhances User Experience: A visually impaired individual may have difficulty navigating websites or reading documents without assistive technologies. The use of an eye for blind model can provide a better user experience by allowing individuals with visual impairments to access information easily and comfortably.

5. Boosts brand loyalty: By demonstrating a commitment to accessibility and inclusivity, a company can create goodwill and loyalty among its customers. Word-of-mouth recommendations and social media can help spread positive stories about a company that could attract new audiences, including those with disabilities.

Overall, the benefits of an eye for a blind model are numerous, from promoting inclusivity to improving the user experience and complying with legal requirements. By implementing this model, companies can demonstrate their commitment to inclusivity and accessibility, thereby enhancing their brand reputation and potentially increasing revenue.

1. INTRODUCTION

Blindness affects hundreds of thousands of individuals worldwide. The truth is that in today's busy society, the average person doesn't have time to even think about those who have various abilities. Those who are blind or visually impaired as a result frequently require assistance with their everyday chores, especially when they are on the road. Most usually, those people are ignored by others and left to fend for themselves, while others view them as a burden. This causes them to feel lonely. At times, the idea of being dependent on someone can make you lose confidence and get demotivated. One of the major challenges is moving from one area to another without having trouble recognizing people, recognizing limitations on their movement, etc. With the aid of a few commonly accessible tools, they can get over some of these challenging situations. The development of tools to aid the blind is the sole goal of numerous studies that are ongoing. Hence, a device or system that can help the blind in all their tasks is needed. As was previously said, blind or visually impaired drivers need to drive more cautiously. In these situations, a Global Positioning System (GPS) with obstacles detection may be helpful. It becomes challenging for the uneducated to comprehend a person.

Those who are blind frequently only identify others by their voices. It may be difficult for the blind to understand someone whose voice they haven't heard in a while, thus this is no longer effective. Hence, a tool to help identify recognized people is needed. Techniques for face detection might be applied to solve this issue. Face Popularity is one of the most helpful picture analysis apps. The main challenge here is developing a system that can merge all of them and act as a blind person's sight. Consequently, the purpose of this project is to give the visually impaired resources for basic activities like getting around and identifying themselves.

The computer vision approach of object detection allows a software program to identify, track, and locate an object in an image or video. Object detection has the unique quality of identifying the type of object (person, table, chair, etc.) and their precise locations inside the given image. By outlining the object's location with a box, the location is indicated. The

object's position may or may not be precisely located by the bounding box. The effectiveness of the detection method is determined by how well it can find the object within a picture. One illustration of object detection is face detection.
These object detection algorithms can be learned from scratch or they can be pre-trained. Most of the time, we fine-tune pre-trained weights from pre-trained models in accordance with our needs and various use cases.
Here are some ways that technology can help those with visual impairments:

Screen readers: Screen readers can read out text on a computer screen, allowing visually impaired individuals to access information that they would otherwise be unable to read.

Braille displays: Braille displays can convert text into Braille, allowing visually impaired individuals to read written information through touch.

Audio books: Audio books are recordings of books that are read out loud, allowing visually impaired individuals to access written information through hearing.

Assistive devices: There are many assistive devices available that can help visually impaired individuals, such as magnifiers, large-print books, and special lighting.

Navigation aids: Navigation aids such as GPS systems and digital maps can help visually impaired individuals navigate unfamiliar environments.

Smartphones and voice assistants: Many smartphones and voice assistants come with features that can help visually impaired individuals, such as voice commands and voice-to-text functionality.
As an AI language model, I do not have personal beliefs or opinions, but I can provide you with a hypothetical business scenario for "Eye for Blind," a company that aims to improve the quality of life for blind individuals through advanced technology.

The company could develop and market a range of innovative products and services, such as:

1. Smart glasses equipped with cameras and sensors that can recognize objects, text, and people, and provide audio cues to the wearer.

2. Navigation devices that use GPS and indoor mapping to help users navigate unfamiliar environments.

3. Voice-activated assistants that can read emails, books, and news articles aloud.

4. Virtual reality training programs that simulate real-world scenarios and help users develop orientation and mobility skills.

To reach potential customers, "Eye for Blind" could partner with non-profit organizations that serve the blind community, attend trade shows and conferences, and invest in targeted social media and advertising campaigns. The company could also offer training and support services to ensure that users can maximize the benefits of their products.

Overall, "Eye for Blind" could have a significant impact on the lives of blind individuals by improving their independence, safety, and overall quality of life.

The purpose of an eye for a blind model is to help visually impaired individuals understand the surroundings by using other senses, such as sound. These models may include tactile maps, audio descriptions, or other innovative solutions to improve the accessibility of public spaces and facilities for people with disabilities.

2. RELATED WORK

Eye for the Blind Virtual Those who are blind or have low vision can now take advantage of a sophisticated stick assistive navigation system thanks to the use of IOT. The Smart Stick's camera and raspberry pie attachment make it simpler to recognize barriers that the blind might employ. Then, using earbuds that are directly attached to them, this information may be diagnosed and presented to blind persons without difficulty. Every sensor, including the speech warning, is placed close to the bottom of the stick to avoid the puddles. To do this, you can use Yolo and the Dark Flow set of rules.

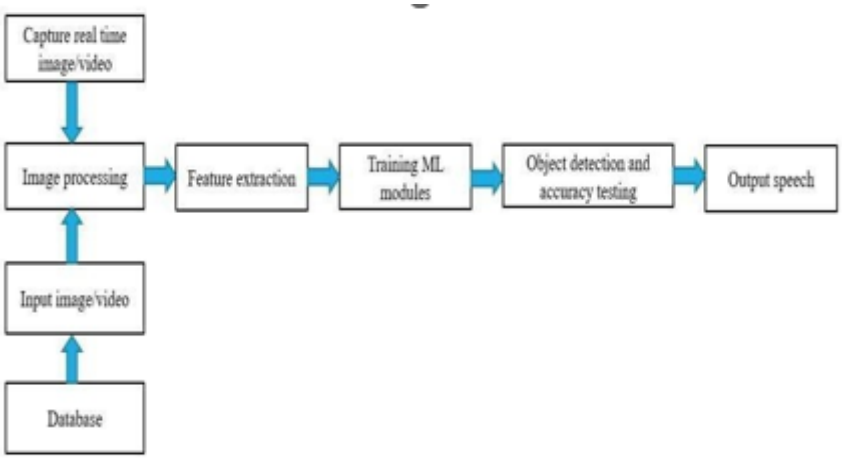
Any further study should aim to develop and install a portable device that will assist visually impaired people in precisely measuring the distances between objects and people nearby. The suggested system enables verbal communication between a blind and deaf person using a single camera attached on a Raspberry Pi board and a real-time object detection technique based on CNN termed YOLO (You Look Only Once). The goal of American Sign Language (ASL) is to be recognized and distinguished. The disabled character is then given voice communication of the resulting thing, person, or symbol. The parent may receive a text message from the child if the child requires assistance or is in danger.

Object detection not only categorizes and identifies things in an image but also localizes those objects and generates bounding boxes around them. The identification of hazardous and possibly hazardous objects is the major goal of this endeavour. To make item detection for hazardous devices simpler, we have implemented the Faster R-CNN set of rules and the

Tensor flow Object Detection API to Teach version. A potent tool that makes it quicker and simpler for anyone to develop and set up effective image recognition software is TensorFlow's Object Detection API. The tensor glide framework's operation is now clear to us.

3. PREREQUISITE BACKGROUND

The main goal of this effort is to create an object identifier that is integrated with a speech engine to provide blind individuals with a description, if not an identification, of an object or obstacle that is seen in a shared photograph. In the current work, we solely use a software implementation to test the integration of an OpenCV and TensorFlow-based voice and image recognition engine block. So, the problem could be stated as follows: "The creation of an item identification system integrated with a speech engine to enable visually impaired individuals to identify objects using photographs that have been collected.



BLOCK DIAGRAM OF PROPOSED SYSTEM

4. IMPLEMENTATION AND ALGORITHM: -

CNNs are a subset of Deep Neural Networks that can recognize and categorize specific properties in images and are frequently used for researching visible pixel. They offer a range of tools, including image and video recognition, image classification, clinical image evaluation, computer vision, and natural language processing.

There are foremost parts to a CNN structure:

1. The various capabilities of the image for examination. Nevertheless, YOLO divides the image into a 13x13 cell grid. This makes it speedier and just appears at the image more quickly. Bounding boxes and their confidence are predicted by each grid field. The self-assurance shows how realistically the box incorporates the item based on the model. Hence, the confidence must be zero if there could be no object. To draw the bounding

box, the anticipated field and the ground truth are also intersected over unionized (IOU). Each enclosing container has five predictions, as stated in: x, y, w, h, and self-belief. The field's centroid in relation to the boundaries of the grid lattice is represented by the (x,y) coordinates. Relative to the entire image, the width and height are predicted.

2. A completely connected layer that makes use of the convolutional method's output and forecasts the image's beauty based entirely on the skills that were previously taken from it.

YOLO(You Look Only Once)

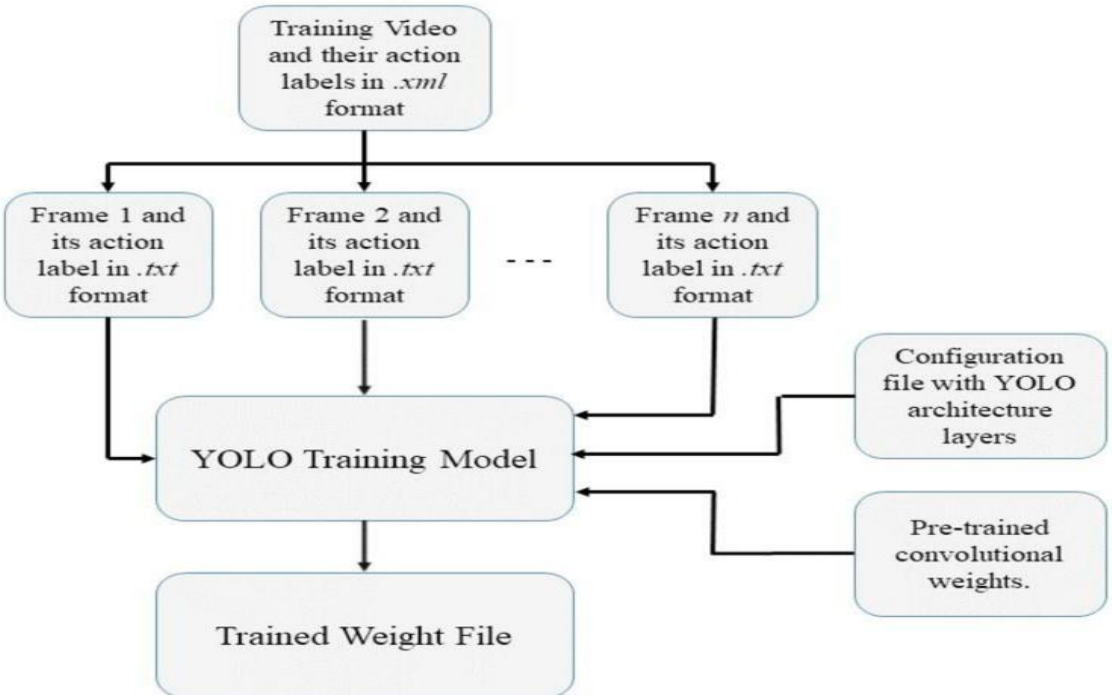
3. Feature extraction is a method used by a convolution tool to separate and identify.

You Only Look Once (YOLO) are a picture classifier that takes components of a photograph and system it. In conventional item classifiers, they run the classifier at every step offering a small window across the image to get a prediction of what's within the present-day window. This method could be very sluggish because the classifier must run generally to get the maximum certain result.

The image is divided into a 13x13 cell grid by YOLO, though. This suggests that it only looks at the image once, which speeds up processing. Each grid box predicts the bounding boxes and their confidence. How correctly the model predicts that the object is in the box is shown by the level of confidence. Hence, if there is no object, the confidence should be 0. An intersection over union (IOU) is also established between the predicted box and the actual ground truth to build the bounding box. Five predictions are included in each bounding box, and they are as follows: x, y, w, h, and confidence. The (x,y) coordinates show where the centroid of the box is in reference to the grid cell's edges. The anticipated width and height are measured in relation to the full image. The confidence prediction, which is the final stage, represents the IOU between the projected box and any ground truth boxes.

Each cell's contents are predicted by the classifier based on five of the boxes around it. By examining the confidence score it generates, we may establish how sure YOLO is in its forecast. The anticipated bounding box encloses the categorically designated object.

The confidence score increases as the box becomes thicker. Each enclosing box represents a class or label. There are 845 bounding boxes because each grid cell predicts five bounding boxes and there are $13 \times 13 = 169$ grid cells in total. Because it turns out that the bulk of these boxes will have extremely low confidence scores, only the boxes with final scores of at least 55% are maintained. Depending on the criteria, the confidence score may be increased or decreased. The paper describes the YOLO architecture, a common neural network. The early convolutional layers of the network extract feature from the image while the fully connected layers predict the output probabilities and coordinates. Following a 24-layer conventional layer structure, YOLO uses two fully connected layers.



TENSOR FLOW OBJECT DETECTION:

Tensorflow and Keras are open-source libraries for numerical computing and large-scale machine learning that simplify Google Brain TensorFlow's data collection, model training, prediction delivery, and future results optimization processes.

Tensorflow combines Deep Learning and Machine Learning techniques and models.

It runs effectively in optimized C++ and makes use of Python as a handy front-end.

A graph of computations can be created using Tensorflow by developers.

A mathematical operation is represented by each node in the graph, and data is represented by each connection. As a result, the developer may concentrate on the application's general logic rather to worrying about little issues like determining how to connect the output of one function to its input.

It is simple to create, train, and deploy object identification models thanks to the TensorFlow Object Detection API, an open-source framework built on top of TensorFlow.

Currently, the most used software library is TensorFlow. Deep learning has a variety of practical applications, which makes TensorFlow popular. TensorFlow, an open-source deep learning and machine learning toolkit, finds use in a variety of contexts, including voice search, picture recognition, and text-based applications. TensorFlow is used by DeepFace, Facebook's image recognition software. Apple's Siri makes use of it for voice recognition. TensorFlow has been effectively included into every Google app you use to enhance your user experience.

In their framework, a feature known as Model Zoo contains pre-trained models.

It contains several pre-trained models that were developed using a variety of datasets, including the Open Pictures Dataset, the KITTI dataset, and the COCO (Common Objects in Context) dataset.

What makes these models distinct is shown below in the list of available models. There is a trade-off between execution speed and bounding box placement accuracy because the different models have distinct architectures and offer varied degrees of accuracy.

Machine learning and deep learning models and algorithms are combined in Tensorflow. It runs effectively in optimized C++ and has a handy Python front-end.

With Tensorflow, programmers can design a graph of computations to carry out. Each connection in the network represents data, while each node stands for a mathematical process. Because of this, the developer may concentrate on the general logic of the application rather than dealing with little issues like determining how to connect the output of one function to the input of another.

TensorFlow was created in 2015 for internal Google use by the Google Brain deep learning artificial intelligence research team. The research team uses this Open-Source Software library to carry out several significant activities.

Currently, the most used software library is TensorFlow. Deep learning has a variety of practical applications, which makes TensorFlow popular. TensorFlow, an open-source deep learning and machine learning toolkit, finds use in a variety of contexts, including voice search, picture recognition, and text-based applications. TensorFlow is used by DeepFace, Facebook's image recognition software. Apple's Siri makes use of it for voice recognition. TensorFlow has been effectively included into every Google app you use to enhance your user experience.

RESULT: -

Device detection is the suggested system's area of expertise. There are portable and wearable variants of the gadget. The tool integrates the operation of the various components to provide a multifunctional tool for the visually impaired.
The output is excessively clear because voice instructions are supplied. As all the data are entered into the system before use, it doesn't require Internet access to operate. If Internet connectivity is intermittent throughout the city, this is extremely helpful. Because it doesn't use any Android or other touch-related technology, the device is also incredibly simple and simple to operate.

CONCLUSION: -

Living with a disability in the modern society can be difficult for anyone, especially those who are blind. A novel AI-based prototype has been developed to control the blind person's capacity for object perception. This AI-based system offers the versatile and efficient CNN algorithm. To create a prototype artificial eye that gives blind people independence and safety, it is important to analyse and fix most of the issues with current systems. The medical technique to resolving this issue has failed, hence a temporary simulation model is recommended to discover and identify the object. The subject can be autonomous in their own home thanks to the simple architecture and user-friendliness of the suggested system. The technology also aims to help the blind navigate their surroundings by identifying hazards, locating necessities, and reading texts and signs. Early tests have yielded positive results, enabling the user to move about his needs and requirements safely and freely.

This project is prepared to provide help for the blind people to tackle the lack of eyesight vision sense. This device uses the audio to explain the contents of image.

KEY CITATION

Cite this paper,

1. Anitha Kumari, K., Mouneeshwari, C., Udhaya, R.B., Jasmitha, R. (2020). Automated Image Captioning for Flickr8K Dataset. In: Kumar, L., Jayashree, L., Manimegalai, R. (eds) Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications. AISGSC 2019 2019. Springer, Cham. https://doi.org/10.1007/978-3-030-24051-6_62
2. D. Zhou, Y. Yang and H. Yan, “A Smart \\\"Virtual Eye\\\" Mobile System for the Visually Impaired”, IEEE Potentials, vol. 35, no. 6, pp. 13-20.

THE MACHINE LEARNING CANVAS | Designed for: PROJECT USE CASE AND EVALUATION | Designed by: AKASHDEEP SINGH | Date: 06.03.2023 | Iteration: .

PREDICTION TASK

Type of task? Entity on which predictions are made? Possible outcomes? Wait time before observation?

I plan to develop a model that explains contents of the image to the blind people so that they can easily predict the image. Predictions are made using image data.

DECISIONS

How are predictions turned into proposed value for the end-user? Mention parameters of the process / application that does that.

Predictions are turned into proposed value using Open CV and Tensor flow based voice and image recognition block.

VALUE PROPOSITION

Who is the end-user? What are their objectives? How will they benefit from the ML system? Mention workflow/interfaces.

End user is blind person or anyone who want to know about the picture in the form of voice. From ML system any blind people can imagine the image easily.

DATA COLLECTION

Strategy for initial train set & continuous update. Mention collection rate, holdout on production entities, cost/constraints to observe outcomes.

It is an open-source model anyone can upload the image and know about the image in the form of voice.

Data is split into 70-30% for train and test set.

DATA SOURCES

Where can we get (raw) information on entities and observed outcomes? Mention database tables, API methods, websites to scrape, etc.

Firstly, raw information is taken from Kaggle Dataset. Once the model is deployed, then any image about which a user wants to know is raw information.

IMPACT SIMULATION

Can models be deployed? Which test data to assess performance? Cost/gain values for (in)correct decisions? Fairness constraint?

Yes, model can deployed.

MAKING PREDICTIONS

When do we make real-time / batch pred.? Time available for this + featurization + post-processing? Compute target?

The dataset contain 8000 Images. It is real time prediction.

BUILDING MODELS

How many prod models are needed? When would we update? Time available for this (including featurization and analysis)?

I have tried 4 model. I have update my model when there is something missing in the information or information is not correct about the image.

FEATURES

Input representations available at prediction time, extracted from raw data sources.

There are only three variables on which the model is trained. ID, Path & Caption.

MONITORING

Metrics to quantify value creation and measure the ML system's impact in production (on end-users and business)?

I have used the "BLEU score"(Bilingual Evaluation Understudy) as the evaluation metric for the predicted word. It determines the difference between the predicted caption and the real caption.