

SOFTWARE REQUIREMENTS SPECIFICATION DOCUMENT

SUBJECT: SYSTEM DESIGN PRACTICE

**PROJECT TITLE : IMAGE CAPTION GENERATOR
USING DEEP LEARNING (CNN/LSTM)**

Group :

Member - 1 : CE-131 Jay M Sohagiya

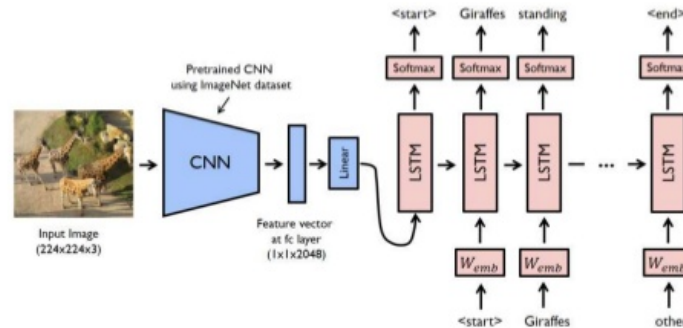
Member - 2 : CE-087 Dev J Patel

Introduction :

Image caption Generator is used to create a sentence description for given image. Our project model will take a picture as input and generate an English sentence as output, describing the contents of the image. It has attracted much research attention in cognitive computing in recent years. The task is quite complex because the concepts of both computer vision and tongue processing domains are combined. In this research study, we proposed a model based on CNN (Convolution neural network) and LSTM (Long- and Short-Term Memory) algorithms to describe image in the sentence form. The CNN works as an encoder to extract features from images and LSTM works as a decoder to generates words describing the image. After the caption generation phase, we use BLEU Scores to gauge the efficiency of our model. Thus, our system helps the user to urge descriptive captions for the given input images.

Detail model Diagram:

Model



Convolutional Neural Network :

Convolutional Neural networks are specialized deep neural networks that process the data that has input shape like a 2D matrix. CNN works well with images and is easily represented as a 2D matrix. Image classification and identification are often easily done using CNN. It can determine whether an image is a bird, a plane or Superman, etc. An important feature of an image can be extracted by scanning the image from left to right and top to bottom and finally, the features are combined to classify images. It can affect the pictures that are translated, rotated, scaled, and changes in perspective.

Long Short-Term Memory :

LSTM is a type of RNN (Recurrent Neural Network) that is well suited for sequence prediction problems. One can predict the next words based on the previous text used. This shows how effectively limitations of RNN are overcome. LSTM can carry out relevant information throughout the processing, it discards nonrelevant information.

Image Caption Generator Model(CNN-RNN model) = CNN + LSTM.

CNN- To extract features from the image.

LSTM- To generate a description from the extracted information of the image.

Dataset :

The Flickr 8K dataset is used for the model training of image caption generators. The dataset can be downloaded directly from the below links. The downloading process takes some time due to the large size of the dataset. The most important file is Flickr 8k.token which is storing all the image names with the captions respectively.

Development Environment :

For development Google Colab is used. Colab is really helpful and hands on to work with due to some amazing features like Efficient UI, importing datasets from kaggle, sharing notebooks to github, preprocessing data, GUI availability ,etc.

FUNCTIONAL REQUIREMENTS:

In software package engineering, a useful demand defines the operation of a code or its element. An operation is represented as a collection of inputs, the behavior, and outputs. useful necessities are also calculations, technical details, information manipulation and process and alternative specific practicality that outline what a system is meant to accomplish. activity necessities describing all the cases wherever the system uses the useful necessities square measure captured in use cases Here, the system has to perform the following tasks:

- It does the Data Cleaning i.e. the Image Processing.
- The model is trained using the above processed image.
 - The processed image is then trained and tested using the validation dataset.
- After finding the accuracy, it is retrained.
- After the required accuracy, it is available for the deployment.

➤ Then it is deployed using Tkinter using python based GUI.

NON-FUNCTIONAL REQUIREMENTS:

1. Performance:

To ensure that our system delivers a good performance, we must take care of certain requirements like a good dataset that is cleaned which can be used for processing, good hosting platform for our website, and a fast and reliable model.

2. Scalability

For our system to be scalable we must work on improving the best available computing power so as to train the model faster and replace it with automatic control as much as possible.

3. Reliability

To make our system reliable we must use the best quality dataset. The dataset should be updated as per the performance and then the model will be able to generate the best caption for the given image.

System Requirements:

- OS: Windows 7 and above, Recommended: Windows 10. CPU: Intel processor with 64-bit support. Disk Storage: 8GB of free disk space.
- Input device : Standard Keyboard and Mouse
- Output device : VGA and High resolution Monitor