**An Efficient Deep Learning based Hybrid Model for Image Caption Generation**

In the recent yeas, with the increase in the use of different social media platforms, image captioning approach play a major role in automatically describe the whole image into natural language sentence. Image captioning plays a significant role in computer-based society. Image captioning is the process of automatically generating the natural language textual description of the image using artificial intelligence techniques. Computer vision and natural language processing are the key aspect of the image processing system. Convolutional Neural Network (CNN) is a part of computer vision and used object detection and feature extraction and on the other side Natural Language Processing (NLP) techniques help in generating the textual caption of the image. Generating suitable image description by machine is challenging task as it is based upon object detection, location and their semantic relationships in a human understandable language such as English. In this paper our aim to develop an encoder-decoder based hybrid image captioning approach using VGG16, ResNet50 and YOLO. VGG16 and ResNet50 are the pre-trained feature extraction model which are trained on millions of images. YOLO is used for real time object detection. It first extracts the image features using VGG16, ResNet50 and YOLO and concatenate the result in to single file. At last LSTM and BiGRU are used for textual description of the image. Proposed model is evaluated by using BLEU, METEOR and RUGE score.

**EXISTING SYSTEM:**

In the existing system the amount of data that is generated on the Internet, it has read to the problem of increase in the difficulty in organizing the data. Regardless of the availability of huge amount of raw data it is not feasible to convert them to resourceful data because of its highly unorganized nature.One of the prime issues with the initial image captioning models were that, they considered the image in the form of a single feature vector and thus, they were not able to perceive their relative relationships with the surroundings which may have resulted in a lot of data being wasted.

**DISADVANTAGES OF EXISTING SYSTEM:**

* They were not able to perceive their relative relationships.
* This model was not able to perceive the visual geometry with respect to the image in spite of succeeding in achieving the spatial relationships.

**Algorithm**: Decision Tree(DT), Random Forest(RF), Logistic Regression(LR)

**PROPOSED SYSTEM:**

The model has two Neural Network Architecture convolution neural network (CNN) and the other being Recurrent Neural Network (RNN). An exceptional type of RNN called as LSTM is used here which comprises of memory cell. The reason behind using LSTM is to keep the information for a long period time.We want our model to take an image as an input and give the output as the text description for it.Thus, the input image will be processed by CNN architecture which will give us the output for the CNN model and act as an input for the LSTM model. This will allow the LSTM to generate an output which is describing the image in text format. With this we have also use a pre-trained CNN called ResNet-50 architecture. This will help us to get the spatial information in the provided image.

**ADVANTAGES OF PROPOSED SYSTEM:**

* Model is built on Flickr 8k dataset which has optimal size, large enough to get a good accuracy.
* A combination of ResNet and LSTM is used.
* The image is extracted form a Pre-trained model called ResNet in CNN which is acts as an input of the RNN for the generation of the caption.

**Algorithm**: Convolution Neural network (CNN), Recurrent Neural Network (RNN), Natural Language Processing, Computer Vision, ResNet, Long Short-Term Memory (LSTM).

**SYSTEM REQUIREMENTS:**

**HARDWARE REQUIREMENTS:**

* System : Intel Core i5.
* Hard Disk : 500GB.
* Monitor : 15’’ LED
* Input Devices : Keyboard, Mouse
* Ram : 32GB.

**SOFTWARE REQUIREMENTS:**

* Operating system : Windows 10.
* Coding Language : Python
* Tool : PyCharm, Visual Studio Code
* Database : SQLite

**REFERENCE:**