IMAGE CAPTIONING AID FOR VISUALLY IMPAIRED PEOPLE USING CONVOLUTIONAL NEURAL NETWORK

Mr.Venkatesan.M[1], Anbarasu.A[2],Balajikrishnan.G.S[3],Harishbabu.A[4],Jothieshwaran.M[5]

[1]Assistant Professor, [2,3,4,5]UG Scholors, Department of Electronics and Communication Engineering, Adhiyamaan College of Engineering, Hosur, Tamil Nadu, India.

venkateshace82@gmail.com[1], aanbuanbuanbarasu@gmail.com[2], gsbalajikrishnan@gmail.com[3], harisharul09@gmail.com[4], jothieshwaran2599@gmail.com[5]

Abstract—*Getting comfortable with the difficulties that visual deficiency makes can assist located individuals with understanding their issues and the significance of this undertaking. Our group stepping forward to make an extension between outwardly impeded way of life with typical way of life. The proposed thought is going to make a wearable gadget that can control an outwardly debilitated individual in everyday life. This gadget is mounted a camera the gadget takes the contribution from outside climate as a picture, and produce an important yield reasonable by outwardly debilitated individual. We are actualizing picture subtitling calculation utilizing Convolutional Neural Network (CNN). In the main stage, the venture will actually want to create yield through speaker justifiable by outwardly debilitated person. After effective execution of the primary stage, the task will be redesigned with voice yield utilizing LSTM engineering and a book to discourse generator.*

Key Words: CPU, TensorFlow, CNN, LSTM, Webcam

1.INTRODUCTION

Therefore, portraying the substance of pictures using typical language is a fundamental and testing task. With the progress in enrolling power alongside the openness of massive datasets, building models that can create captions for an image has gotten possible. Of course, individuals can without a doubt portray the conditions they are in. Given a picture, it's normal for a person to explain a tremendous proportion of experiences concerning this image with a brief glance. But unprecedented improvement has been made in PC vision, tasks, for instance, seeing a thing, action portrayal, picture gathering, property plan additionally, scene affirmation are possible yet it is a for the most part new task to permit a PC to depict an image that is shipped off it as a human-like sentence. For this target of picture recording, considering semantics of pictures should be gotten here and conveyed in the ideal kind of typical lingos. It has a remarkable impact in the veritable world, for instance by supporting ostensibly debilitated people better fathom the substance of pictures on the web.

2.RELATED WORK

In this paper perhaps the most famous profound neural organizations is the Convolutional Neural Network (CNN) is clarified. There are different layers in CNN, for example, convolutional layer, and nonlinearity layer, &pooling layer and completely associated layer also. The CNN has an incredible presentation in AI issues and quite possibly the most widely recognized calculations. In this paper clarify about the profound neural organization calculation long momentary Memory (LSTM). LSTM is neighborhood in both space just as on schedule; the computational intricacy is per season of step and furthermore the weight design portrayal In contrast with other calculation LSTM prompts a lot more effective runs, and learn a lot quicker It's even address mind boggling, counterfeit long delay assignments that have never been tackled by past repetitive organization. Naturally portraying the substance of an image using properly coordinated English sentences is an exceptional testing task, yet it could is something amazingly principal for supporting apparently blocked people Present day phones can take the photographs, which can help in taking incorporating pictures for ostensibly hindered social classes. Here pictures as data can make engravings that can be uproarious enough so that apparently upset can hear, so they can improve sensation of things introducing there enveloping. Here Christoper Elamri uses a CNN model to eliminate features of an image. These features are then dealt with into a RNN or a LSTM model to make a depiction of the image in etymologically right English sentences depicting the ecological components.

3.METHODOLOGY

PC vision has gotten widespread in our overall population, with applications in a couple of fields. In this undertaking, we base on one of the visual affirmation parts of PC vision, i.e picture captioning. As a result of the new degrees of progress in the field of article area, the task of scene depiction in an image has gotten more straightforward. We can make a wearable thing for the outwardly debilitated which will guide them going in the city without the assistance of some other individual. This is done by

interfacing a camera sensor with PC i.e OpenCV Starter Kit, that gets persistent pictures and feed those image edges to CPU. By then a CNN based Image Captioning Algorithm reasoning inside CPU, takes the data picture diagrams as an information data for successfully arranged Convolution Neural Network (CNN). The CNN by then creates a huge yield which is transported off external environment using actuators, that can be recognized by an outwardly debilitated person.

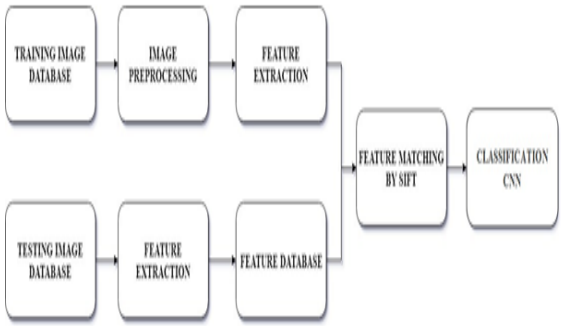

Fig -1: Block diagram of Image Captioning system

4.EXISTING METHOD

Visual Impairment Is Something That Any Person Does Not To Have And It Does Not Have Any Kind Of Temporary Fixes. Visual Impairment Is Nothing But Disability To See Which Makes The Problem Not Fixable By Temporary Means. As per The World Health Organization, 285 Million People Are Visually Impaired Worldwide, Including Over 39 Million Blind People. Living Without One Of The Most Useful Sensory Organ In A Technologically Developing World Where Even The Smallest Piece Of Work Would Require Sight Is Very Difficult. Living In The Era When The Technology Sector Is Booming There Can Be Many Developments Made That Could Improve The Lives Of Visually Impaired. There Are Several Ways In Which The Technology Can Provide An Aid To The Visually Impaired One Way Is By Detecting The Objects From An Image And Providing A Meaningful Caption That Would Be Read Out Loud Which Would Help The Person Using This System

5.PROPOSED SYSTEM

We executed a significant discontinuous designing that normally conveys short portrayals of pictures. Our models use a CNN, which was pretrained on ImageNet, to get pictures features. We by then feed these features into either a vanilla RNN or a LSTM association make a depiction of the image The profound learning methods to the picture subtitle age task. At first concentrate picture highlights utilizing a CNN. Picture subtitle age might actually give dazzle individuals nonstop continuous data. LSTMs could be utilized in mix with CNNs to make an interpretation of Image to Audio.

5.1 CNN-based Image Feature Extractor

For highlight extraction, we utilize a CNN. CNNs have been generally utilized and read for pictures assignments, and are as of now cutting edge strategies for object acknowledgment and recognition [20]. Solidly, for all information pictures, we remove highlights from the fc7 layer of the VGG-16 organization pretrained on ImageNet, which is all around tuned for object discovery. We got a 4096-Dimensional picture highlight vector that we decrease utilizing Principal Component Analysis (PCA) to a 512-Dimensional picture include vector because of computational limitations. We feed these highlights into the primary layer of our RNN or LSTM at the main emphasis

5.2 LSTM-based Sentence Generator

Despite the fact that RNNs have demonstrated fruitful on undertakings like content age and discourse acknowledgment [25, 26], it is hard to prepare them to learn long haul elements. This issue is likely because of the evaporating and detonating angles issue that can come about because of engendering the inclinations down through the numerous layers of the repetitive organizations. LSTM organizations (Figure 3) give an answer by joining memory units that permit the organizations to realize when to fail to remember past secret states and when to refresh covered up states when given new data [24]. At each time-step, we get an information xt ∈ R D and the past secret state ht−1 ∈ R H, the LSTM additionally keeps a H-dimensional cell state, so we likewise get the past cell state ct−1 ∈ R H. The learnable boundaries of the LSTM are a contribution to-covered up framework Wx ∈ R 4HxD, a covered up to covered up lattice ∈ R 4HxH, and an inclination vector b ∈ R 4.

6. EXPERIMENTAL RESULTS

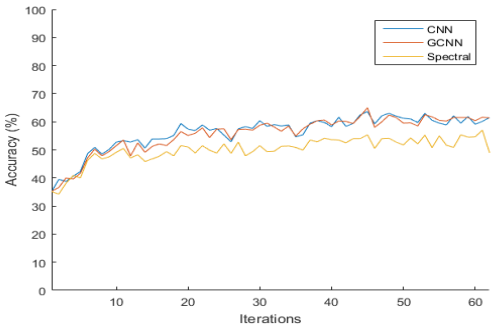

Fig -2: Result of Accuracy Variation

Fig -3: Result of Camera detection

7.CONCLUSION

We Presented This Paper To Help Visually Impaired People By Using Deep Learning Techniques. Systems Like Convolutional Neural Networks (CNN) And Feature Maps That Get Generated Using Such Neural Nets Help Us To Recognize Objects And Later Generate Sentences Using Recurrent Nets Such As Long-Short Term Memory (LSTM). The CNN And LSTM Are Currently The State-Of-The-Art Techniques For Article Detection, Scene Representation And Scene Description Such That The Generated Captions Are Highly Distinct Of The Objects Depicted On The Images. By virtue of The High Quality of the Generated Image Depictions, Visually Impaired People Can Greatly Benefit and Get a Better Sense of Their Surroundings Utilizing Text-To-Speech Technology. The Further Study And Research For This Project Is To Generate Captions For The Live Video To Give Real-Time Understanding And Describing The Scene To The User Instead Of Portraying The Captured Static Images That Can Only Provide Blind People With Information About One Explicit Instance Of Time.

REFERENCES

[1] Micah Hodosh, Peter Young, and Julia Hockenmaier. Framing image description as a ranking task: Data, models and evaluation metrics. J. Artif. Int. Res., 47(1):853–899, May 2013

[2] Andrej Karpathy and Li Fei-Fei. Deep visual-semantic alignments for generating image descriptions. IEEE Trans. Pattern Anal. Mach. Intell., 39(4):664–676, Apr. 2017

[3] Polina Kuznetsova, Vicente Ordonez, Alexander C. Berg, Tamara L. Berg, and Yejin Choi. Collective generation of natural image descriptions. pages 359–368, 2012.

[4] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Raanan, Piotr Dollar, and C. Lawrence ´ Zitnick. editors, Computer Vision – ECCV 2014, pages 740– 755, Cham, 2014. Springer International Publishing.

[5] S. Liu and W. Deng. Very deep convolutional neural network based image classification using small training sample size. pages 730–734, Nov 2015.

[6] Chen, Xinlei and C. Lawrence Zitnick. Learning a Recurrent Visual Representation for Image Caption Generation. CoRR abs/1411.5654 (2014). Web. 19 May 2016

BIOGRAPHY:

Mr. M. Venkatesan,
Assistant Professor,
Electronics And Communication Engineering Department,
Adhiyamaan College of Engineering,
Anna University.