Ilkogretim Online - Elementary Education Online, 2022; Vol 21 (Issue 2): pp. 354-368
http://ilkogretim-online.org
doi: 10.17051/ilkonline.2021.02.25

Smart Vision System For Visually Impaired People

Dr. D.Balasubramanian[1]-Professor - Department of Master of Computer Application.

M.Kumaran[2] –Asst. Professor - Department of Computer Science Engineering.

Dr.V.Seedha Devi[3]-Associate Professor, Department of Information Technology.

Mr.S.Alangaram[4]**,** Asst. Professor, Department of Information Technology.

Jaya Engineering College, Thiruninravur, Tamil Nadu -602024.

[1]balatab3@gmail.com, [2]kumaran@jec.ac.in, [3]seethaitjec@gmail.com, [4]alangaram1985@gmail.com

ABSTRACT

Globally, 1 billion people have a vision impairment that could have been prevented or has yet to be addressed. This 1 billion people includes those with moderate or severe distance vision impairment or blindness due to unaddressed refractive error (123.7 million), cataract (65.2 million), glaucoma (6.9 million), corneal opacities (4.2 million), diabetic retinopathy (3 million), and trachoma (2 million), as well as near vision impairment caused by unaddressed presbyopia (826 million). In terms of regional differences, the prevalence of distance vision impairment in low- and middle-income regions is estimated to be four times higher than in high-income regions. With regards to near vision, rates of unaddressed near vision impairment are estimated to be greater than 80% in western, eastern and central sub-Saharan Africa, while comparative rates in high-income regions of North America, Australasia, Western Europe, and of Asia-Pacific are reported to be lower than 10%. Population growth and ageing are expected to increase the risk that more people acquire vision impairment. In this paper in order to facilitate the blind we develop an artificial vision system in which the blind person can have a device with them which can guide the blind about the surrounding environment and help them lead a safer life as well increase awareness about the surroundings. This is been achieved by using advanced image captioning techniques implementing efficient net algorithms and tokenization methods where the scenes with different captions are learned by the machine. Whenever an image is captured via the camera are been recognized and predicted by the machine. After the prediction, it is been sent as an audio output to the user which can help them identify the scene happening around. Thus, with the help of this paper provide an artificial vision to the blind, which can help them gain confidence while travelling alone.

1. INTRODUCTION

The human eye is like a camera that collects, focuses, and transmits light through a lens to create an image of its surroundings. In a camera, the image is created on film or an image sensor.
In the eye, the image is created on the retina, a thin layer of light-sensitive tissue at theback of the eye

Like a camera, the human eye controls the amount of light that enters the eye. The iris (the colored circular part of the eye) controls the amount of light passing through the pupil. It closesup the pupil in bright light and opens it wider in dim light. The cornea is the transparent, protective surface of the eye. It helps focus light, as does the lens, which sits just behind the iris. When light enters the eye, the retina changes the light into nerve signals. The retina then sends these signals along the optic nerve (a cable of more than 1,000,000 nerve fibers) to the brain. Without a retina or optic nerve, the eye can't communicate with the brain, making visionimpossible.

Many people have some type of visual problem at some point in their lives. Some can no longer see objects far away. Others have problems reading small print. These types of conditions are often easily treated with eyeglasses or contact lenses.

But when one or more parts of the eye or brain that are needed to process images become diseased or damaged, severe or total loss of vision can occur. In these cases, vision can't be fully restored with medical treatment, surgery, or corrective lenses like glasses or contacts.

The American Foundation for the Blind estimates that 10 million people in the United States are visually impaired. Visual impairment is a term experts use to describe any kind of vision loss, whether it's someone who cannot see at all or someone who has partial vision loss.

Some people are completely blind, but many others have what's called legal blindness. They haven't lost their sight completely but have lost enough vision that they'd have to stand 20 feetfrom an object to see it as well as someone with perfect vision could from 200 feet away.

2. TECHNOLOGIES USED:

Deep Learning

Deep learning is a computer software that **mimics the network of neurons in a brain**. It is a subset of machine learning and is called deep learning because it makes use of deep **neural networks.**

Deep learning algorithms are constructed with connected layers.

The first layer is called the Input Layer
The last layer is called the Output Layer
All layers in between are called Hidden Layers.

The word deep means the network joinneurons in more than two layers.

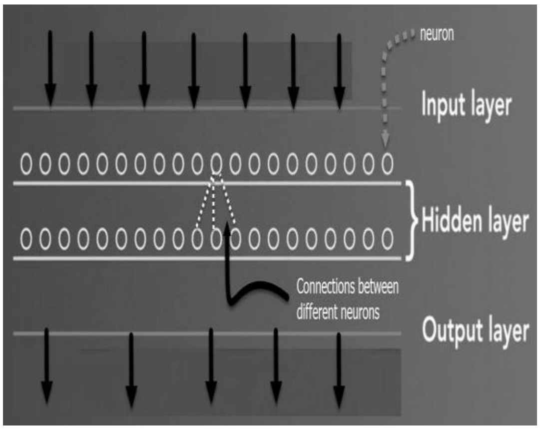

Figure 1 Deep Learning Layers

Each Hidden layer is composed of neurons. The neurons are connected to each other. The neuron will process and then propagate the input signal it receives the layer above it. The strength of the signal given the neuron in the next layer depends on the weight, bias and activation function.

The network consumes large amounts of input data and operates them through multiple layers; the network can learn increasingly complex features of the data at each layer.

2.2 IMPORTANCE OF DEEP LEARNING

Deep learning is a powerful tool to make prediction an actionable result. Deep learning excels in pattern discovery (unsupervised learning) and knowledge-based prediction. Big datais the fuel for deep learning. When both are combined, an organization can reap unprecedented results in term of productivity, sales, management, and innovation.

Deep learning can outperform traditional method. For instance, deep learning algorithms are 41% more accurate than machine learning algorithm in image classification, 27 % more accurate in facial recognition and 25% in voice recognition.

2.3 DEEP LEARNING PROCESS

A deep neural network provides state-of-the-art accuracy in many tasks, from object detection to speech recognition. They can learn automatically, without predefined knowledge explicitly coded by the programmers.

Figure 2 Deep Learning Process

To grasp the idea of deep learning, imagine a family, with an infant and parents. The toddler points objects with his little finger and always says the word 'cat.' As its parents are concerned about his education, they keep telling him 'Yes, that is a cat' or 'No, that is not a cat.'The infant persists in pointing objects but becomes more accurate with 'cats.' The little kid, deep down, does not know why he can say it is a cat or not. He has just learned how to hierarchies' complex features coming up with a cat by looking at the pet overall and continue to focus on details such as the tails or the nose before to make up his mind.

A neural network works quite the same. Each layer represents a deeper level of knowledge, i.e., the hierarchy of knowledge. A neural network with four layers will learn morecomplex feature than with that with two layers.

The learning occurs in two phases.

The first phase consists of applying a nonlinear transformation of the input and create astatistical model as output. The second phase aims at improving the model with a mathematical method known asderivative. The neural network repeats these two phases hundreds to thousands of time until it hasreached a tolerable level of accuracy. The repeat of this two-phase is called an iteration.

2.4 CLASSIFICATION OF NEURAL NETWORKS

Shallow neural network: The Shallow neural network has only one hidden layer between theinput and output.

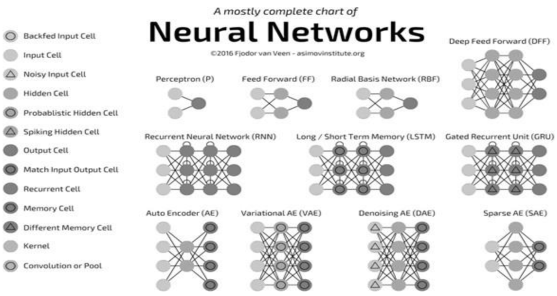

Figure 3 Types of Deep Learning Networks

Deep neural network: Deep neural networks have more than one layer. For instance, GoogleLeNet model for image recognition counts 22 layers.

Nowadays, deep learning is used in many ways like a driverless car, mobile phone, Google Search Engine, Fraud detection, TV, and so on.

2.5 CONVOLUTIONAL NEURAL NETWORKS (CNN)

CNN is a multi-layered neural network with a unique architecture designed to extract increasingly complex features of the data at each layer to determine the output. CNN's are well suited for perceptual tasks

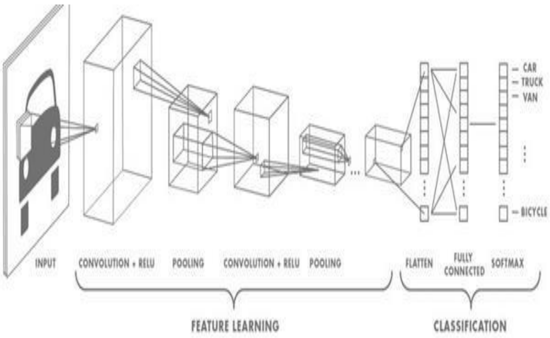

Figure.4 CNN

CNN is mostly used when there is an unstructured data set (e.g., images) and thepractitioners need to extract information from it

For instance, if the task is to predict an image caption:

The CNN receives an image of let's say a cat, this image, in computer term, is a collection of the pixel. Generally, one layer for the greyscale picture and three layers for a color picture.
During the feature learning (i.e., hidden layers), the network will identify unique features, for instance, the tail of the cat, the ear, etc.
When the network thoroughly learned how to recognize a picture, it can provide a probability for each image it knows. The label with the highest probability will become the prediction of the network.

A Convolutional Neural Network (**CNN**, or **ConvNet**) are a special kind of multi-layer neural networks, designed to recognize visual patterns directly from pixel images with minimal pre-processing. The **Image Net** project is a large visual database designed for use in visual object recognition software research. The Image Net project runs an annual software contest, the **Image Net Large Scale Visual Recognition Challenge (ILSVRC)**, where software programs compete to correctly classify and detect objects and scenes.

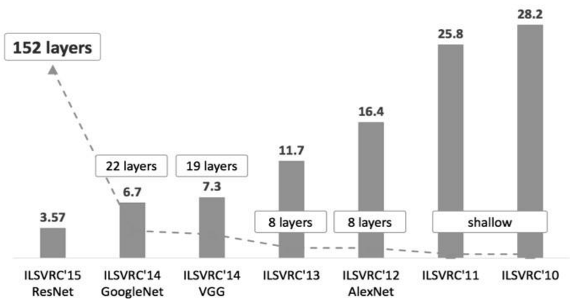

Figure 5 Types of CNN

3. Collecting and Sending Information

This means sensors. Sensors could be temperature sensors, motion sensors, moisture sensors, air quality sensors, light sensors, you name it. These sensors, along with a connection, allow usto automatically collect information from the environment which, in turn, allows us to make more intelligent decisions.

Figure 6 Soil moisture sensor

(Soil moisture sensor) On the farm, automatically getting information about the soil moisturecan tell farmers exactly when their crops need to be watered. Instead of watering too much (which can be an expensive over-use of irrigation systems and environmentally wasteful) or watering too little (which can be an expensive loss of crops), the farmer can ensure that cropsget exactly the right amount of water. More money for farmers and more food for the world! Just as our sight, hearing, smell, touch, and taste allow us, humans, to make sense of the world,sensors allow machines to make sense of the world.

Receiving and Acting on Information

We're all very familiar with machines getting information and then acting. Your printerreceives a document and it prints it. Your car receives a signal from your car keys and the doors open. The examples are endless.

Whether it's a simple as sending the command "turn on" or as complex as sending a 3D model to a 3D printer, we know that we can tell machines what to do from far away. So what?

The real power of the Internet of Things arises when things can do both of the above. Things that collect information and send it, but also receive information and act on it.

Doing Both

Let's quickly go back to the farming example. The sensors can collect information about the soil moisture to tell the farmer how much to water the crops, but you don't actually need the farmer. Instead, the irrigation system can automatically turn on as needed, based on how much moisture is in the soil.

You can take it a step further too. If the irrigation system receives information about the weather from its internet connection, it can also know when it's going to rain and decide not to water the crops today because they'll be watered by the rain anyways.

And it doesn't stop there! All this information about the soil moisture, how much the irrigation system is watering the crops, and how well the crops actually grow can be collected and sent to supercomputers that run amazing algorithms that can make sense of all this information.

And that's just one kind of sensor. Add in other sensors like light, air quality, and temperature, and these algorithms can learn much more. With dozens, hundreds, thousands of farms all collecting this information, these algorithms can create incredible insights into how to make crops grow the best, helping to feed the world's growing population.

4. EXISTING SYSTEM

In this traffic-scene-modeling study, they propose an image-captioning network which incorporates element attention into an encoder-decoder mechanism to generate more reasonable scene captions. A visual-relationship-detecting network is also developed to detect the relative positions of object pairs. Firstly, the traffic scene elements are detected and segmented according to their clustered locations. Then, the image-captioning network is applied to generate the corresponding description of each traffic scene element. The visual- relationship-detecting network is utilized to detect the position relations of all object pairs in the sub region. The static and dynamic traffic elements are appropriately selected and organizedto construct a 3D model according to the captions and the position relations. The reconstructed 3D traffic scenes can be utilized for the offline test of unmanned vehicles. The evaluations and comparisons based on the TSDmax, KITTI and Microsoft's COCO datasets demonstrate the effectiveness of the proposed framework.

4.2. DISADVANTAGES OF EXISTING SYSTEM

Only the datasets are used for validating the proposed model.
Not real time applicable as the accuracy obtained in less.
Only the traffic scene prediction is been done which makes the system inefficient forother use.
DBSCAN algorithm does not work well in case of high dimensional data.

5. PROPOSED SYSTEM

Globally, 1 billion people have a vision impairment that could have been prevented or has yet to be addressed. This 1 billion people includes those with moderate or severe distance vision impairment or blindness due to unaddressed refractive error (123.7 million), cataract (65.2 million), glaucoma (6.9 million), corneal opacities (4.2 million), diabetic retinopathy (3 million), and trachoma (2 million), as well as near vision impairment caused by unaddressed presbyopia (826 million). In terms of regional differences, the prevalence of distance vision impairment in low- and middle-income regions is estimated to be four times higher than in high-income regions. With regards to near vision, rates of unaddressed near vision impairment are estimated to be greater than 80% in western, eastern and central sub-Saharan Africa, while comparative rates in high-income regions of North America, Australasia, Western Europe, and of Asia-Pacific are reported to be lower than 10%. Population growth and ageing are expected to increase the risk that more people acquire vision impairment. In this project in order to facilitate the blind develop an artificial vision system in which the blind person can have a device with them which can guide the blind about the surrounding environment and help them lead a safer life as well increase awareness about the surroundings. This is been achieved by using advanced image captioning techniques implementing efficient net algorithms and tokenization methods where the scenes with different captions are learned by the machine. Whenever an image is captured via the camera are been recognized and predicted by the machine. After the prediction, it is been sent as an audio output to the user which can help them identify the scene happening around. Thus, with the help of this project provide an artificial vision to the blind, which can help them gain confidence while travelling alone.

5.1 ADVANTANGES OF PROPOSED SYSTEM

Cheap and effective solution for the blind to become aware of the surroundings.
Effective scene prediction model is developed.
Provides an artificial vision to the visually challenged people.

SYSTEM ARCHITECTURE

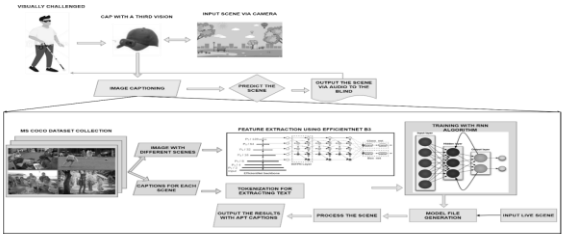

Figure 7 Proposed System Architecture

7. WORKING

In this project in order to facilitate the blind have developed an artificial vision system with the help of deep learning techniques. Initially COCO dataset is collected to train the model, before training the model pre-processing of dataset is done so that the dataset can be directly applied to the deep Learning algorithms for further processes. After pre-processing feature extraction in done using efficient net algorithm. Once the feature extraction is completed the model is trained using RNN algorithm. Recurrent neural networks (RNN) are the state of the art algorithm for sequential data and are used by Apple's Siri and Google's voice search. It is the first algorithm that remembers its input, due to an internal memory, which makes it perfectly suited for machine learning problems that involve sequential data. It is one of the algorithms behind the scenes of the amazing achievements seen in deep learning over the past few years. Finally whenever the user wants to know about the surrounding movements, they can ask the model to recognize what is happening around by saying any predefined keyword like “what is happening in front of me”, so that the model can start recognizing the surrounding movements. This is achieved by using advanced image captioning

8. Training with RNN Algorithm

After tokenization, it will be fed for training with the RNN algorithm.

A recurrent neural network (RNN) is one of the basic architectures. Many of the advanced architectures today are inspired by RNNs. The key feature of an RNN is that the network has feedback connections, unlike a traditional feed forward neural network. This feedback loop allows the RNN to model the effects of the earlier parts of the sequence on the later part of the sequence, which is a very important feature when it comes to modeling sequences.

There are many different architectures for RNNs. One of the key differences in the architectures is the feedback within the network. Typically, RNNs can be “unfolded” in time and trained using back-propagation through time, where the same set of weights is used for a layer across multiple time steps and updated using the gradients similar to the back- propagation algorithm. A generic structure of an RNN.

A different method of deep learning is recurrent neural networks. Recurrent neural networks are neural networks that, in addition to its inputs, use an internal state to perform a task. Since the new internal state is calculated from the old internal state and the input, it can be understood as one part of the output of the neural network. If we understand it in this way, recurrent neural networks get part of their output as input for the next time step. This recurrent behavior gave recurrent neural networks their name. The information that is stored in this internal state is automatically selected by the algorithm. Recurrent neural networks are, hence,a deep learning approach. The internal state, often referred to as "memory," makes them useful for sequential data, such as time series, audio, or text, and gives them the ability to handle inputs of different sequence length.

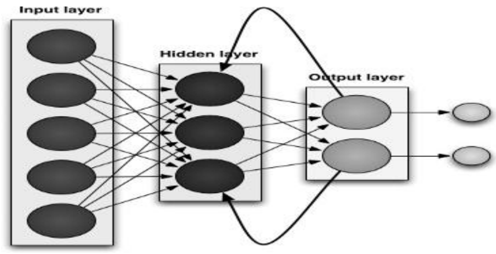

Figure 8 RNN DFD

9. Prediction of scene and audio output

After training the model with algorithm, a live scene is captured via the camera. This capturedscene will be recognized and the output model file will be generated.

The movements in the scene will be predicted and

a caption will be generated according to thescene.

After the prediction, an audio output based on the caption will be sent to the user through thisthe user can identify the movements happening around them.

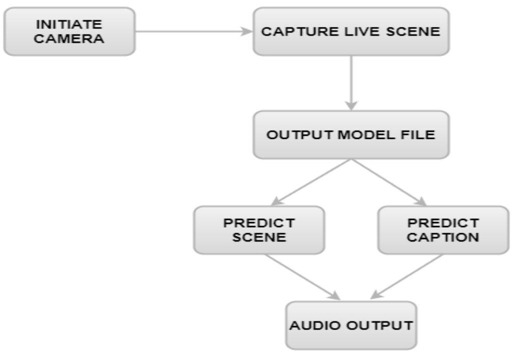

Figure 9 Prediction of scene and audio output DFD

10. RESULTS AND DISCUSSIONS

To begin with, testing of the trained model, can split our project into modules ofimplementation that is done.

Dataset collection involves the process of collecting image caption dataset.
The dataset has been collected for the project and the below figure can be seen as follows:

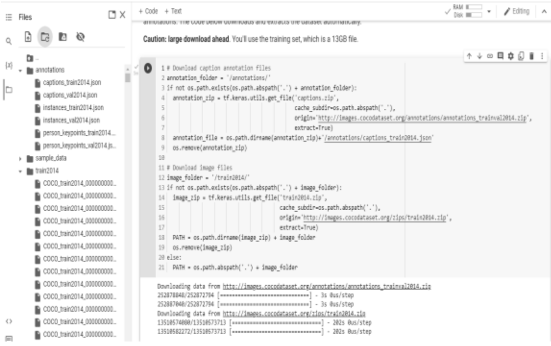

Figure 10 Dataset Collection

Then these datasets are pre-processed from convert the images into required size formatso that it can be made ready for training with the model.
The below figure shows the simple pre-processing techniques used for image resizing.

Figure 11 Simple Pre-processing

After pre-processing, Feature extraction process it includes several convolution layers followedby max-pooling and an activation function. For extracting the features we will be using EfficientNetB3architecture.

```
<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]],
```

Figure 12 Feature Extraction

Then the captions are pre-processed using tokenization. Tokenization is the process ofconverting the words into integers.

```
tf.Tensor(b'<start> A plate of meat, fries and vegetable salads <end>', shape=(), dtype=string)
tf.Tensor(b'<start> A blue and white plate topped with meat and vegetables. <end>', shape=(), dtype=string)
tf.Tensor(b'<start> a plate of fish , fries and coleslaw. <end>', shape=(), dtype=string)
tf.Tensor(b'<start> Pork chops, french fries, and cooked vegetables on a blue plate. <end>', shape=(), dtype=string)
tf.Tensor(b'<start> a blue and white hold two pieces of meat, fries, and salad. <end>', shape=(), dtype=string)

tf.Tensor(
[   3    2   56    6  889  984   10  746    1    4    0    0    0    0    0    0    0    0
    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
    0    0    0    0    0    0    0    0    0    0    0    0    0    0], shape=(50,), dtype=int64)
tf.Tensor(
[   3    2   51   10   21   56  308    9  390   10  636    4    0    0    0    0    0    0
    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
    0    0    0    0    0    0    0    0    0    0    0    0    0    0], shape=(50,), dtype=int64)
tf.Tensor(
[    3     2    56     6  2016  2785   984    10     1     4     0     0     0     0
     0     0     0     0     0     0     0     0     0     0     0     0     0     0
     0     0     0     0     0     0     0     0     0     0     0     0     0     0
     0     0     0     0     0     0     0     0], shape=(50,), dtype=int64)
tf.Tensor(
[    3  3946     1  1047  2418    10   876   317     5     2    51   196     4     0
     0     0     0     0     0     0     0     0     0     0     0     0     0     0
     0     0     0     0     0     0     0     0     0     0     0     0     0     0
     0     0     0     0     0     0     0     0], shape=(50,), dtype=int64)
tf.Tensor(
[    3     2    51    10    21   959    16   642     6   889  2418    10  1968     4
     0     0     0     0     0     0     0     0     0     0     0     0     0     0
     0     0     0     0     0     0     0     0     0     0     0     0     0     0
     0     0     0     0     0     0     0     0], shape=(50,), dtype=int64)
```

Figure 13 Tokenization

After extracting features from dataset, training is performed with the Deep learningalgorithm such as Recurrent Neural Network is used for training the datasets.

```
    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
```

Figure 14 Training using Recurrent Neural Network

The below figure shows the extracted features using EfficientNetB3.

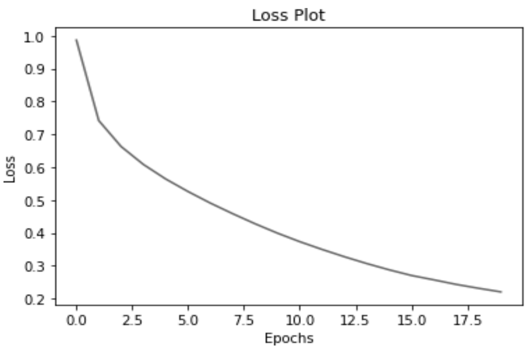

Figure 15 Feature Extraction

11. CONCLUSION:

In this project in order to facilitate the blind develop an artificial vision system in whichthe blind person can have a device with them which can guide the blind about the surrounding environment and help them lead a safer life as well increase awareness about the surroundings. In the existing system, they proposed an image-captioning network which incorporates elementattention into an encoder-decoder mechanism to generate more reasonable scene captions. But only the traffic scene prediction is been done which makes the system inefficient for other use. In order to avoid this, have developed a model using advanced image captioning techniques implementing efficient net algorithms and tokenization methods where the scenes with differentcaptions are learned by the machine.
Thus, with the help of this project provide an artificial vision to the blind, which can help them gain confidence while travelling alone.

12. FUTURE WORK:

In the coming future, will review the scope of this of the project in the medical field, and try to enhance this technique in other fields. And there are more chance to develop or convert this project in many ways. Thus, this project has an efficient scope in coming future to guide the blind people about the surrounding environment.

13. RESULT:

The below figure shows the login page of the mobile app.

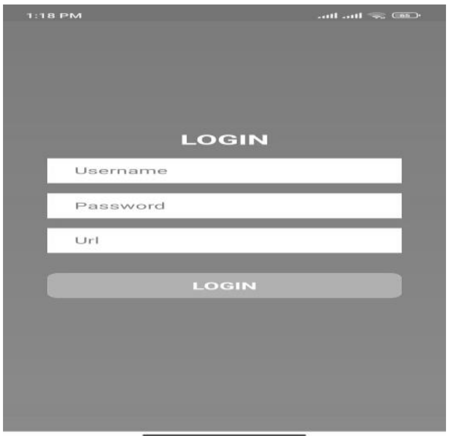

Figure 16 Login Page

In this app the user can upload an image to get the output

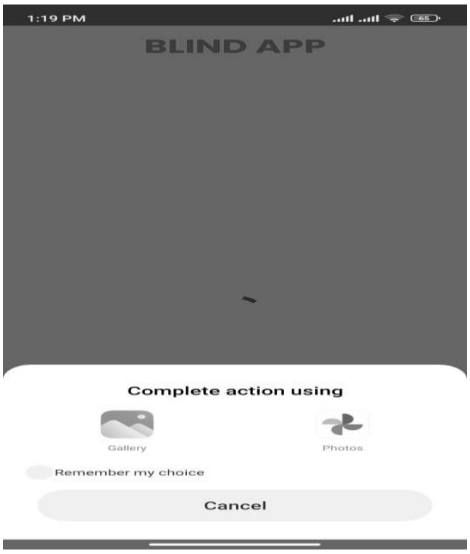

Figure 17 Image Uploading Page

After training the model, it will be waiting for our command to start capturing the movements around us. Whenever give a command similar to "what is in front of me" it will capture the movement in-front of us and starts predicting, after prediction it will return the movement as a caption which will be outputted as an audio.

Figure 18 Prediction

Whenever an image is captured via camera those images are recognized and predicted by the model. After prediction, the model will generate a caption based on the captured scene and an audio output based on the caption will be sent to the user through this the user can identify the movements happening around them. So that the user can identify the scene happening around them. Thus, with the help of this project provide an artificial vision to the blind, which can help them gain confidence while travelling alone.

REFERENCE

LIST OF BOOKS

[1] Xiaojuan Cheng; Jiwen Lu; Bo Yuan; Jie Zhou, “Face Segmentor-Enhanced Deep Feature Learning for Face Recognition”, IEEE Transactions on Biometrics, Behavior, and Identity Science [Vol no: 1, 2019]

[2] Lei Zhang; Ji Liu; Bob Zhang; David Zhang; Ce Zhu, “Deep Cascade Model-Based Face Recognition: When Deep-Layered Learning Meets Small Data”, IEEE Transactions on Image Processing [Vol no: 29, 2019]

[3] Yuqi Zhang; Yongzhen Huang; Shiqi Yu; Liang Wang, “Cross-View Gait Recognition by Discriminative Feature Learning”, IEEE Transactions on Image Processing [Vol no: 29, 2019]

[4] Christian Galea; Reuben A. Farrugia, “Matching Software-Generated Sketches to Face Photographs With a Very Deep CNN, Morphed Faces, and Transfer Learning”, IEEE Transactions on Information Forensics and Security [Vol no: 13, 2018]

[5] Ran He; Xiang Wu; Zhenan Sun; Tieniu Tan, “Wasserstein CNN: Learning Invariant Features for NIR-VIS Face Recognition”, IEEE Transactions on Pattern Analysis and MachineIntelligence [Vol no: 41, 2019]

[6] Jiwen Lu; Gang Wang; Jie Zhou, “”,Simultaneous Feature and Dictionary Learning forImage Set Based Face Recognition, https://ieeexplore.ieee.org/document/7944673

[7] Gaurav Goswami; Mayank Vatsa; Richa Singh, “Face Verification via LearnedRepresentation on Feature-Rich Video Frames”, https://ieeexplore.ieee.org/document/7850956

[8] Jiwen Lu; Junlin Hu; Yap-Peng Tan, “Discriminative Deep Metric Learning for Faceand Kinship Verification”,

[10] Https://ieeexplore.ieee.org/document/
7953665

[11] https://aws.amazon.com/what-is-aws

[12]https://www.youtube.com/watch?v=2MMFNLJkjOo

[13] https://azure.microsoft.com/en-in/overview/what-is-azure/

[14]https://developer.amazon.com/public/solutions/alexa/alexa-skills-kit/docs/developing-an-alexa-skill-as-a-lambda-function

[15] https://sbstjn.com/custom-alexa-skill-for-amazon-echo-with-lambda.html

[16]http://docs.aws.amazon.com/cli/latest/userguide/installing.html

[17] Kher Chaitrali S., Dabhade Yogita A., Kadam Snehal K., Dhamdhere Swati D., Deshpande Aarti V. An Intelligent Walking Stick for the Blind. Published 2015.International Journal of Engineering Research and General Science Volume 3, Issue 1, January-February, 2015.ISSN 2091-2730.

[18] Piyush Chanana1, Rohan Paul2, M Balakrishnan1 and PVM Rao1 Assistive technology solutions for aiding travel of pedestrians with visual impairment.Journal of Rehabilitation and Assistive Technologies Engineering
Volume 4:1–16! The Author(s) 2017 Reprintsandpermissions:sagepub.co.uk/journalsPermissions.nav DOI.

[19] Abdelsalam (Sumi) Helal,Steven Edwin Moore,Balaji Ramachandran.Drishti: An Integrated Navigation System for Visually Impaired and Disabled. Computer & Information Science & Engineering, University of Florida, Gainesville, FL-32611.
[20] Vigneshwari, Vimala, Sumithra.Sensor Based Assistance System for Visually Impaired.International Journal of Engineering Trends and Technology (IJETT) –Volume 4
Issue 10 - Oct 2013.
[21] RAMA MURTHY. N, P. N. SUDHA. Smart Navigation System For Visually Challenged People.International Journal of Industrial Electronics and Electrical Engineering, ISSN: 2347-6982 Special Issue, Sep.-2016.
[22] Sneha Venkateswar, Ninad Mehendale.Intelligent Belt for the Blind.International Journal of Scientific &
Engineering Research, Volume 3, Issue 11, November20121 ISSN 2229-5518.