**ABSTRACT**

In the recent years, with the increase in the use of different social media platforms, image captioning approach play a major role in automatically describe the whole image into natural language sentence. Image captioning plays a significant role in computer-based society. Image captioning is the process of automatically generating the natural language textual description of the image using artificial intelligence techniques. Computer vision and natural language processing are the key aspect of the image processing system. Convolutional Neural Network (CNN) is a part of computer vision and used object detection and feature extraction and on the other side Natural Language Processing (NLP) techniques help in generating the textual caption of the image. Generating suitable image description by machine is challenging task as it is based upon object detection, location and their semantic relationships in a human understandable language such as English. In this paper our aim to develop an encoder-decoder based hybrid image captioning approach using VGG16, ResNet50 and YOLO. VGG16 and ResNet50 are the pre-trained feature extraction model which are trained on millions of images. YOLO is used for real time object detection. It first extracts the image features using VGG16, ResNet50 and YOLO and concatenate the result in to single file. At last LSTM and BiGRU are used for textual description of the image. Proposed model is evaluated by using BLEU, METEOR and RUGE score.