ABSTRACT

Portraying the substance of an image may be troublesome. It requires the discovery and acknowledgment of things, individuals, connections, and related characteristics to empower nitty gritty depiction. Presently, most of study is comprehensive in nature, which can miss particulars relating to key components of a scene. We offer a shiny new locale based profound learning design for making picture portrayals to address this trouble.

One of the applications of deep learning is image caption generation or photo description. In this scenario, the model must be given the image to process. The model then applies its training to process the image and provide captions or descriptions. Sometimes, this prediction is not very accurate and produces a few sentences that have no real sense. To get better results, we need highly powerful computers and a very huge dataset. The dataset and the neural network architecture for the image captions generator will now be covered in some detail.

Experts in both image and natural language processing are interested in the multidisciplinary field of artificial intelligence (AI) research that is picture captioning. Image captioning is a challenging topic since it frequently calls for obtaining data that isn't immediately visible in a scene. It could need either a straightforward interpretation or in-depth familiarity with the subject matter of the image.The most cutting-edge technologies now available are deep learning and machine learning. The human mind and artificial intelligence are increasingly being compared, with AI performing better than humans in several fields.

In this work, we use CNN and LSTM to comprehend the caption of the image. An image caption generation system also has to be familiar with computer vision standards and natural language processing in order to grasp the link between the image and the English description.Also we are employing a pre-trained model called VGG16 for image detection. VGG16 is already installed in the Keras library. VGG 16 was introduced in 2014 by Karen Simonyan and Andrew Zisserman of the Visual Geometry Group Lab at Oxford University.