Problem Statement:

Automatically describing the content of an image is a fundamental problem in artificial intelligence that connects computer vision and natural language processing. Being able to automatically describe the content of an image using properly formed English sentences is a very challenging task. Indeed, a description must capture not only the objects contained in an image, but it also must express how these objects relate to each other as well as their attributes and the activities they are involved in. Moreover, the above semantic knowledge has to be expressed in a natural language like English, which means that a language model is needed in addition to visual understanding.

Aim:

In this project, we present a generative model based on a deep recurrent architecture that combines recent advances in computer vision and machine translation and that can be used to generate natural sentences describing an image. The model trained to maximize the likelihood of the target description sentence given the training image.