



# MALLA REDDY UNIVERSITY

(As per Telangana State Private Universities Act No.11 of 2018 and G.O.Ms.No.14, Higher Education (UE) Department)

## **AD Prototype**

**Name of the Guide : A Siva Kumar**

**Batch Number : ZT18**

**Title of the Project : Image Caption Generator Using LSTM & CNN**

### **Prototype Review:**

#### **Introduction:**

Image Captioning is a task where each image must be understood properly and are able generate suitable caption with proper grammatical structure. Here it is 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 there is 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. We used an encoder-decoder based model that is capable of generating grammatically correct captions for images. This model makes use of VGG16(Visual Geometry Group) as an encoder and LSTM as a decoder. 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 but we use CNN and LSTM to capture dependencies and tell both the spatial relationships of images and contextual information of captions and generate contextually relevant captions.

#### **Problem Statement:**

Image captioning presents a significant challenge in Artificial Intelligence, requiring systems to understand images and generate grammatically correct descriptions. The existing methods facing issues in accurately describing images with proper context. To address this, a hybrid system Convolutional Neural Networks (CNN) for feature extraction and Long Short-Term Memory (LSTM) networks for caption generation is proposed. However, the existing system struggling to capture suitable image differences and contextual nuances. Therefore, there is a need for an improved image captioning system that effectively integrates CNN and LSTM models to accurately describe images with contextual relevance. This research aims to develop such a system, utilizing the Flickr8K dataset for training and evaluation. The system efficiency will increased by generating suitable captions for given images.

## Process of the Project:

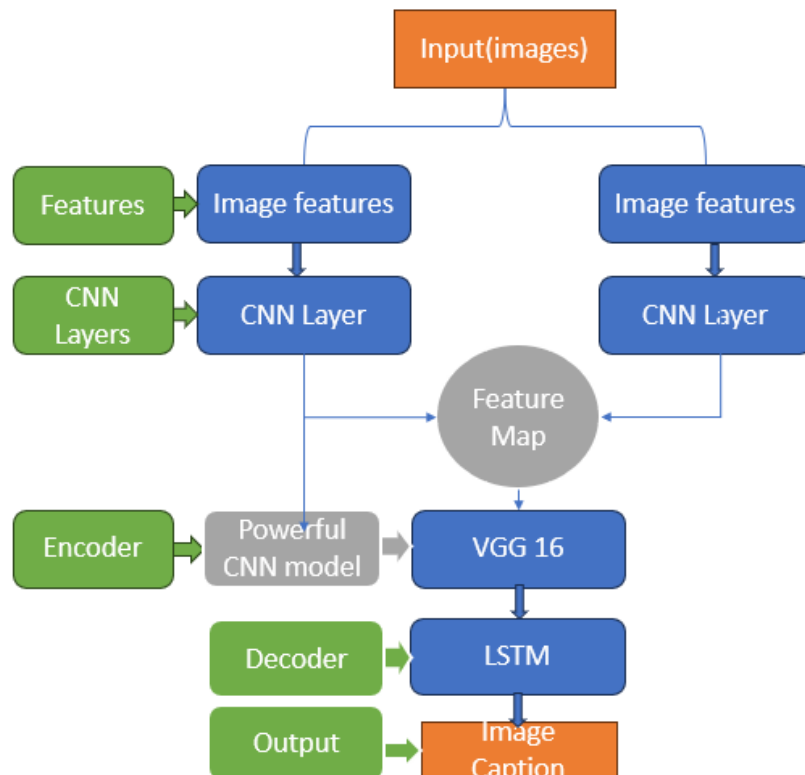


Figure 1- Process of the Project

## Results:




Figure 2- A predicted caption generated by model by using actual captions as training data

```
[ ] from nltk.translate.bleu_score import corpus_bleu
# validate with test data
actual, predicted = list(), list()

for key in tqdm(test):
    # get actual caption
    captions = mapping[key]
    # predict the caption for image
    y_pred = predict_caption(model, features[key], tokenizer, max_length)
    # split into words
    actual_captions = [caption.split() for caption in captions]
    y_pred = y_pred.split()
    # append to the list
    actual.append(actual_captions)
    predicted.append(y_pred)

# calculate BLEU score
print("BLEU-1: %f" % corpus_bleu(actual, predicted, weights=(1.0, 0, 0, 0)))
print("BLEU-2: %f" % corpus_bleu(actual, predicted, weights=(0.5, 0.5, 0, 0)))
```

100%  810/810 [11:00<00:00, 1.43it/s]

BLEU-1: 0.533333  
BLEU-2: 0.306618

Figure 3- BLEU score of our Model

## Conclusion:

In conclusion, we have successfully developed a prototype for Image Caption Generator using deep learning techniques. By leveraging a pre-trained convolutional neural network model and transfer learning, we have created a robust system capable of accurately identifying the caption from input images and captions. The prototype demonstrated promising performance, achieving high accuracy on various breeds across different test scenarios. However, there is room for further improvement, such as fine-tuning the model on a larger and more diverse dataset to enhance its generalization capabilities. Overall, this project lays the foundation for building more advanced and practical solutions for automated image identification, with potential applications in image identifying and captcha generators for a better security.