Name: Fahad Ahmed Pranto

Id: 1612607042

Paper name: **Bangla language textual image description by hybrid neural network model**.

**Abstract:**

The role of automated image captioning in various languages is a complex task that has still not been well studied due to the lack of dataset and models that are efficient. For comprehensive semantic representation of images for natural language image descriptors, it also includes clear comprehension of scene and conceptual embedding. They created a new Bangla dataset of images matched with the target language name, called the Bangla natural language image to text (BNLIT) dataset, in order to produce an image descriptor in Bangla. They suggest a hybrid encoder-decoder model based on the encoder-decoder architecture to cope with image comprehension, and the model is tested on the newly generated dataset. Our hybrid model uses the convolutionary neural network as an encoder, whereas the long-term symmetric memory is used for the description of sentences that reduces the computational difficulties without sharing the descriptor precision.

**Introduction:**

Image interpretation is simply the written description of an image based on expression, and has been an active area of computer image processing and natural language processing. Due to its many realistic uses, such as text-based image search, image curation, allowing visually impaired people to better understand the actual image search, image captioning has attracted a lot of researchers' attention. One of the key difficulties is to construct a dataset in the target language to create an image captioning model. It is important to enhance the visual significance of the image descriptor of an image to establish an image caption generation model. Also it is essential to note how to apply contextual semantic embedding to a picture's multiple scenarios. They suggest a hybrid encoder-decoder paradigm to address these difficulties in captioning tasks, and the difficult part of encoder-decoder architectures is to model the application that controls the details, Flow between the model constructs of applied CNN [14], long short-term memory (LSTM)[15] and bi-directional neural networks (BRNN)[16, 17].

**Methods:**

To confirm that each image stays in the same pixels, they resize the full dataset images. Firstly, using CNN and VGG16 characteristics, they classify the full dataset. They are doing the 30-class classification. BRNN is primarily used to generate text from the input images provided. BRNN is used primarily for producing text from the input images provided. They finally, integrated their dataset with both model features of CNN, LSTM, and BRNN and train up the full model.

**Datasets:**

A new dataset was created, named BNLIT and 8,743 images are included in that dataset. To create a new data set, they choose Bangladesh perspective images.

**Results:**

Using the BNLIT dataset, they implemented their hybrid model. Using their self-made Bangla dataset, they observed that their model provides good precision. They see how this new dataset can obtain learning from dataset and image categorization using VGG16 during the categorization of the use of CNN. They get better accuracy, which is 0.794538 accuracy of training time and 0.782161 which is the accuracy of validation for the CNN result BNLIT dataset. In addition, in the BRNN and LSTM period, they received 0.8739 training time accuracy. Then both the model and the full dataset were integrated again and their accuracy finally reached 0.942546 for training time and 0.758651 for validation.

**Conclusion:**

A complex hybrid neural network model is adopted in this research, which illustrates an exceptional ability to create a single sentence representation based on Bangla natural language from a given test image. The model is able to identify images with encoded multimodal and semantic difficulties and is capable of generating a description of the natural language based on the image context.