Name: Fahad Ahmed Pranto

Id: 1612607042

Paper name: **Oboyob: A sequential-semantic Bengali image captioning engine.** Paper link: //www.sci-hub.ren/https://content.iospress.com/articles/journal-of-intelligent-and-fuzzy-systems/ifs179351

**Abstract:** An effective and challenging research subject in machine learning and natural language processing is to understanding the meaning of the generation of textual description from an input image. In the context of the Bengali language, moreover, the issue remains undeveloped. By subsampling the machine readable dataset, they examine a standard method for Bengali image caption generation. After, with the state-of-the-art CNN-LSTM architecture- based pre-processing techniques, they use some models. For further research with 16 distinct models built in the whole training process, the training description subsampled dataset is constructed for both Bengali and English languages. With respect to many caption assessment criteria, the trained models for both languages are evaluated.

**Introduction**: While not missing its underlying meaning, there are several ways to convey an image. Different explanations may provide different viewpoints on how the viewer considers an image. Keeping these things into consideration, it has become a challenge for researchers in machine learning and natural language processing to achieve the sentiment analysis definition of an image in various languages automatically. These methods of sentence representation of an image are known as image captioning, which specifically addresses two difficulties. The first difficulty is to recognize objects in an image, and the second is to establish a connection in the field of natural language processing between objects and sentence-level descriptions.

**Literature review:** For automatic captioning, there are two approaches: bottom-up and top-down. Various words are accumulated to correspond with a picture and the words construct sentences depending on the first method. However, due to different continuity, certain features remain overlooked at the same time. The state-of - art policy, moreover, is the top-down approach that all relevant stimuli is collected via the Recurrent Neural Network (RNN).

**Methods:** Several approaches are used, from function extraction to final proposed scheme. Firstly, their dataset is analyzed and the Bilingual caption selection metric has been used. Secondly, for CNN feature extraction, they employed Inception-ResNet and VGG-16. Finally, they create their model architecture that includes description in visual and linguistic terms.

**Dataset:** The experiment is performed on the standard Flickr-8 K dataset, including many changes to the Bengali language adaptation. It Consist of a total of 8092 images taken fromFlickr2. The corresponding image IDs are categorized into training, validation and testing, where 6000 are used for preparation, 1000 for validation and 1000 for testing.

**Result:** For languages, they find mixed findings. The external pre-trained word embedding description, regarding English language, leads inner vocabulary due to having varied vocabulary for robust description, while internal vocabulary shows better scores for Bengali language in the statistics. This outcome effectively shows that the existing Bengali pre-trained embedding model still needs to be improved with various language tokens.

**Conclusion:** In image captioning research, this study links the Bengali language by implementing a traditional experimental analysis and provides a comparison analysis of recent developments and techniques currently used in this sector. Fix resource constraints in Bengali that have higher genetic difficulty and build a dataset for machine translation. Then they implement a new bilingual sentence selection method.