Artificial Vision Using Blind for Alexa

E.Sakthivel[1], Anjali.V.R[2], Dharshini.R[3], Kaviya.S[4]
[1,2,3,4] *Electronics and Communication Engineering, Jeppiaar Engineering College*

Abstract: One billion people worldwide suffer from a vision problems that should have been avoided or is yet unaddressed. In terms of geographical differences, low- and middle-income countries are projected to have four times the prevalence of distant vision impairment as high-income regions. In terms of near vision, rates of untreated near impaired vision are projected to be bigger than 80percentage points in western, eastern, and sector Directly Africa, while percentages in high-income regions such as Asia Pacific, Australasian, Western Europe, and Asia-Pacific are less than 10%. The probability of more people developing vision impairment is predicted to rise as the population grows and ages. In order to assist the blind, we created a device imaging system in which a blind person can carry an audio device with them that will guide them through their surroundings and help them live a safer life while also increasing awareness of their surroundings. This was accomplished by applying innovative image captioning techniques that included the use of effective net B3 procedures and tokenization approaches, in which the machine learnt situations with different captions. When a picture is taken with a camera, the CPU recognizes it and predicts it. Following the estimate, it will be given to the alexa microphone, which will provide an auditory output to the user, allowing them to recognize the scene that is unfolding around them. As a result of this study, we are able to deliver synthetic eyesight to the blind, allowing them to acquire confidence when travelling alone.

Keywords: Blind People, Real time Application, Raspberry pie, Alexa, Python Programming Language, Artificial Intelligence, Deep Learning, Image Capture, Audio Output, identification of visual relationships.

I. INTRODUCTION

Unmanned vehicles have piqued the interest of many academics as a hot subject in the area of intelligent transport systems [1]. The capacities of autonomous vehicle algorithms have improved thanks to recent advances in deep learning [2]. Verifying those techniques is a challenge worth thinking about. Direct testing on real-world road circumstances, on the other hand, is time-consuming and costly. In recent years, offline testing of drone vehicle algorithms has grown popular[3]. To address the issues raised above, various driving simulation programmers also including PreScan [4] as well as Carla have been created for offline testing. We may test and evaluate uav vehicles algorithms in offline situations using these platforms that can create and mimic traffic scenarios [5]. Unmanned vehicles may safely mimic millions of kilometers in a short amount of time in virtual mode, providing a solid foundation for further testing in actual traffic.

The human eye functions similarly to a camera, collecting, focusing, and transmitting light through with a lens to produce pictures of its surroundings. The image is generated on films or an imaging system in a camera. The retina, a small covering of light-sensitive tissue in the back of the eye, is where the image is produced. The human eye, like a camera, regulates the intensity of the light the eye. The quantity of light that passes through the pupil is controlled by the iris (the colourful circular region of the eye). In strong light, it narrows the pupil, but in dim light, it widens it. The cornea is the eye's transparent and protective surface. It, together with the lens that sits just behind iris, helps concentrate light. The retina converts photons into nerve signals as it reaches the eye. The retina then conveys these impulses to the optic nerve to the brain (a cable with over 1,000,000 nerve fibres). The eye cannot connect with the brain without the need for a retinal or optic nerve, rendering vision impossible. Many people experience vision difficulties that at a certain point in life. Some people have lost their ability to see items that are far away. Others struggle to read small print. Sunglasses or contact lenses are frequently used to address these disorders. However, significant or total vision loss can occur when one or more components of the eyeball or brain that processing images becomes diseased or damaged. Hospital attention, surgical, or corrected lenses like spectacles or

contacts aren't enough to restore eyesight in many circumstances. And according to American Foundation for something like the Blind, there are 10 million visually impaired people in the Usa. Experts use the term "visual impairment" to describe any type of loss of vision, whether someone can't see at all or has incomplete vision loss. Some people are fully blind, but many more suffer from a condition known as legal blindness. They haven't entirely lost their eyesight, but it has deteriorated to the point where they would have to approach 20 feet away from an item to see it as clearly as somebody with perfect eyesight could out of 200 feet away.

Main Objective of our Project

- The project's goal is to provide blind people artificial vision so they can become more aware of their environment.
- Implement a strong scene prediction system that may be used in a variety of settings.
- In order to obtain successful outcomes, Alexa must be integrated.

The remainder of the paper was organized. Parts of the system include Chapter I: Introduction, Chapter II: Literature Review, Chapter III: Methodology, Chapter IV: Results and Discussion, and Chapter V: Conclusion. This section concludes the references section.

II. REVIEW OF LITERATURE

Iqbal, A et al., " A low cost artificial vision system for visually impaired people". For visually challenged people, a low-cost navigation system based on the AT89C52 microcontroller has been created. Ultrasonic sensors are utilized to calculate the distance between the blind person and the obstructions in their way, guiding the user to the available path. The transmission is in the format of a voice that the blind person would understand, such as right, left, and so on. The device will be designed to distinguish items utilizing image processing methods in its advanced mode. The results are offered to demonstrate the system's validity and performance.

Kalaivani. K et al., "An Artificial Eye for Blind People". Visually impaired persons are always attempting to live in harmony with their surroundings. Their day-to-day activities, however, are severely limited due to their loss of vision. Those walking sticks have the ability to detect the objects with which they come into contact. To address this challenge, the authors used Raspberry Pi to create an electronic help in the form of an intelligent stick. For blind persons, this proposed stick provides artificial vision, object detection, and real-time GPS guiding. The suggested gadget detects an objects in its surroundings and provides speech information, earphone alerts, and GPS navigation to a specific place. The intelligent stick's ultimate goal is to provide a subtle amount of and efficacious orientation trying to find and obstruction detecting assistance for the blind, providing an interpretation of imitation vision by providing information about the environmental situation of stationary and moving objects nearby, allowing them to walk independently.

Caraiman. S et al., "Computer vision for the visually impaired: the sound of vision system". This paper describes a perceptual displacement device for the sight handicapped based on computer vision. Its primary goal is to provide users with a three-dimensional depiction of the surroundings around them, which is delivered through the audible and tactile senses. One of the most difficult tasks for this network is to provide pervasiveness, or the ability to work in any inside or outside location and under any lighting condition. This paper explains the equipment (3D competitive bidding) and programming (3D processing pipeline) utilised to create this sensory substitution device, as well as how it can be used in different contexts. Preliminary usability testing with blind users yielded positive findings and provided useful suggestions for system improvement.

Hwan. S et al., "Temporizer, factorize and regularize: Robust visual relationship learning". In this paper, we begin with a simple number of co learning model, which provides a rich formalization for establishing a powerful precondition for remembering visual relationships in theory. While the interpretation problem for determining the regularize is difficult, our major technical accomplishment is to illustrate how current advances in numerical methods may be used to devise effective strategies for a construction scheme that produces highly informative priors. Even without using visual characteristics, the factorization gives limited sample bounds for inference for the underpinning [[object, predicate, object]] connection learning task (under moderate conditions) and outperforms (in certain cases) existing approaches.

We then significantly improve the state-of-the-art by combining it with an end-to-end framework for visual relationship recognition using image data.
Lu, C et al., "Visual relationship detection with language priors". We present a model that takes advantage of this knowledge to train individual visual representations for objects and predicates, then combines them to anticipate complex interactions per image. We build on previous work by fine-tuning the likelihood of a projected relationship using linguistic prior convictions from semantic word embeddings. From a few instances, our algorithm can predict thousands of different sorts of associations. We also use bounding boxes in the image to localise the elements in the predicted relationships. We also show how a better grasp of relationships might help with content-based picture retrieval.

III. PROPOSED WORK

We can see how the approaches are employed to show the system's result in our proposed way. To assist the blind, we designed an intelligent vision system in which all the blind person should keep an Audio device that would guide them around it and allow them to live a supposed to protect while sharpening their awareness of their surroundings. This is accomplished by employing modern captioning techniques that employ effective mesh computations and tokenization approaches, in which the machine learns scenes with different captions. The processor recognises and predicts images acquired by the camera every time they are captured. Following the prediction, it is transferred to the Alexa microphones, which provides the user with an audio output that can assist them in identifying the scene that is currently taking place in the area. For example, we are delivering artificial vision to blind persons as part of this initiative, which will assist them acquire confidence when walking alone.

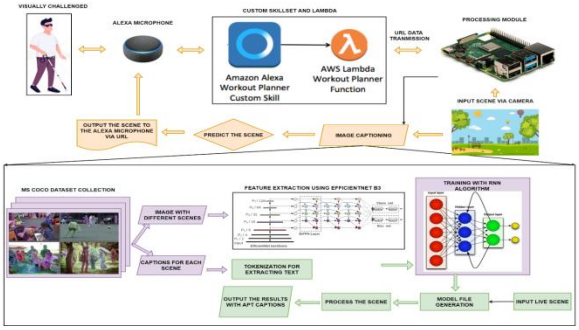


Figure 1 Architecture Diagram of Proposed method

In our project, we may use the five modules to display the system's results in an efficient manner.

Modules are follow as,

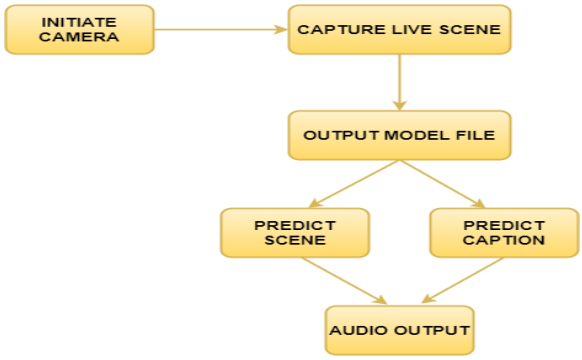
- Dataset collection
- Efficient net B3 Feature Extraction
- Tokenization
- RNN Algorithm Training
- Scenes and audio output prediction

Datasets Collection: The COCO dataset is utilised to train a model in this project. A database set is a set of information. Deep Learning has emerged as the preferred way for tackling a wide range of difficult real-world situations. It is, without a doubt, the most effective strategy for computer vision jobs. These deep learning robots, which have been doing admirably, require a lot of fuel, which is data. Our model performs better when there is more tagged data available. Google has even experimented with the premise of much more material leading to greater performance on a massive scale, with a sample of 300 million photos! When using a Deep Learning approach in a specific implementation, it must be fed data on a regular basis in order to improve its performance.
Feature Extraction: We utilized efficient net B3 to extract the features in this project. The technique by which QAs verify if a software program is behaving in accordance with the pre requirements is known as functional testing. It employs black-box testing methodologies, wherein the tester is unaware of the internal logic of the system. Functional testing is solely concerned with ensuring that a system functions as expected. It's critical to ensure that perhaps the Block chain Application works as planned while also ensuring that the complete ecosystem works as well. This is where one-of-a-kind scenarios are put to the test.

Tokenization: In Natural Language Processing, tokenization is a typical activity (NLP). Both classic NLP approaches like Indicator Measures and Sophisticated Deep Training architectures like Transformers rely on this phase. Tokenization is the process of breaking down a large chunk of data into smaller tokens. Tokens can be words, characters, or sub words in this case. Tokenization can thus be divided into three categories: word, character, and sub word (n-gram characters) tokenization. The most frequent approach of processing raw text is at the token level, because token are the basic building block of Natural Language.

Algorithm Training: During tokenization, it would be put into the RNN algorithm for training. One of the most fundamental structures is the recurrent neural network (RNN). RNNs are the inspiration for many of today's advanced designs. A major aspect of an RNN is that, unlike a standard feed forward neural network, it incorporates feedback connections. This feedback loop enables the RNN to represent the impacts of early portions of the sequencing on later sections of the series, which is a critical aspect when modeling sequences. RNN architectures come in a variety of shapes and sizes. The interaction within the network is one of the fundamental differences between the architectures. RNNs are typically "unfolded" in period and educated using back-propagation across time, in which the same set of weights is utilized for a layer across successive time steps and updated using gradients, much like the back-propagation algorithm.

Prediction of scene and audio output: A live scenario is taken via the camera once the model has been trained with the algorithm. This photographed scene will be identified, and a modeling file will be created as a result. The scene's movements will be predicted, and a commentary will be generated based on the scene. Following the prediction, the user will receive an auditory object depends on the caption, allowing them to recognize the movements taking place around them.

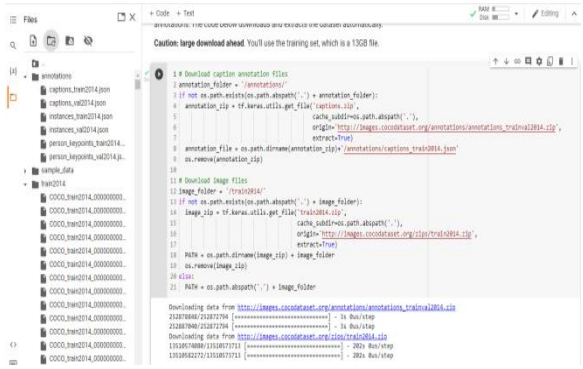


IV. RESULTS AND DISCUSSION

The practical results acquired while carrying out the project are discussed in this chapter. The hardware structure of our proposed approach of the system is shown in Figure. It depicts the suggested method's hardware block diagram. The Raspberry pie is used to demonstrate the system's result and assistance to blind individuals.

Figure Hardware Diagram of our Proposed method

To begin, we can divide our program into components of implementation that have already been completed. The practice of collecting picture caption datasets is known as dataset collection. The dataset for the project has been collected, as shown in the diagram below:



Following pre-processing, the feature extraction procedure consists of numerous convolution layers,

max-pooling, and a transfer functions. Efficient-NetB3 architecture will be used to extract the features. The retrieved features using EfficientNetB3 are shown in the diagram below.

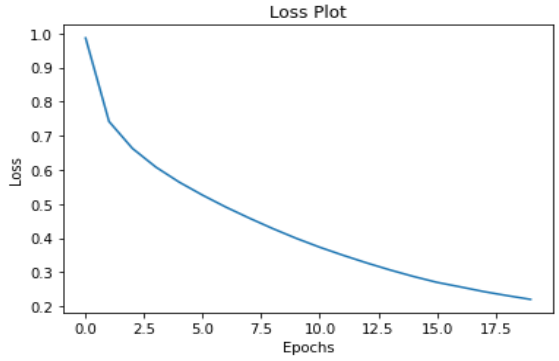
```
<tf.Tensor: shape=(16, 81, 1536), dtype=float32, numpy=
array([[[ 3.0222054 , -0.19645698, -0.2157049 , ...,  0.01563778,
         -0.1171177 , -0.17346299],
        [ 2.3809533 , -0.16479115, -0.09443058, ...,  1.1427802 ,
         -0.21753368, -0.27796447],
        [ 1.2134335 , -0.13580592, -0.12243932, ...,  1.722553  ,
         -0.25641885, -0.25629777],
        ...,
        [-0.23769625, -0.20917192, -0.12009052, ..., -0.25629568,
         -0.27204078, -0.2643148 ],
        [-0.25624245, -0.25263324, -0.16926424, ..., -0.2782392 ,
         -0.2411697 , -0.27770087],
        [-0.27149257, -0.27620935, -0.24727543, ..., -0.26705113,
         -0.17990382, -0.2741223 ]],

       [[ 0.8291164 ,  1.4152273 , -0.27841523, ..., -0.2056691 ,
         -0.0648527 , -0.27737674],
        [ 2.0562167 ,  2.0905025 , -0.27029803, ...,  0.162285  ,
         -0.15099198, -0.25881913],
        [ 1.8235472 ,  1.587184  , -0.2497857 , ..., -0.11006434,
         -0.169515  , -0.25980118],
        ...,
        [ 0.5055626 ,  0.40773052, -0.27745295, ..., -0.25190958,
         -0.24964881, -0.12933512],
        [ 0.21281683, -0.05790276, -0.27810192, ..., -0.27751723,
         -0.19192086,  0.9206407 ],
        [-0.17503569, -0.27758336, -0.26463798, ..., -0.26287085,
         -0.17614678,  0.90730304]],
```

After collecting features from the dataset, the datasets are trained using a deep learning method such as the Recurrent Neural Network.

```
    if epoch % 5 == 0:
      ckpt_manager.save()

    print(f'Epoch {epoch+1} Loss {total_loss/num_steps:.6f}')
    print(f'Time taken for 1 epoch {time.time()-start:.2f} sec\n')
```

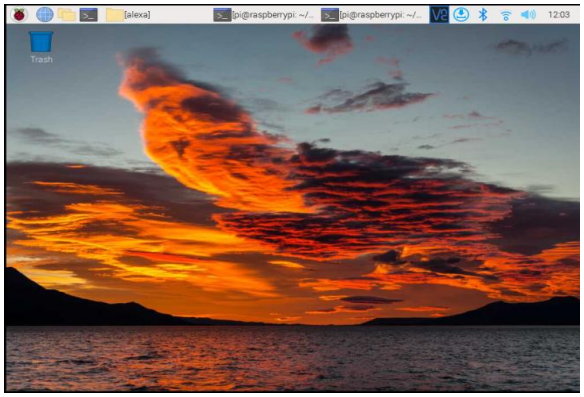
```
Epoch 1 Batch 0 Loss 1.9054
Epoch 1 Batch 100 Loss 1.0788
Epoch 1 Batch 200 Loss 0.9712
Epoch 1 Batch 300 Loss 0.8087
Epoch 1 Loss 0.986307
Time taken for 1 epoch 169.96 sec

Epoch 2 Batch 0 Loss 0.8091
Epoch 2 Batch 100 Loss 0.7494
Epoch 2 Batch 200 Loss 0.7185
Epoch 2 Batch 300 Loss 0.7824
Epoch 2 Loss 0.741672
Time taken for 1 epoch 70.97 sec

Epoch 3 Batch 0 Loss 0.6530
Epoch 3 Batch 100 Loss 0.6692
Epoch 3 Batch 200 Loss 0.6711
Epoch 3 Batch 300 Loss 0.6448
Epoch 3 Loss 0.662739
Time taken for 1 epoch 70.97 sec
```

The graph below shows a decrease in loss as the number of epochs increases during training.

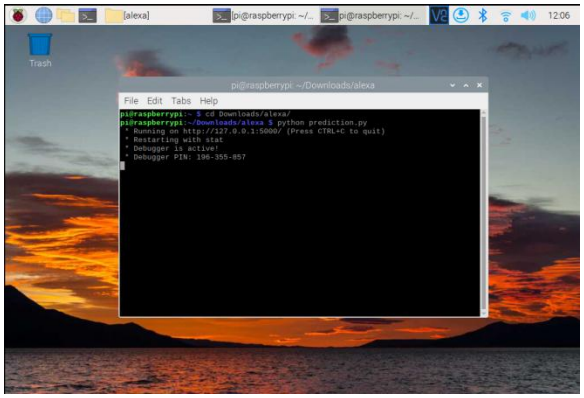


The image caption validation is shown in the diagram below.

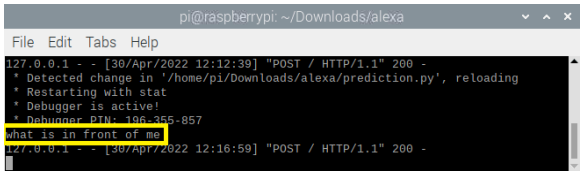
The image below explains how to connect to a Raspberry Pi using a remote desktop connection.



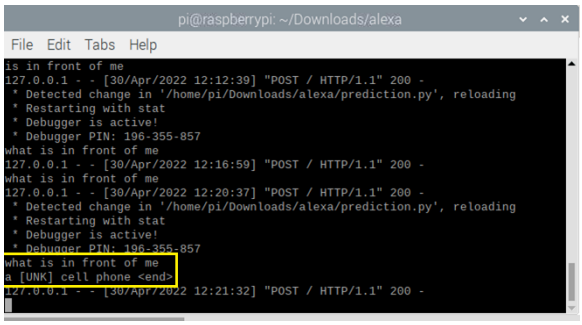
After creating a skill in the Amazon Developer Console, the picture captioning code must be run on the Raspberry Pi. The following diagram depicts the execution of code.



The speech will be sent to the Raspberry Pi as a text once you give a voice control using Amazon Echo. The audio to text output is shown in the diagram below.



Output of our Proposed Method



Final prediction output

V.CONCLUSION

Thus, in order to assist the blind, we developed a virtual camera system where the blind person could perhaps carry a device with them that can guide them about the all around it environment which helps them contribute a safer life while also increasing awareness of the surroundings using advanced graphics captioning techniques. We will assess the scope of work in the medical area in the near future, and try to improve this methodology in other fields. There are also more opportunities to develop or adapt this project in a variety of ways. As a result, this research has a bright future ahead of it in terms of guiding blind people through their surroundings.

REFERENCE

[1] L. Claussmann, M. Revilloud, D. Gruyer, and S. Glaser, “A review of motion planning for highway autonomous driving,” IEEE Trans. Intell. Transp. Syst., vol. 21, no. 5, pp. 1826–1848, May 2020.

[2] Y. Ma, Z. Wang, H. Yang, and L. Yang, “Artificial intelligence applications in the development of autonomous vehicles: A survey,” IEEE/CAA J. Automatica Sinica, vol. 7, no. 2, pp. 315–329, Mar. 2020.

[3] K. Menda et al., “Deep reinforcement learning for event-driven multiagent decision processes,” IEEE Trans. Intell. Transp. Syst., vol. 20, no. 4, pp. 1259–1268, Apr. 2021.

[4] M. Liu, F. Zhao, J. Niu, and Y. Liu, “Reinforcement Driving: Exploring trajectories and navigation for autonomous vehicles,” IEEE Trans. Intel. Transp. Syst., vol. 22, no. 2, pp. 808–820, Feb. 2021

[5] Y. Xing et al., “Advances in vision-based lane detection: Algorithms, integration, assessment, and perspectives on ACP-based parallel vision,”. May 2021.

[6] Iqbal, A., Farooq, U., Mahmood, H., & Asad, M. U. (2020, December). A low cost artificial vision system for visually impaired people.

[7] Kalaivani, K., RR, Y. V., PAA, M. B., CH, K. B., & Mahalaxmi, R. (2022, February). An Artificial Eye for Blind People. In 2022 IEEE Delhi Section Conference (DELCON) (pp. 1-5). IEEE.

[8] Caraiman, S., A. (2020). Computer vision for the visually impaired: the sound of vision system. In Proceedings of the IEEE International Conference on Computer Vision Workshops (pp. 1480-1489).

[9] Hwang, S. J., Ravi, S. N., Tao, Z., Kim, H. J., Collins, M. D., & Singh, V. (2018). Tensorize, factorize and regularize: Robust visual relationship learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1014-1023).

[10] Lu, C., Krishna, R., Bernstein, M., & Fei-Fei, L. (2020, October). Visual relationship detection with language priors. In European conference on computer vision (pp. 852-869). Springer, Cham.