Name: Fahad Ahmed Pranto

Id: 1612607042

Paper name: **Hybrid deep neural network for Bangla automated image descriptor.**

Paper link: <https://www.researchgate.net/profile/Md_Jishan/publication/343323240_Hybrid_deep_neural_network_for_Bangla_automated_image_descriptor/links/5f232ee7299bf1340494b63b/Hybrid-deep-neural-network-for-Bangla-automated-image-descriptor.pdf>

**Abstract:** Automated text generation for an image is a computer vision activity that involves a sufficient understanding of the syntactic and semantic context of an image in order to produce a meaningful description. Due to the lack of graphic-descriptor datasets and functional models to capture underlying complexities involving characteristics of an image, it has been researched to a limited extent. A unique dataset was created in this study by generating a Bangla textual descriptor from visual input, named Bangla Natural Language Image to Text (BNLIT), including 100 annotation classes. For image description generation, a deep neural network-based image captioning model was suggested. The model uses the Convolutional Neural Network (CNN) to identify the entire dataset, while the sequential semantic representation of text-based sentences is captured by the Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) and the related description is created based on the structural characteristics of a picture.

**Introduction:** In view of the rapid developments in natural language processing, this study aims to produce Bangla textual captions of contextual images to benefit the Bangla speaking population. Almost all of the research into image captioning is carried out in English. A main subject that produces a simple language text from the input dataset is the image captioning model. The key objective is to define the entire dataset and then, using a particular optimization technique, apply a hybrid model for producing text. To construct a framework for creating image captions, the visual interface of the image narrator needs to be improved and promoted. For instance, why the models retrieve the background, detect an image region, and then create the caption of the image that is compatible with the image data.

**Literature review:** They resize the images of the datasets through training and testing process to confirm higher multi-inclusive statements and to preserve a strategic distance from any numerical irregularity. They prefer to use raw data set image documents near the examples of CNN and VGG16.To remove the concern about the RNN , they picked one bangla annotation for each picture.

**Methods:** The Convolutional Neural Network (CNN) is a deep learning algorithm that extracts and differentiates features from those in the image taken. They used four main layers of the architecture of CNN: Convolutional Layer, Pooling Layer, Rectified non-linear unit, and Fully-Connected Layer.

**Dataset:** They introduced a new dataset titled BNLIT [12] that is comprised of a gallery of 8,743 photos representing the life, heritage, ethnicity & culture of our country Bangladesh where every image speaks with its language.

**Results:** A hybrid neural image captioning model has been implemented that is capable of producing a realistic representation of the given image based on Bangla text. In order to get accustomed with the relation between better bits of the image alongside the related segment of the sentences, they prepared our model. We used measures of grouping precision to calculate accuracy. After successfully training up, they evaluated their model and represent that result to the readers.

**Conclusion**: They suggested a complex neural network with enough deep structure that, by understanding complexities in the content of an image, conveniently generates a Bangla natural language sentence. Using the combination of CNN, RNN, and LSTM architecture, they implemented their proposed model and achieved benchmark precision for our self-made dataset.