IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 6, May 2023

Image Caption Generator Using Deep Learning Approach

Prof. A. S. Narote[1], Kunal Vispute[2], Harshit Himanshu[3], Rajas Bhagatkar[4], Sneha Jadhav[5]

[1]Professor, Smt. Kashibai Navale College of Engineering Vadgaon Bk, Pune, India

[2,3,4,5]Student, Smt. Kashibai Navale College of Engineering Vadgaon Bk, Pune, India

Abstract: *The creation of captions for a picture is the focus of the image caption generator. The image's semantic meaning is extracted and translated into plain language. Also, built-in programs create and offer a caption for a certain image. Image captioning is the process of creating a description of a picture in the form of a caption. The system must recognize and develop connections between things, people, and animals. This study uses deep learning to find, identify, and produce interesting captions for a given image. The practice of creating textual descriptions of a given image using computer vision and natural language processing methods is known as image captioning. This suggests a strategy for deriving a caption for a picture that highlights particular objects in the image.*

Keywords: Deep Learning, Feature Extraction, Thresholding, Image Segment, CNN Model, NLP.

I. INTRODUCTION

Image caption generation is a task that involves image processing and natural language processing concept to detect the context of an image and describe them in natural language. While human beings can do it easily, it takes a lot of computational power for a computer system. The mechanism must detect and establish relationships between objects, people, and animals. There are several steps to generating captions, such as understanding the visual representation of objects, establishing relationships among the objects, and generating captions both linguistically and semantically correlated.

Image caption generation is a challenging task that involves the use of advanced techniques in both computer vision and natural language processing. The goal is to create a system that can automatically generate a natural language description of an image, accurately capturing the content and context of the scene. This requires the system to be able to identify and recognize various elements within the image, such as objects, people, and animals, and to understand the relationships between these elements and their context within the scene. To achieve this, the system must be able to extract meaningful features from the image data and use these features to generate a coherent and semantically meaningful description.

Advances in deep learning have enabled significant progress in this field, allowing for the development of more sophisticated and accurate image captioning systems. These systems use neural networks to learn complex representations of image data and natural language, enabling them to generate accurate and fluent captions. Despite these advances, there are still many challenges to be overcome in this field, including improving the accuracy and fluency of generated captions and developing systems that can handle a wider range of images and scenes.

II.LITERATURE SURVEY

1] “Image Caption Generator” by Megha J Panicker and Vrinda Mathur explores the use of a combination of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to generate captions for images. The authors found that by using these two types of neural networks in conjunction, they were able to effectively identify relationships between objects within images and generate coherent and semantically meaningful captions. In addition to the approach proposed by Panicker and Mathur, several other hybrid image captioning systems have been developed that combine different neural network architectures to improve the accuracy and fluency of generated captions.

DOI: 10.48175/IJARSCT-10150

[2] “Comparative Evaluation of CNN Architectures for Image Caption Generation” Author: SulabhKatiyar and Samir Kumar Borgohain., Aided by recent advances in Deep Learning, Image Caption Generation has seen tremendous progress over the last few years. Most methods use transfer learning to extract visual information, in the form of image features, with the help of pretrained Convolutional Neural Network models followed by transformation of the visual information using a Caption Generator module to generate the output sentences.

[3]Image Caption Generator, Author: Liya Sunny and Sara Susan Joseph, Image Caption Generation involves training a Machine Learning model to learn to automatically produce a single sentence description for an image. For human beings, it is a trivial task. However, for a Machine Learning method to be able to perform this task, it has to learn to extract all the relevant information contained in the image and then convert this visual information into a suitable representation of the image which can be used to generate a natural language sentence description of the image.

[4]Image Caption Generating Deep Learning Model Author: AishwaryaMaroju, Sri Doma, LahariChandarlapati, Image captioning is the process of generating descriptions of what is going on in the image. By the help of Image, Captioning descriptions are built which explain about the images. Image Captioning is basically very much useful in many applications like analyzing large amounts of unlabeled images and finding hidden patterns for Machine Learning Applications for guiding Self-driving cars and for building software that guides blind people. This Image Captioning can be done by using Deep Learning Models.

[5]Image Captioning using Deep Learning, Author: Murk Chohan, Adil Khan, Muhammad Saleem Mahar, Saif Hassan, Abdul Ghafoor, Mehmood Khan, Auto Image captioning is defined as the process of generating captions or textual descriptions for images based on the contents of the image. It is a machine-learning task that involves both natural language processing (for text generation) and computer vision (for understanding image contents). Auto image captioning is a very recent and growing research problem nowadays. Various new methods are being introduced daily to achieve satisfactory results in this field.

[6]Bottom-Up and Top-Down Attention for Image Captioning and Visual Question Answering, Author: Peter Anderson, Xiaodong He, Chris Buehler, Damien Teney, Mark Johnson, Stephen Gould1 Lei Zhang, Top-down visual attention mechanisms have been used extensively in image captioning and visual question answering (VQA) to enable deeper image understanding through fine-grained analysis and even multiple steps of reasoning. In this work, we propose a combined bottom-up and top-down attention mechanism that calculates attention at the level of objects and other salient image regions.

[7]Image Caption Generator Author:Parth Kotak , Prem Kotak, Automatically creating the description or caption of an image using any natural language sentences is a very challenging task. It requires both methods from computer vision to understand the content of the image and a language model from the field of natural language processing to turn the understanding of the image into words in the right order.

[8]Visual Image Caption Generator Using Deep Learning, Author: Priyanka Kalena, NishiMalde, Aromal Nair, Saurabh Parkar, Image Caption Generation has always been a study of great interest to researchers in the Artificial Intelligence department. Being able to program a machine to accurately describe an image or an environment like an average human has major applications in the field of robotic vision, business, and many more. This has been a challenging task in the field of artificial intelligence throughout the years.

[9]Image Caption Generator Using Convolutional Neural Network Algorithm, Author: Shaik Parvez, It is a very difficult challenge to automatically describe an image using a sentence from any natural language, such as English. It necessitates knowledge of both natural language processing and picture processing. The fusion of computer vision and natural language processing has received a lot of interest recently thanks to the advent of deep learning. This field is exemplified by image captioning, which teaches a computer to understand an image's visual information using one or

DOI: 10.48175/IJARSCT-10150

more phrases. In addition to the ability to recognize the item and the scene, high-level image semantics also needs the ability to analyze the state, the properties, and the relationship between these things.

III. OBJECTIVES

1. Our project aims to learn the concepts of CNN and LSTM models and build a working model of an Image caption generator by implementing CNN with LSTM.
2. This System is based on pre-processing of an image caption generator which will generally result in less effort and time.

IV. IMPLEMENTATIONS

Details: The user can use the prediction system through browser on Desktop, Laptop, Tab or Smart Phone. For the execution we need some packages to reduce the extra coding work. Packages will help to minimize the extra coding work to keep the program neat and clean.

For the hosting we are using FLASK which is a micro web server. With the help of flask api, we have created connections between different web pages.

For the SQL data connection, to store the data and to fetch the data, we are using the SQL Alchemy.

The web pages are dynamic pages in which we don’t need to change header and footer for every single page. The pages will work in container manner, for this we used JINJA TEMPLATE from Flask

V.CONCLUSION

This study offers a deep learning approach for creating image captions with neural networks; the suggested technique now incorporates a Flickr 8k dataset. Compared to existing image caption generators, the suggested deep learning technology produced captions with more descriptive meaning. Further research could lead to the creation of a hybrid photo caption generator model for more accurate captions.

REFERENCES

[1] Megha J Panicker, Vikas Upadhayay, Gunjan Sethi, Vrinda Mathur: “Image Caption Generator” International Journal of Innovative and Exploring Engineering (IJITEE) JAN-2021.

[2] Murk Chohan, Adil Khan, Muhammad Saleem Mahar, Saif Hassan, Abdul Ghafoor, Mehmood Khan: “Image Captioning using Deep Learning: A Systematic Literature Review” IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 5, 2020.

[3] Peter Anderson, Xiaodong He, Chris Buehler, Damien Teney, Mark Johnson, Stephen Gould, Lei Zhang: “Bottom-Up and Top-Down Attention for Image Captioning and Visual Question Answering” 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.

[4] Liya Ann Sunny, Sara Susan Joesph, Sonu Sarah Geogy, K.s>Ssreelakshmi, Abin T. Abraham: “Image Caption Generator” International Journal of Recent Advances in Multidisciplinary Topics (IJRSMT) APRIL2021.

[5] Aishwarya Maroju, Sneha Sri Doma, LahariChandarlapati: “Image Caption Generating Deep Learning Model” International Journal of Engineering Research & Technology (IJERT), September2021.

[6] Sulabhkatiyar,Samirkumarborgohain:” Comparative Evaluation of CNN Architectures for Image Caption Generation”(IJACSA) International Journal of Advanced Computer Science and Applications,2020.

[7] Parth Kotak and Prem Kotak:”Image Caption Generator”(IJERT) International Journal of Engineering Research & Technology,November-2021.

[8] Priyanka Kalena ,NishiMalde ,Aromal Nair ,Saurabh Parkar:” Visual Image Caption Generator Using Deep Learning” International Conference on Advances in Science & Technology (ICAST-2019).

DOI: 10.48175/IJARSCT-10150