

Project Overview

This project focuses on solving the problem of automatically generating captions for an image using deep learning techniques. We make use of advanced deep learning concepts such as CNN and LSTM (Long Short Term Memory) for better performance. We will be making use of the Flickr8k dataset that contains 8000 well annotated images for training of this model. The dataset is used as it contains different sources of descriptions for the images.

Problem Definition

Automatically describing the content of an image is a fundamental problem in artificial intelligence that connects computer vision and natural language processing. Image captioning is a challenging problem owing to the complexity in understanding the image content and diverse ways of describing it in natural language. It is interesting because it aims at endowing machines with one of the core human intelligence to understand the huge amount of visual information and to express it in natural language.

Objective

- To implement an algorithm for automatic captioning of images.
- To reduce the human intervention needed for captioning of images.
- To improve the performance of image captioning models.
- To check the performance of combining LSTM layer with CNN (Convolutional Neural Networks)