**SCENE PORTRAYER**

**ABSTRACT**

**Problem Scenario**

You saw an image and your brain can easily tell what the image is about, but can a computer tell what the image is representing? With the advancement in Deep Learning techniques and availability of huge datasets and computer power, it is possible to build models that can generate captions for an image.

**Problem Solution**

A Scene Portrayer is a Machine Learning application, which identifies the action portrayed in the given image. The generated caption will describe about the image that will say what kind of actions is taking place in it.

This project involves Computer Vision and Natural Language Processing concepts to recognize the context of an image and describe them in a natural language like English.

The objective of the project is to build a working model of Caption Generator by implementing CNN with LSTM.

The basic working of the application is that the features are extracted from the images using pre-trained **VGG16 model** and then fed to the **LSTM model** along with the captions to train.

The trained model is then capable of generating captions for any images that are fed to it.

**Methodology**

**Convolutional Neural Networks (CNN)**

With the help of CNN Algorithm, features are extracted from the images with the help of pre-trained VGG16 model.

CNN Algorithm is being used in this project as they are specialized deep neural networks which can process the data that has input shape like a 2D matrix. Images are easily represented as a 2D matrix and CNN is very useful in working with images.

**LSTM (Long Short Term Memory)**

With the features being extracted from the images with the help of CNN Algorithm, they are now fed into the LSTM model which will be responsible for generating the image captions.

LSTM is a type of RNN which is well suited for sequence prediction problems. With the help of LSTM, it can predict what the next word will be. LSTM will use the information from CNN to help generate a description of the image.

**Model Generation**

Once both the algorithms are grouped together, the images are loaded onto the Pretrained CNN model. With the features being extracted from it, it is now loaded onto LSTM model which will generate the captions for the images. Thus, the generated caption will describe what kind of action is being portrayed in the image.

**Tools and Technologies**

1. **Frontend:** HTML5, CSS3 and Bootstrap4
2. **Backend:** Flask 1.x
3. **Language:** Python 3.7
4. **Storage (Dataset)**: Google Drive
5. **Python libraries**: Matplotlib, Tensorflow GPU, Pre-trained VGG-16 weights, NLTK, Keras, Numpy, Pandas, Pillow
6. **Flickr\_8K dataset**

**Application in the real world**

1. If the caption is converted to speech, it will be helpful for the visually impaired person.
2. Auto subtitling.