ABSTRACT

The convolution neural network method is among the most widely used deep learning-based object detection methods. This system mainly focuses on image captioning based on research papers. Different Captioning metrics are used for evaluation of the sentences generated by the system. The scores tells about the accuracy of the words obtained. Different methods are compared which tells the efficiency of the LSTM method to be 80%.This provides best results on Visual Genome Dataset. The output generated can have few limitations i.e, the paragraph can contain upto 500 words or 4-5 lines.Hence, this system provides a clear view of how a paragraph is generated from an image. The scope of the paper is limited to LSTM approach only. In future, the scope of the work can be extended so that the system can be more efficiently used by all the researchers. In order to address these problems, a new cluster method for estimating the initial width and height of the predicted bounding boxes has been developed.

INTRODUCTION

Consistently a great deal of photographs is received from the climate, web-based media and magazines. One can know the photograph. As people, it can decided to take photographs without clarifying them, however then again, machines need a preparation picture, and afterward give a composed photograph. Video depictions can be helpful for an assortment of reasons, for instance, supporting the visually impaired through informal utilizing genuine plans to envision the condition of the camera, and advancing general wellbeing by changing over photograph sharing to a public site. furthermore, instant messages. Assist youngsters with mastering things and foster language abilities. The data on every photograph on the world's web is quick and permits you to look for and show definite photographs. Picture depictions come in different bundles in biomedicine, trade, online pursuit, boats and considerably more. It is feasible to remark on photographs on open locales like Instagram and Facebook, and the principle reason for this review is to have less insight and a more profound learning style. This paper utilizes two systems to make pictures: CNN and LSTM.

OBJECTIVE

The objective of our project is to learn the concepts of a CNN and LSTM model and build a working model of Image caption generator by implementing CNN with LSTM.

In this Python project, we will be implementing the caption generator using ***CNN (Convolutional Neural Networks)*** and LSTM (Long short term memory). The image features will be extracted from Xception which is a CNN model trained on the Flikr8k dataset and then we feed the features into the LSTM model which will be responsible for generating the image captions.

Literature survey

Title 1: Camera2Caption: A real-time image caption generator (2017)

Author: Pranay Mathur, Aman Gill, Aayush Yadav, Anurag Mishra, Nand Kumar Bansode

Proposed methodology:

Proposed Deep Learning Based Advanced Technique Deep Reinforcement Learning that's led by Computer Vision and machines translation based on deep learning model. Dataset used in this model is MS-COCO.

Conclusions:

The proposed model based on deep learning, well optimize and perform in real time environment (mobile devices) and produce high quality captions by using help of tensorFlow by google.

Title 2: Image Captioning: Transforming Objects into Words. June 2019

Author: Simao Herdade, Armin Kappeler, Kofi Boakye, Joao Soares

Proposed Object Relation Transformer model, focuses on spatial relationship between objects of images through used of faster R-CNN with ResNet101. Mainly focuses on Improve the relationship between objects. Dataset used in this model is MS-COCO with Pycharm IDE

Conclusion: The proposed model encodes positions and size and relationship between detected objects in images and extracted features by building upon the bottom-up and topdown image captioning approach and CNN.

Title 3: Automatic Image Captioning Using Convolution Neural Networks and LSTM November 2019

Author: R subhash

Proposed Deep Learning based Convolution Neural Networks and Natural Language Processing (NLP) Techniques reasonable sentences are framed and inscriptions are produced. Dataset used in this model is MS-COCO.

Conclusion: Proposed model having convolution neural network whose output is paired to Long Short Term Memory network which helps us generate descriptive captions for the image. Also model don't require huge dataset to produce caption of images.

Title 4: Image Caption Generator Using Deep Learning

Author: B.Krishnakumar,K.Kous alya, S.Gokul,R.Karthikeyan, D.Kaviyarasu

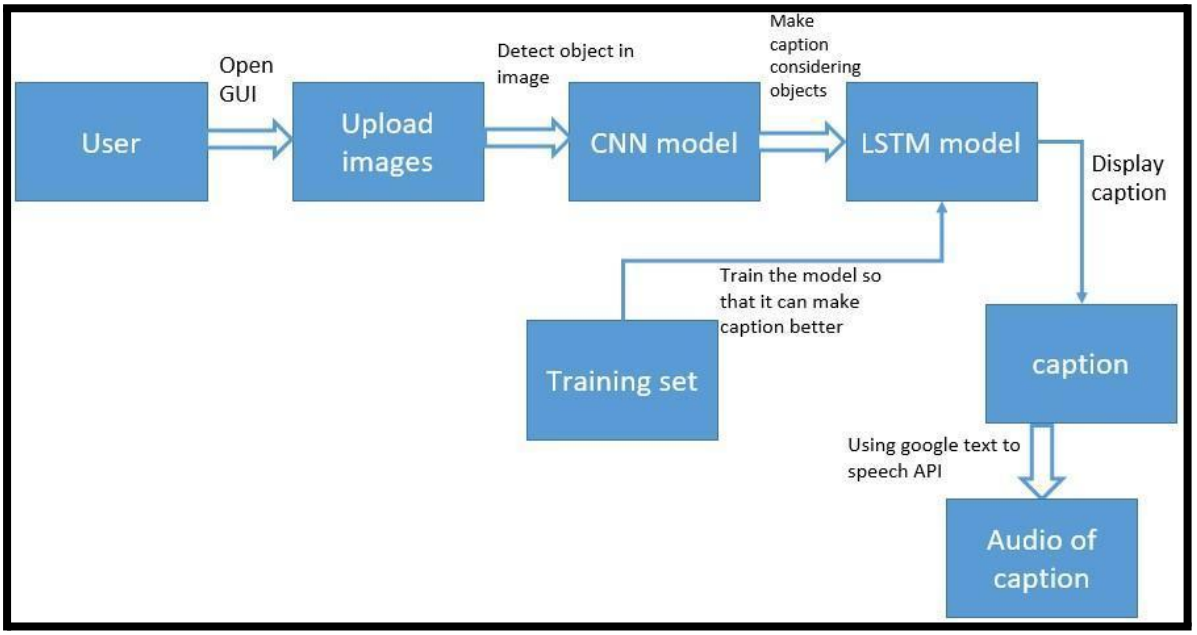
Proposed Deep Learning based Convolution Neural Networks to identify objects in the images using OpenCv. Detected Images converted into audio using GTTP and then converted to text using to Long Short Term Memory network.They used Pre-trained model VGG16 as a baseline model

Conclusion: Proposed Model successfully trained to generate captions of images using CNN technique, model is depends on data and used small data set. The model generate caption by using Keras Framework used in Jupyter notebook and also conclude keras has strong support for multiple GPU's.

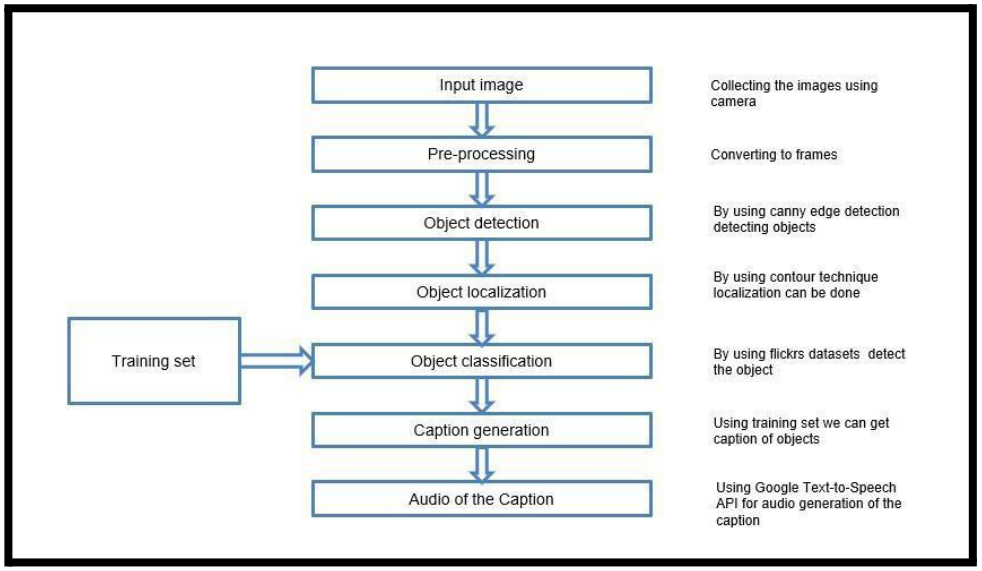
Proposed system

- This paper proposes a new method for initializing the width and height of predicted bounding boxes. Our proposed method has a larger Avg IOU and smaller running time on the MS COCO dataset.
- The Avg IOU is 60.44%, which is 0.56% higher than original YOLOv3 method, and the running time is 1/297 that of the original YOLOv3 method. For the PASCAL VOC dataset, the average IOU is 67.45%, which is 0.13% higher than original YOLOv3 method, and the running time is 1/81 that of the original YOLOv3 method.
- It exhibits better performance in terms of initializing the width and height of predicted bounding boxes, as well as in terms of choosing the representative initial width and height. Besides, we randomly selected some images from the test set of the MS COCO dataset for detection.
- The object detection results indicate that our proposed method detected more objects in some test images. It also has better performance in terms of detecting small objects. Our proposed method also outperforms the original YOLOv3 method in terms of recall, mean average precision, and F1-score.

Architecture diagram



Flow chart



Models

1. Collect the datasets
    - From the camera we can get video, and some data set collecting in the form images.
2. Feature extraction
    - Applying edge and localization on the images.
3. Applying algorithm
    - Yolov3 and MobileNet algorithms are applied.
4. Classification analysis
    - Object can be classified.

DIGITAL IMAGE PROCESSING

Digital Image Processing :-

- Digital Image Processing is a software which is used in image processing. For example: computer graphics, signals, photography, camera mechanism, pixels, etc.
- Digital Image Processing provides a platform to perform various operations like image enhancing, processing of analog and digital signals, image signals, voice signals etc.
- It provides images in different formats.

Digital Image Processing allows users the following tasks :-

- **Image sharpening and restoration:** The common applications of Image sharpening and restoration are zooming, blurring, sharpening, grayscale conversion, edges detecting, Image recognition, and Image retrieval, etc.
- **Medical field:** The common applications of medical field are Gamma-ray imaging, PET scan, X-Ray Imaging, Medical CT, UV imaging, etc.
- **Remote sensing:** It is the process of scanning the earth by the use of satellite and acknowledges all activities of space.
- **Machine/Robot vision:** It works on the vision of robots so that they can see things, identify them, etc.

Characteristics of Digital Image Processing :-

- It uses software, and some are free of cost.
- It provides clear images.
- Digital Image Processing do image enhancement to recollect the data through images.
- It is used widely everywhere in many fields.
- It reduces the complexity of digital image processing.
- It is used to support a better experience of life.

Advantages of Digital Image Processing :-

- Image reconstruction (CT, MRI, SPECT, PET)
- Image reformatting (Multi-plane, multi-view reconstructions)
- Fast image storage and retrieval
- Fast and high-quality image distribution.
- Controlled viewing (windowing, zooming)

Disadvantages of Digital Image Processing :-

- It is very much time-consuming.
- It is very much costly depending on the particular system.
- Qualified persons can be used.

History of Photography :-

The first process of photography was called **heliography** which was invented by **Nicephore Niepce** in the year **1824**. In this process, images are obtained after several days by bitumen of Judea on a silver plate.

In 1829, Niépce relates to Louis Jacques Mandé Daguerre in his research.

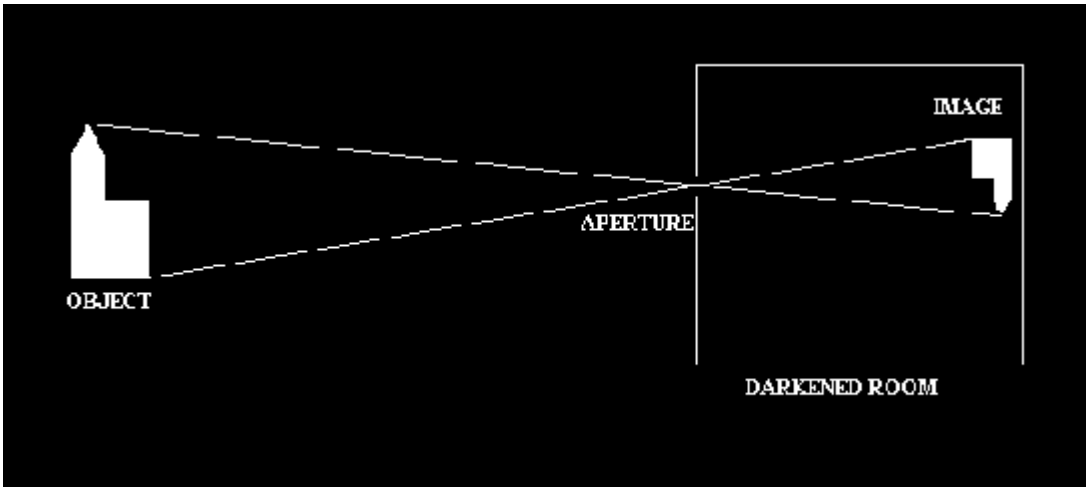
Origin of the camera :-

There were several scientific inventions that took place for the betterment of people. One of the best achievement was the invention of the camera which can catch human activity. Photographic technology is changed over many generations.

It starts with camera obscura and then extends with the introduction of new technology including **calotypes, dry plates, daguerreotypes, and today digital cameras**. Digital cameras have completely changed the history of photography because it produces the best quality of images.

Camera Obscura :-

In Latin word, the Obscura means a dark room, which was used to take the picture from the other side of the camera. This concept was taken from a Chinese concept to where images are projected on a surrounding wall.



an example of camera obscura