International Journal of Future Generation Communication and Networking
Vol. 13, No. 3, (2020), pp. 3247–3251

Object Detection and Automatic Image Captioning Using Tensorflow

Raj kadam[1] , Uday Kumbhar[2] , Onkar Gulik [3], Dr Makrand Shahade[4]

[1] *Scholar, Department Of Computer Engineering, JSPM's RSCOE Pune*

[2] *Scholar, Department Of Computer Engineering, JSPM's RSCOE Pune*

[3] *Scholar, Department Of Computer Engineering, JSPM's RSCOE Pune*

[4] *Assoc Prof, Department Of Computer Engineering, JSPM's RSCOE Pune*

[1]rajkadam91098@gmail.com, [2]udaykumbhar619@gmail.com, [3]onkargulik999@gmail.com, [4]manu1509.shahade@gmail.com

Abstract

The recent advancement in the field of object detection and image captioning has influenced us to develop this venture. The process of determining objects present in the image is called as object detection. Multiple objects can be detected in a given image. We as humans can easily spot the objects present in the image. The machine needs to understand the specifications of image. Generating the descriptions of images has become popular. It is a difficult task because the image needs to be understood properly and the visual knowledge needs to be converted into descriptive sentence. It is important to detect the objects in the image before captioning it. Object detection falls under one of the international popular research fields. Large scale image classification is possible by building very deep convolutional neural network (CNNs). Pre-trained model is used to predict category of image. The main motive implies training computer to classify an image.

Keywords: *Deep learning, Feature extraction, Object, Object detection, Image.*

1. Introduction

Automatically generating the captions for images uploaded by social media users is gaining popularity with the advancement in object classification methodologies. Identification of objects in images and videos relates to object detection. There is slight difference in working of object detection and recognition. Enhancing object and object parts is a vital task. Humans have the natural capability to spot details present in image or video through naked eyes. We concentrate on parallel discovering and localization of common objects present in real world images. The main goal is to improve the accuracy of system. It becomes more challenging when it comes to computer to understand the complex scenes. We have used the deep learning approach to train the computer.

2. Related Work

- Object detection from images using convolutional neural networks, Olavi Stenroos.

This paper helped us in answering questions like, why we got into deep learning rather than machine learning ,As Object detection is a subfield of machine learning, but for past decade this field is being dominated by so called deep neural networks. We were able to take advantage of improvement in computing power and data availability. How different features of Fast R-CNN or CNN like identifying edges, corners and colour differences across the image and to combine them to form different shapes.

ISSN: 2233-7857 IJFGCN

- A region-based image caption generator with refined descriptions, Philip Kinghorna, Li Zhanga, Ling Shao

The method of automatic image captioning has gained popularity, Karpathy and Li proposed system for natural language descriptions for image regions. This paper tells how to avoid holistic techniques at the start process of project.

3. Motivation

We got motivated from various aspects of existing system. Most important task is to compare the user object with database image using scale invariant feature transform. In this system first preprocessing on the image is done, after that feature extraction is performed by comparing object with database. The motive of object detection is to locate (localize) all known objects in image. In computer machine object recognition is vital part. It implies determining different aspects present in the images and videos. Extraction features and learning algorithms are frequently used by object recognition methods to recognize instances of object belonging to an object category. The relationship between different objects present in the image is established. Each object in the image or video has distinct characteristics. Image retrieval, security, surveillance are some are the fields in which concept of object recognition is used. Process of automatic generation of captions has influenced us to develop a product that will save users time to a great extent.

4. Existing System

The most popularly studied computer vision problem include localization and object detection in images. Due to presence of intra-class variation, inter-class diversity, and noisy annotations in wild images, it becomes challenging task. To achieve better performance, data is vital to train detectors. Earlier the process of image captioning was divided into two parts. Template matching is first part. Actions, scenes, objects, attributes insert them into hand designed template. Output generated from template matching technique are not always fluent and expressive. The later part depends on retrieval based approach, it first incurs the like images from website to match the given statement. It lacks in updatability, performance, flexibility and scalability.

5. Proposed System

Proposed architecture easily deals with image retrieval problem without any complications . Deep neural network can solve the problems occurring in both the issues, by generating suitable, expressive and fluent captions. Efficient computation of automatic metrics is possible using deep learning. It speeds up the development process of image captioning. However, these automatic metrics roughly correlate with human judgment. Task is to built a model to recognize category of object by first annotating and then comparing it with images present in pre labelled training set. CNN and RNN is used to fully classify the image. Existing system allows the social media user to upload the image of their choice of any dimensions and having complexity. In the system proposed by us, social media user does not have to waste time on searching captions suitable for image on google. Our system provides a user friendly platform for the social media user to upload the image of their choice. User doesn't have to type the caption manually for the uploaded image. The System extracts the captions from website that we have linked to the software. After uploading the image, as soon as generate caption button is pressed a suitable caption is generated. The neural network that we have used is in the system is convolutional neural network.

6. Disadvantages of Existing System

1. Excellent image quality is required as input

ISSN: 2233-7857 IJFGCN

2. Hard to detect features from low quality image . Difficult to analyze complex scene

3. Usage of proxy is done for speeding up the image retrieval process

4. Time consuming when the input image is complex

7. Advantages of Proposed System

1. Proposed framework can solve the image retrieval issues.

2. Neural networks can handle all the issues by generating suitable, expressive and highly fluent caption using tensorflow and algorithm.

3. Efficient computation of automatic metrics is possible.

4. As the captions are generated automatically there is no need to waste time on searching

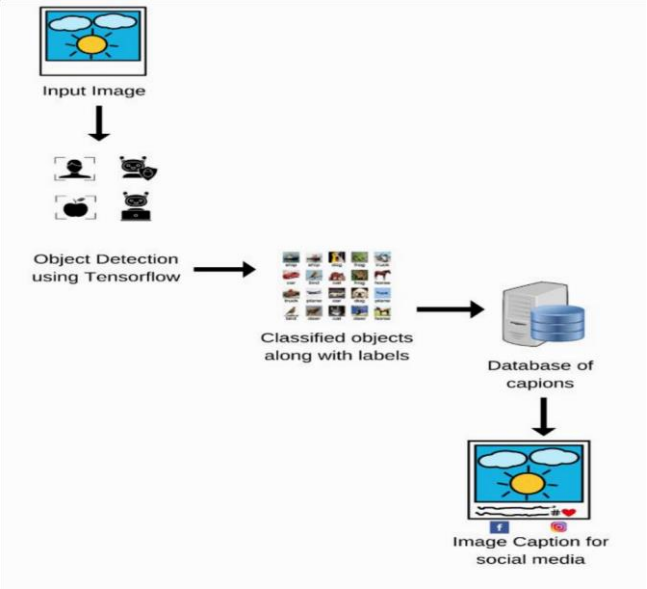

Figure 1. Architecture of System

ISSN: 2233-7857 IJFGCN

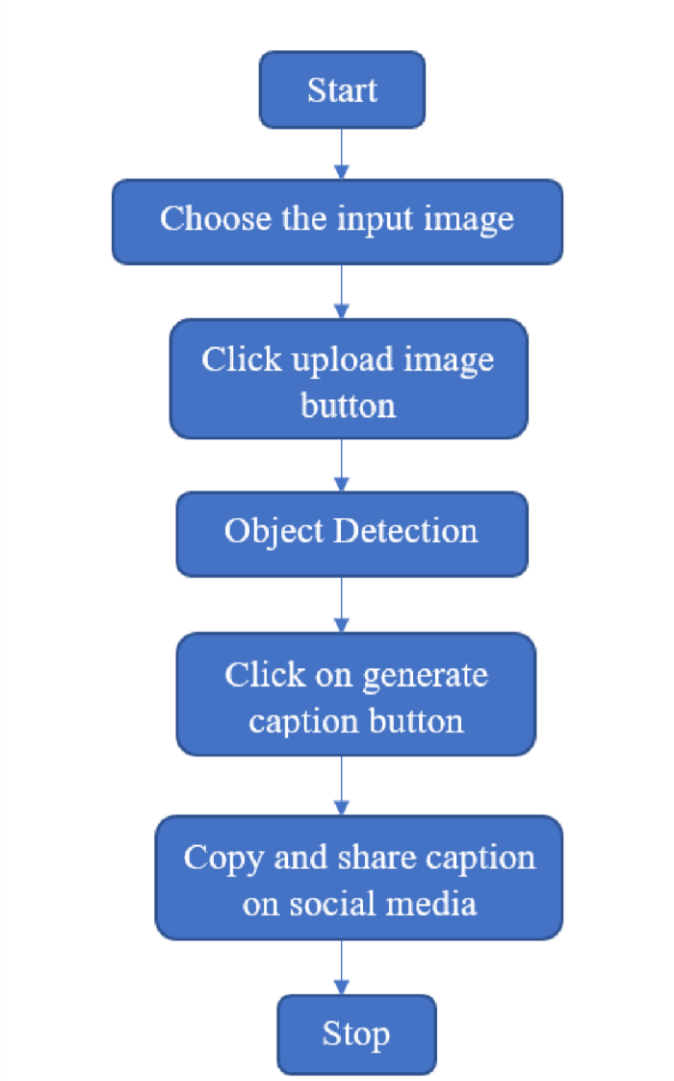

Figure 2. Flowchart of Proposed System

8. Conclusion

Taking advantages of existing systems and algorithm we can propose a system with much good efficiency and resulting in better image recognition. Using the proposed system, it becomes easier for social media user to generate caption automatically suitable for the uploaded Image without wasting much time. The produced architecture can efficiently deal with image/instance retrieval, these automatic metrics can be computed efficiently. Efficient algorithm of image recognition requires better and faster image captioning algorithm, this condition is satisfied by the proposed architecture.

9. Acknowledgements

We are deeply grateful to our project guide Assoc Prof Dr M. R. Shahade for his contribution in the field of deep learning.

ISSN: 2233-7857 IJFGCN

References

1. Philip Kinghorn, Li Zang, “a region based image caption generator with refined descriptions” , Elsiver B V, 6 july 2017, Ling Shao University Northumbria New castle NE 1, United Kingdom.[1]

2. Olavi Stenroos, “Object detection from images using convolutional neural networks”, master’s thesis, july 28 2017, Aalto University, School Of Science.[2]

3. Luis Fernando, Rafel E Banchs, “Automatic labelling of touristic pictures”,IEEE publication, 2017 Institute For Infocom Research,Connex South Tower,Singapore.[3]

4. Xudie Ren, Huanon guo, Shengong li, Shilin Jianhua li, Shangai tong, “a novel image classification method using CNN”, 15 October 2019, university of China.[4]

5. Tianmei Guo, Jiwen Dong, Henjian Yunxin Gao, “simple convolutional neural network on image classification” IEEE publication, 26 decdember 2018, Department of computer and technology, Computing University of China.[5]

6. Raffela Bernardi, Ruket Cacki, Demond Elliot, Frank Keller, Adrian Muscat, “automatic description generation from images”, arxiv publication, 19 march 2017, University of Copenhagen.[6]

ISSN: 2233-7857 IJFGCN