**IMAGE CAPTION GENERATOR USING CNN & LSTM**

**Objective:**

Based on the content of an Image, this application will generate a caption for any natural image. Such an application might help blind people to see the world full with images. This model automatically generates natural language captions which then can be utilized for indexing and searching of images, tagging in social media, helping the visually impaired etc. The uses of such an application is immense. In this paper, we will build such an application using Convolution Neural Networks (CNN) for feature extraction and Long Short Term Memory (LSTM) for generating the captions.

**Project flow:**

The flow for the project is given below:

List of words

Input Image

**Sequence Processer**

**Feature Extractor**

CNN model Xception

Embedding Layer (Reducing size to 256 nodes)

Feature Vector (2048 features)

LSTM Layer

Reduction to 256 nodes using Dense

Merging using Dense layer

softmax

Output Caption

**Advantages of Project:**

* Image searching using context.
* Visually impaired can listen to the caption of an Image.
* Classifying different photos in our phone gallery automatically for better management.
* Self driving vehicles like Tesla, Waymo etc. can use image captioning to better understand the images.
* Social media platforms like Facebook uses image captioning to understand the users and their behaviors.

**Software and Hardware Requirements:**

* Python v3.6+
* RAM: 4GB minimum (8GB recommended especially if using only CPU to process data).
* Dedicated graphics card (optional)