IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

GENERATING CAPTIONS FOR IMAGES GIVEN BY USERS

[1]Soniya Jadhav, [2]Kavach Shah, [3]Swarnim Mudaliar, [4]Prof. Torana Kamble

[1]Student, [2]Student, [3]Student, [4]Professor.
[1]Dept. of Computer Engineering,
[1]Bharati Vidyapeeth College of Engineering, Belapur,
Navi Mumbai, Maharashtra, 400614, India

Abstract: Recent developments in the field of computer vision and object detection have allowed for an increased interest in the concept of image caption generation. Various researchers have worked on this topic and have produced impressive results. In this paper, we discuss the latest techniques used for the generation of captions from any given input image. The paper elaborates on the latest approach towards solving the problem of caption generation, namely by focusing on the use of a Convolutional Neural Networks (CNN) based model for feature extraction and a Long-Short Term Memory (LSTM) based Recurrent Neural Network model for generating meaningful captions. Furthermore, the existence of huge datasets like Flickr30k , Kaggle and MS COCO, along with greater computational power, allow for increased accuracy in the built models.

Index Terms **- Computer Vision ,Object Detection Convolutional Neural Network (CNN) , Recurrent Neural Network (RNN) , Long Short Term Memory (LSTM) .**

I. INTRODUCTION

Inspired by recent work in machine translation and object detection , this Image Caption Generator is an attention based model which automatically learns to describe the content of images . As the name suggests , this web application is used to generate caption or a description of the image . One of the assumptions generally made is that image caption generation is an easy concept and one can implement it without any difficulties . But everyone forgets that the scene understanding capacity of human and that of a computer system cannot be compared . Scene understanding capability comes naturally to a human being but it must be taught to a computer system . Automatically generating captions for an image is a similar to the scene understanding which is one of the primary goals of computer vision . The caption generation models must be able to solve the computer vision challenges of determining what objects are in an image and they must also be powerful enough to capture and express their relationships in human language . The computer system must have the remarkable human ability to compress huge amount of relevant visual information into descriptive language and is thus an important challenge for machine learning and AI research but this Image Caption Generator overcomes exactly that obstacle flawlessly to give the user appropriate description of the image .

At present , the mainstream image caption generation algorithm is based on the combination of CNN image recognition model and RNN structured model .

In Image Caption Generator , the user will give an image as an input and the description of image will be the output . The user will upload the image on the user interface i.e. website and the output will also be displayed on the website . After getting the input , the system will identify the objects in the image . For this , the system makes use of CNN i.e. Convolutional Neural Networks . Then , it describes the objects and establishes relationship between objects . The system uses RNN i.e. Recurrent Neural Networks to convert it into human language or text . The final output i.e. the description of the image will be displayed on the same website . Processing of data is done on cloud storage . The dataset used for this project comprises of more than 30 ,000 images having 5-6 descriptions each .

The contributions of the paper are as follows :

- One of the objectives of Image Caption Generator is describing the image which the user has given as an input .
- Another objective of this project is making use of Convolutional Neural Network for object detection .
- Neural language detection model using Recurrent Neural Networks .
- Special focus is given on recurrence of a specific domain using Long Short-Term Memory (LSTM) .
- To provide appealing description or caption generator which provides apt description of the images .

II. LITERATURE REVIEW

[1] Songtao Ding , Shiru Qu , Yuling Xi , Arun Kumar Sangaiah and Shaohua Wan , in their reseach paper , Image Caption Generation with High-Level Image Features , uses an approach which involves Convolutional Neural Networks and Recurrent Neural Networks . The drawbacks of this paper was since it was not on cloud storage , it consumed a large amount of space .

[2] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhutdinov, Richard Zemel and Yoshua Bengio have published a paper , Show, Attend and Tell: Neural Image Caption Generation with Visual Attention in Cornell University Journal . The techniques involved in this paper are Convolutional Neural Networks , Recurrent Neural Networks and Long Short Term Memory Models . This system is similar to the proposed system but the only drawback of this paper is that it is related to a specific domain .

[3] Akash Verma , Harshit Saxena , Mugdha Jaiswal , Dr. Poonam Tanwar in their paper , Intelligence Embedded Image Caption Generator using LSTM based RNN Model have worked on image caption generation using CNN , LSTM models and RNN . This system has some similarity to the proposed system . The drawbacks of this paper were not only was this paper related to a specific domain but the dataset used was also very small .

Table 1. Literature review comparison

Sr No	Name	Publisher	Author	Techniques Involved	Drawbacks
1.	Image caption generation with high-level image features	ELSEVIER	Songtao Ding Shiru Qu Yuling Xi Arun Kumar Sangaiah Shaohua Wan	CNN RNN	No scope for memory retention in domain specific analysis
2.	Show, Attend and Tell: Neural Image Caption Generation with Visual Attention	Cornell University Journal	Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhutdinov, Richard Zemel, Yoshua Bengio	CNN RNN LSTM models	Owing to the added focus on specific domains using the LSTM model, the model fails to perform for more general use cases if it is unrelated to the specific domain.
3.	Intelligence Embedded Image Caption Generator using LSTM based RNN Model	IEEE	Akash Verma Harshit Saxena Mugdha Jaiswal Dr. Poonam Tanwar	CNN RNN LSTM models	Because the model uses both a general approach and a domain specific approach, the data used is very small as compared to other models

IV. CONCLUSION

The aim of computer vision is scene understanding i.e. the system has to perceive each and every object as well as its surroundings as easily as human brain . Not only must caption generation models be able to solve the computer vision trials of determining what objects are in an image, but they must also be powerful enough to capture and express their relationships in human language . In this paper , we have reviewed deep learning based image captioning methods . We have done research on 3 technical papers wherein 1st one involved CNN and RNN . Its drawback was since it was not on cloud storage , it consumed a large amount of space . The next paper involved CNN , RNN and LSTM model but this was related to a specific domain . Hence its application was limited . The last paper also involved CNN , RNN and LSTM model . It was similar to the 2nd paper but in this , the dataset was also small .

The system in this paper which we have developed overcomes all these drawbacks and provides a reliable platform for image caption generation . Then the paper focuses on how to develop it and the generic block diagram for the system .

V. ACKNOWLEDGEMENT

We take this opportunity to express our deepest gratitude and appreciation to all those who have helped us directly or indirectly towards the successful completion of this project . It is a great pleasure and moment of immense satisfaction for us to express our profound gratitude to out dissertation Project Guide, Prof. Torana Kamble whose constant encouragement enabled us to work enthusiastically. We are highly indebted to her for her invaluable guidance and ever-ready support in the successful completion of this dissertation in time. Working under her supervision has been a remarkable and unforgettable experience. Despite of her busy schedule, she was always available to give us advice, support and guidance during the entire period of our project. The finalization of this project would not have been achievable without her morale-boosting stimulation, patient guidance and consistent support. We would extend our gratitude towards our parents without whom we would not get the confidence to be here.

VI. REFERENCES

[1] Image caption generation with high-level image features . Publisher–ELSEVIER (https://www.researchgate.net/publication/331999950_Image_Caption_Generation_with_High-Level_Image_Features)

[2] Show, Attend and Tell: Neural Image Caption Generation with Visual Attention . Publisher –Cornell University Journal (https://arxiv.org/abs/1502.03044)

[3] Intelligence Embedded Image Caption Generator using LSTM based RNN Model . Publisher– IEEE (https://ieeexplore.ieee.org/document/9489253)