Name: Fahad Ahmed Pranto

Id: 1612607042

Paper name: **Automated image captioning.**

Paper link: <https://ieeexplore.ieee.org/abstract/document/9129930>

**Abstract**: Research into the understanding of machine learning algorithms and techniques appropriate for the processing of images and the processing of natural languages. The facilities used in the implementation of machine learning algorithms with current packages. The introduction of an algorithm that takes a picture and explains in lengthy sentences following the assessed parameters, for familiarization with the domain and potential models, a bibliographic analysis was carried out. This paper introduces a hybrid model that consists of a deep model that convolutionary neural network for extraction of features that makes transform learning use and a recurrent neural network for constructing the details.

**Introduction:** The purpose of this article is to obtain useful textual details in the form of actual descriptions, from pictures. People can extract knowledge from pictures as well. If a picture showing a flower is shown to an individual, provided that the individual has seen a flower before, she or he will be able to recognize the object. But if that person is visually impaired, they can't see that picture at all. The outcomes can be furthermore, it will be run through a text-to - speech engine to give complete Sustainable. In this way, it could be a fully independent experience, Delivered to persons with this impairment. These people might feel like they are losing out, particularly when it comes to their relatives and close friends, with the increase of social media interactions posting online images.

**Literature review:** The deep neural network decoder receives from the encoder the encrypted characteristics. That recurrent neural network architecture is a Gated Recurrent Unit (GRU) type. For predict the next term, the decoder watches over the picture. They performed the data preprocessing part Skipping this segment results in an error being raised by the model as it does not receive the planned data. If the dataset is small sufficient, it can be transferred in one batch, in its totality, to the algorithm, but the dataset being used is very large and involves batching. Batching means that the dataset is split into many equal sections and pieces feed them to the algorithm one at a time.

**Methods**: In any machine learning model, data preprocessing reflects a very significant step. Two different kinds of data preparation are needed for this application: one for the deep convolutional neural network encoder and one for the deep recurrent neural network decoder. Then, Feature extraction via the CNN encoder. Passing the output of the encoder, initializing the hidden state to 0 (zero) and the input of the decoder (the "start" token) to the decoder. The hidden state of the decoder is returned in a loop. Calculating the loss is used in the design and its assumptions. Calculate the gradient, add it to the optimizer and backpropagate it via the algorithm. The selected loss function is cross-entropy with logs. To measure the actual error values for each lot, it is initialized to zero before passing each set. At each epoch, the cumulative loss is initialized to zero.

**Datasets:** MSCOCO, which stands for 'Microsoft-Common Objects in Context', is the dataset used to train and evaluate the model. It was first published in 2014 and its last revision was in 2015.

**Results:** The computer on which the models are trained does not have a lot of computing power and it took a lot of time to train for a relatively low number of epochs. To solve this problem their input data size is decreased. After reducing the number of training data, the number of remains 20, printing the actual and predicted captions. A woman playing tennis is represented in one of the images, while the other shows people in boats across a river. The first should be regarded easy to read, while the second is a little harder to understand. The results are terrible, but the algorithm work. Next, the number of epochs is increased to 50. The first image predictor has improved the ball is detected, while the prediction of the second image is off. This makes it clear that over sufficient iterations, the models were not educated.

**Conclusion:** The study portion of the paper was successful, enabling one to interpret the findings of both machine learning and deep learning algorithms. The basic elements of artificial neural networks, such as feed-forward, backpropagation of the error via a neural perceptron, were studied. Network, potential loss function, and features for activation. The composite model, consisting of a convolutionary neural network, was chosen as the encoder. This uses learning algorithm, thereby lowering training time and computational complexity and, as the decoder, a recurrent neural network.When it comes to coping with textual data, which is the top of the line, especially anticipating the next word in a phrase