Captionizer: Automated Caption Suggestions

Priyanka Raghuvanshi[1], Priyanshi Karamchandani[2], Sakshi Verma[3], Ruchi Deshmukh[4]

[1,2,3,4]Acropolis Institute of Technology and Research, Indore, (M.P), India

ABSTRACT

More than 3.78 billion people use Social Media and every day millions of images are posted on these platforms. Many users need quotes and phrases for their respective images for which they surf on the internet even for hours which consumes time and energy. So our project will provide automatic quotes and phrases based on your image provided using machine learning and database searching.

Keywords: caption generator, machine learning, object detection, captionizer

INTRODUCTION

The basic ability of human beings is the tendency to describe an image with an ample amount of information about it by just a quick glance [1]. Creating a computer system to simulate the abilities of human beings is a long time researcher goal in the fields of machine learning and artificial intelligence. There are several research progress made in the past such as the detection of objects from a given image, attribute classification, image classification, and classification of actions by human beings. Making a computer system to detect the image and produce a description using natural language processing is an exigent task, which is called an image caption generator system. Generating a caption for an image involves various tasks such as understanding the higher levels of semantics and describing the semantics in a sentence by which human can understand. In order to understand the higher levels of semantics, the computer system must learn the relationships between the objects in a given image. Usually, communication in human beings occurs with the help of natural language, so developing a system that produces descriptions that can be understandable by human beings is a challenging goal. There are several steps to generate captions, such as understanding visual representation of objects, establishing relationships among the objects and generating captions both linguistically and semantically correct. This paper aims at detection, recognition and generating captions using deep learning.

Whenever an image appears in front of us, our brain is capable of annotating or labeling it. But, what if a website can do this for us? With the enhancement of Computer Vision and Deep learning algorithms, availability of relevant datasets, and AI models, it becomes easier to build a relevant caption generator for an image. Captionizer is a web-based application which provides quotes to the user according to their inputs (this input can be image or keyword).It recognizes the context of an image and annotates it with relevant captions using deep learning, and computer vision .It includes the labeling of an image with English keywords with the help of datasets provided during model training. It also provides quotes to users on the basis of keywords and if anybody wants to share their quotes they can.

REVIEW OF LITERATURE

There have been several attempts at providing a solution to this problem including template based solutions which used image classification i.e. assigning labels to objects from a fixed set of classes and inserting them into a sample template sentence. But more recent work have focused on Recurrent Neural Networks [2,3]. RNNs are already quite popular with several Natural Language Processing tasks such as machine translation where a sequence of words is generated. Image caption generator extends the same application by generating a description for an image word by word. The computer vision reads an image considering it as a two dimensional array. Therefore, Venugopalan (et al)[9] has described image captioning as a language translation problem. Previously language translation was complicated and included several different tasks but the recent work[10] has shown that the task can be achieved in a much efficient way using Recurrent Neural Networks. But, regular RNNs suffer from the vanishing gradient problem which was vital in case of our application. The solution for the problem is to use LSTMs and GRUs which contain internal mechanisms and logic gates that retain information for a longer time and pass only useful information.

One of the major challenges we faced was choosing the right model for the caption generation network. In their research paper, Tanti (et al)[4] has classified the generative models into two kinds – inject and merge architectures. In the former, we input both, the tokenized captions and image vectors to an RNN block whereas in the latter, we input only the captions to the RNN block and merge the output with the image. Although the experiments show that there is not much difference in the accuracy of the two models, we decided to go with the merge architecture for the simplicity of its design, leading to reduction in the hidden states and faster training. Also, since the images are not passed iteratively through the RNN network, it makes better use of RNN memory.

Implementation Phase

This project is proposed in the form of website which has the following two parts:-

Front-End**:** The front end of the project is built on React Js and the styling and animations will be done by Sass. project will consists of four sections:

- Section where users can get quotes based on image.
- Section where users can get quotes based on keywords.
- Section where users can get random quotes. Add your quotes

The whole purpose of the front-end system is to provide the resources such as image or keywords to the server and display the resulting quotes.

Back-End: The Backend of the project will provide the quotes based on the resources provided by the front end using computer vision with machine learning algorithms and database searching. The server will be built using Django in python and the database will be built using MySQL. The algorithms of computer vision and database will be written in python using pre-built libraries such as OpenCV, Tensorflow, etc.

The implementation phase of a Captionizer project would involve several key steps:

- Taking input from user: this step involves input of user. User can upload images or enter keyword and get captions.
- Processing user input : as soon as backend receive image it will process the image and detect object from the images and give keywords .
- Fetching quotes on the basis of keyword: this involves fetching of quotes from keywords that we got from image detection process.
- Displaying quotes : this involves displaying of quotes

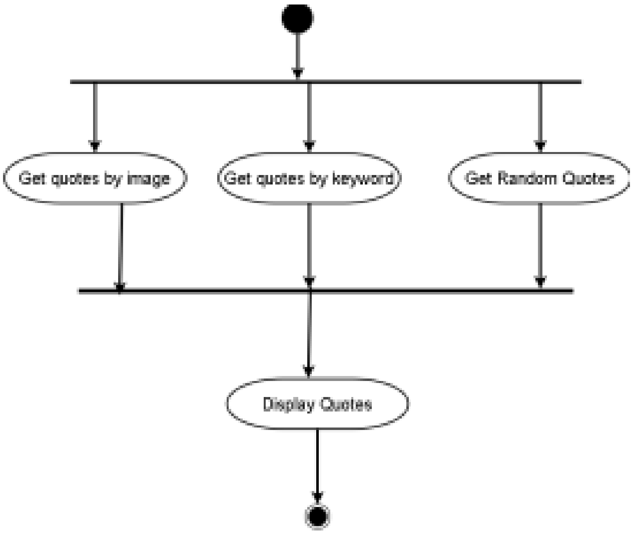

Fig 1. Activity diagram of proposed software

❖ *System Architecture*

The system architecture of a captionizer project can vary depending on the specific implementation and requirements of the project. However, typical system architecture for such a project would involve the following components:

Data collection: This component would be responsible for collecting data.

Data processing and AI algorithms: This component would be the core of the system, and would include the various AI algorithms responsible for object detection . The AI algorithms could include machine learning models, deep learning models, and natural language processing (NLP) algorithms.

Quotes generator: This component would use the output of the AI algorithms to provide quotes .

User interface: This component would provide an interface to display quotes.

❖ *User Interaction*

The user interaction of captionizer project would be designed to be intuitive, user-friendly, and provide captions to user. The user interaction could involve the following components:

- Quotes from image: It provides quotes from image. User can upload image and captionizer will detect object in that image and display all quotes related to that object.
- Quotes from keywords: It provides quotes from keyword.

 Get random quotes : it generate random quotes Add quotes : It allow user to add there quotes.

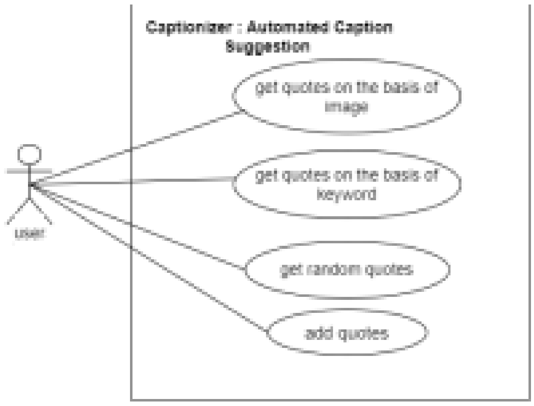
Captionizer : Automated Caption Suggestion
get quotes on the basis of image
get quotes on the basis of keyword
get random quotes
add quotes
user

Fig 2. Use case of proposed software

V. Result and Discussion

The main objective of this project is to provide the captions to the user from image, keyword.There are following three ways in which you can generate a suitable quote for your use: Quote Suggestions based

on your Image Upload your image from your computer or web link that will get you suitable quotes for your given image. Quote Suggestions based on your keyword -In any case you want your quote suggestions based on your given keywords, you can input all the related keywords and in return you will get suitable quotes for your keywords Random Quote Suggestion You will get a random quote which does not depend on your input.

❖ *Possible Limitations*

The proposed system has some challenging problems due to the following limitations:

- The performance of the system depends on the discrete number of quotes that can limit it.
- The system depends on our machine learning algorithms and database and it sometimes may produce unpleasant results.
- The project is in the initial state, so accuracy of the project is less.

VI. Conclusion

In conclusion, many users need quotes and phrases for their respective images for which they surf on the internet even for hours which consumes time and energy. So our project will provide automatic quotes and phrases based on your image provided using machine learning and database searching .object detection are growing fields that have the potential to significantly improve the efficiency and accuracy . The technology involves using artificial intelligence algorithms to detect object and generating quotes .

VII. Acknowledgement

We are sincerely grateful to our college "ACROPOLIS INSTITUTE OF TECHNOLOGY AND RESEARCH" for giving us an opportunity to work on the project which enable us to learn new Technologies and enhance our skills which would be helpful to us in our professional career. We owe a debt of sincere gratitude, the deep sense of reverence and respect to our guide and mentor Prof Anurag Punde AITR, Indore for his motivation, sagacious guidance, constant encouragement, vigilant supervision and valuable critical appreciation throughout this project work, which helped us to successfully complete the project on time.

We express profound gratitude and heartfelt thanks to Dr. Kamal Kumar Sethi, HOD CSE, AITR Indore for his support, suggestion and inspiration for carrying out this project.We are also very much thankful to other faculty and staff members of CSE Dept, AITR Indore for providing us all support, help and advice during the project.

References

1. L. Fei-Fei , A. Iyer , C. Koch , P. Perona. ,What do we perceive in a glance of a real-world scene? J. Vis. 7 (1) (2007) 1–29 service delivery," *Sci. Rep.*, vol. 12, no. 1, p. 3601, 2022.
2. "Every Picture Tells a Story: Generating Sentences from Images." Computer Vision ECCV (2016) by Farhadi, Ali, Mohsen Hejrati, Mohammad Amin Sadeghi, Peter Young, Cyrus Rashtchian, Julia Hocken maier, and David Forsyth
3. Show and Tell: A Neural Image Caption Generator by Oriol Vinyal, Alexander Toshev, Samy Bengio, Dumitru Erhan, IEEE (2015)
4. Where to put the Image in an Image Caption Generator by Marc Tanti, Albert Gatt, Kenneth P. Camilleri.