The combination of CNN and RNN model is best model to generate an automated caption. This model takes image as input and produces captions as output.

Starting every year from 2010 and the MSCOCO Image Captioning Challenge in 2015, the ImageNet Challenge has given rise to new state of the art performances in the task of image classification. A new architecture called Inception was introduced by GoogLeNet, which began a new era in the field of image classification, which is also a fundamental block in this paradigm of image captioning. In order to represent images, the NIC model generates a single network. To identify images using deep convolutionary networks, this network is used.

Describing images with a neural and probabilistic system is the proposed NIC. A robust sequence model was given by the development of statistical machine translation. It will increase the likelihood of

accurate translation by presenting "end to end" input sentences, both the training process and its predictions. A CNN is used to display the input image. They are commonly used for object detection and identification and are presently state-of-the- art. The CNN version used in this demonstration was the one that performed best in the 2014 ILSVRC classification competition.

The training dataset consisted of pictures and sentences explaining those images in English. It used a total of 82783 samples from the MSCOCO dataset. For the test, another 40775 samples were isolated.

Not all translation processes fit the correct phrase to the target language from the source language,

since there can be several correct responses to that. In this case, assessment involves raters to

subjectively assess the utility of each image definition, but this is obviously time-consuming. Both subjective and automated tests have been improved to test the NIC model, and it has

been shown that there is also a connection between these two scores.

Due to a lack of funding, there were many limitations to carrying out this work. It was only possible to

train the last layer of the NIC model and do the demonstration because of the improvement in transfer

learning. It is important to have a strong resource history to perform tests after tuning various

hyperparameters of any model to the degree to which machine learning research is going on right now.

The results provided by current models are very promising, and no indications are known.