International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 01 | Jan 2022 www.irjet.net p-ISSN: 2395-0072

IMAGE CAPTION GENERATOR USING DEEP LEARNING

Chaithra V[1], Charitra Rao[2], Deeksha K[3] , Shreya[4], Dr.Jagadisha N[5]

[1-4]Student, Dept. of Information Science and Engineering, CEC, Karnataka, India
[5]Associate Professor & Head of Dept. of Information Science and Engineering, CEC, Karnataka, India

---

Abstract - *In the last few years, the problem of generating descriptive sentences automatically for images have gained a rising interest in Natural language processing (NLP). Image captioning is a task where each image must be understood properly and are able generate suitable caption with proper grammatical structure. To describe a picture, you need well-structured English phrases. Automatically defining image content is very helpful for visually impaired people to better understand the problem. Here, a hybrid system which uses multilayer CNN (Convolutional Neural Network) for generating keywords which narrates given input images and Long Short Term Memory (LSTM) for precisely constructing the significant captions utilizing the obtained words. Convolution Neural Network (CNN) proven to be so effective that they are a way to get to any kind of estimating problem that includes image data as input. LSTM was developed to avoid the poor predictive problem which occurred while using traditional approaches. The model will be trained like when an image is given model produces captions that almost describe the image. The efficiency is demonstrated for the given model using Flickr8K data sets which contains 8000 images and captions for each image.*

Key Words: **Long Short Term Memory, Convolutional Neural Network, Tensorflow, Deep Learning , Recurrent Neural Network , VGG-16.**

1. INTRODUCTION

Generating a caption is a problem under artificial intelligence where meaningful sentence is generated for a given input image. Includes a model from NLP (Natural Language Processing) for converting image to captions as a meaningful sentence. Image captioning contain various types of applications such as advising for editing applications, for image indexing, use in online personal assistants, in social media, for visually impaired people, and other natural language processing applications.

More recently, it has been shown that deep learning models are able to achieve good results in the field of caption predictions. Instead of be in need of composite data editing or a pipeline of specially designed models, one end-to-end model can be defined to captions, if an image is provided. To test our model, we measure its performance using the Flickr8K datasets. These results show that our proposed model works better than standard models in terms of image captions in performance tests.

2. LITERATURE SURVEY

[1] This paper presents the model that generates caption to a given image using pre-trained deep machine learning. This project is built using the python Jupiter environment. In the Keras system backend is built by using the Tensorflow library for training and building deep neural networks. For image processing it uses VGG net and the Keras 2.0 is use to apply the deep convolutional neural network. The caption is obtained when model's output is contrasted with actual human sentence. This comparison is made using models output and analysis of human given captions for the image. Based on this comparison it is concluded that the generated caption by the model is almost same as human given caption. So the accuracy is about 75%, hence the human given sentence and generated sentence is very similar. This project mainly helps in different areas such as aid to the visually impaired people and for different applications.

[2] This paper implemented image caption generator, using a database named Flickr_8k database. This database has various kinds of images it has different types of situations. Flickr_8k data set contains 8000 images and all the picture has 5 captions. Images are divided into 6000 training, 1000 confirmation, 1000 testing. The model was well trained and tested to produce valid captions for given images. The proposed model is based on the division of many labels using the Convolutional Neural network-Recurrent Neural Network method to create captions in proper grammatical structure where CNN acts as an encoder and RNN acts as decoder.

[3] This paper implemented a model using deep learning approach for image captioning. The model includes 3 phases where the first one was Image Feature Extraction. In this phase, the features of the images was extracted using the Xception model. The dataset considered was Flickr 8k. The second phase was Sequence Processor which acts as a embedding layer of word for handling the text input. The embedded layer contains the rules to extract the features which are necessary and will ignore the other values using masking. After this, the network will then be connected to LSTM for captioning of images. The third and last phase considered was Decoder. In this phase, the model will combine the input extractor phase of image and phase for Sequence processor by applying a method and will be sent to a neural layers and later there will be a

final output that is Dense layer which will produce the generation of immediate word required for caption over the language which was created from the typed data which was obtained Processor phase that is in the sequence. Then they also used BLEU score which is an algorithm used for checking the quality of the text translated by machine, so it was used to check the quality of caption generated. The value of BLEU score lies between 0 to 1 .If the score is higher, the quality of caption is better. This project obtained 55.01% of effective BLEU score for Xception model.

[4] This Paper Implements a model which generates suitable descriptions from images. Model has utilized Flickr8K dataset with 8000 images and for each image consists of five descriptions. This project presents a model, which is implemented using neural network which spontaneously observe an image and generates suitable captions in English. The captions or descriptions generated from the model are classified into: Descriptions in the absence of errors, Descriptions containing minimum errors, Descriptions in some measure related to image, Descriptions distinct from image.

[5] This paper provides a quick subjective evaluation contrasting original with altered captions, which is suited for most cases. This paper represents the proof of concept implementation for caption generation. Test image generates the caption for image that are evaluated as subjectively and for the sake of substantial improvements the BLEU score technique was compared. The NLTK package is used for BLEU score calculation for the evaluation set. For image encoding Inception v3 model is used. This project is designed using hybrid RNN/CNN deep network models and implemented image captioning model using deep fusion concept. The word2vec model and inception model is used for textual caption and encoding training images respectively. With the BLEU score they have evaluated the result both objectively and subjectively.

3. SYSTEM ARCHITECTURE

The below figure 1 shows the model for image caption generator. Initially an input image is given which is processed by CNN (Convolutional Neural Network). This CNN will form a feature vector which will be dense in nature. Another name for dense vector is embedding. This embedding can be given as input to various other available algorithms. As output, a descriptive sentence that is caption which is suitable for inputted image. This embedding obtained will be the description of the inputted image. It will be utilized as a first state of Long Short Term Memory (LSTM) which will be helpful for obtaining suitable caption for the given inputted image.

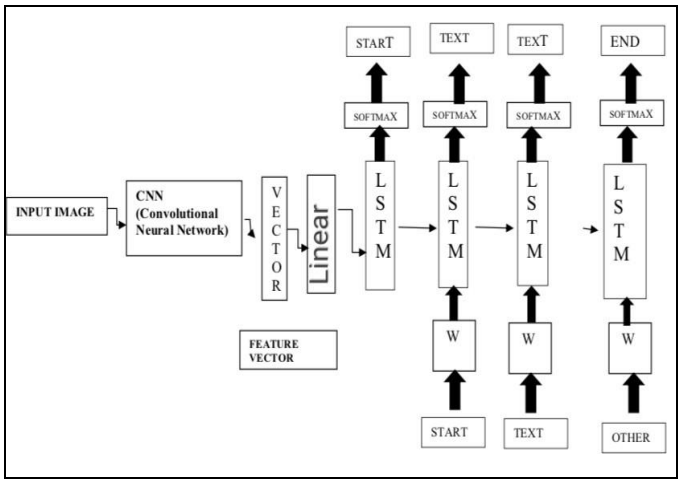

Fig-1: Block diagram of Image Caption Generator

4. PROPOSED IMAGE CAPTION GENERATOR

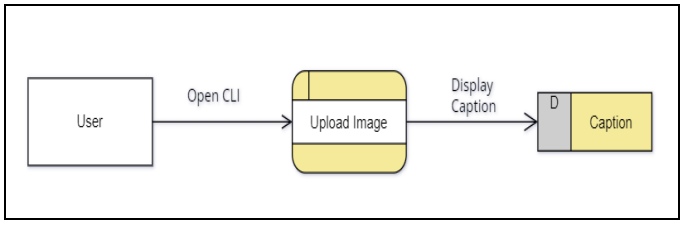

Fig-2: Level-0 DFD

The above Figure 2 is the Data Flow diagram-Level 0, where the user could evaluate the image using command prompt by running the python file with image path as an argument for which he requires caption and later the caption is generated as output.

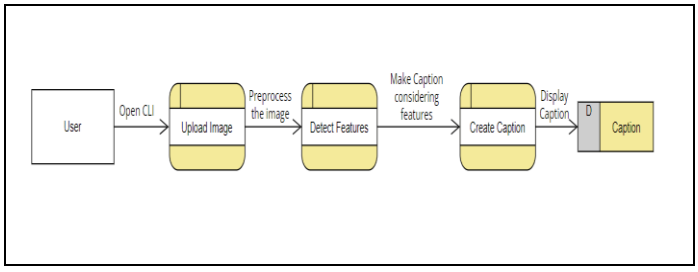

Fig-3: Level-1 DFD

The above Figure 3 is the Data Flow diagram-Level 1, where the user inputs an image in command prompt with image path as an argument for which he requires caption, then the pre-processing of the input image occurs, and later the caption is generated by the features obtained and later the caption is obtained as output.

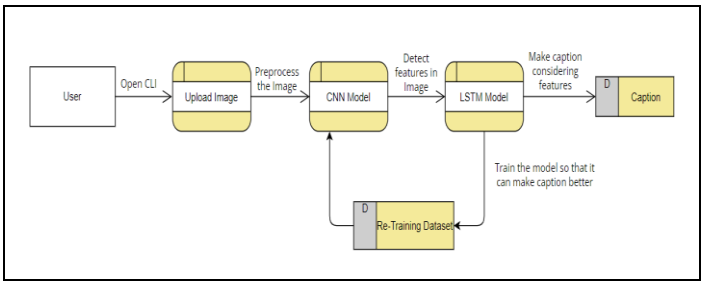

Fig-4: Level-2 DFD

The above Figure 4 is the Data Flow diagram-Level 2, where the user inputs an image in command prompt with image path as an argument for which he requires caption, then the pre-processing of image takes place, and then the CNN (Convolutional Neural Network) model extracts the main features from the image, and considering those features a caption is generated with the help of LSTM (Long Short Term Memory) .During evaluation if re-training is required then the training dataset is again fed into the CNN from LSTM by inputting a particular image else caption is generated as output from the features extracted.

5. CONCLUSION

Image caption methods based on deep learning have made remarkable progress in recent years and it produces high quality captions for every image to be achieved. With emergence of deep learning models, automatically captioning a given input image will always be a functioning study area for some time. The goal of image captioning is very huge in the future since use of social media is rising as the days goes on to upload photos and other purposes. So this project will be of help for greater extent. The proposed model automatically generates captions for an image using Neural Network and Natural Language Processing techniques in Inception V3 model. CNN and LSTM have been combined to work well together and were able to find a connection between objects in images to generate the right caption. Model is trained on Flickr8K dataset and Global Vectors for word representation. The Flickr8k dataset includes about 8000 images, and suitable captions are also saved in a text file. Image Caption Generator is instructed to improvise the probability of the caption when an image is given.

REFERENCES

[1] Sreejith S P, Vijayakumar A (2021) : Image Captioning Generator using Deep Machine Learning.

[2] Preksha Khant, Vishal Deshmukh, Aishwarya Kude, Prachi Kiraula (2021) : Image Caption Generator using CNN-LSTM.

[3] Ali Ashraf Mohamed (2020) : Image Caption Using CNN and LSTM.

[4] Chetan Amritkar, Vaishali Jabade (2018) : Image caption Generation Using Deep Learning Technique.

[5] Subrata Das, Lalit jain, Arup Das (2018) : Deep Learning for Military Image Captioning.

[6] Pranay Mathur , Aman Gill , Aayush Yadav , Anurag Mishra, Nand Kumar Bansode (2017): Camera2Caption :A Real-Time Caption Generator.

[7] Jianhui Chen, Wenqiang Dong, Minchen Li (2015): Image Caption Generator Based On Deep Neural Networks.

[8] Oriol Vinyals, Alexander Toshev, Samy Bengio, Dumitru Erhan (2015) : Show and Tell: A Neural Image Caption Generator.