Paper: **An Automated Bangla Image Captioning System (Chittron)**

Name: Milon Hossain

Id: 1610497042

**Abstract:** Automatic image caption generation intends to automatically create a meaningful explanation of an image in a natural language. Nevertheless, Bangla, the world’s fifth most commonly spoken language, is slightly lagging behind in science and growth of those domain. In comparison, though there is no such resource for Bangla yet for proven data sets relevant to image annotation in English. This paper therefore describes the development in Bangla “Chittron”, an automated image captioning system. I this system there will be a model which will take an image as an input and describe about the image as an output. The model is learned to predict a caption, one word at a time, when the input is an image.

**Dataset:** As we know there not much resources in deep learning or computer vision system for Bangla languages. To address the dataset availability issue, a collection of 16,000 Bangladeshi contextual images has been accumulated and manually annotated in Bangla. This data set is then used to train a model which integrates a pre-trained VGG16 image embedding model stacked with LSTM layers.

**Method:** For this system various deep learning concept had been used like “CNN”, “RNN”, “LSTM”, “GRU” and so on. Image captioning evaluation metrics like BLEU and METEOR also used to generate the most appropriate captions. Using these terms, they made a model and trained it up. Finally, the model generates a caption when the input is an image.

**Complexity:** Main complexity is the available issue for the data set. There are not so many data sets are available for Bangla language. For this reason, contextual image has been managed and annotated in Bangla.

**Improvement:** A significant use of large number of data set into the model can give better output. When the model is being trained if we use huge number of data set the model will be more accurate and show us proper output. Uses of image captioning evaluation metrics can also improve the output accuracy.

**Conclusion:** We can conclude that in this paper represents a system to automatically generating Bangla image captioning using deep neural networks and the collection of Bangla Lekha Image Captions data set consists of 16000 images, with a single overly-descriptive caption per image. This data set is used to train a model employing a pre-trained VGG16 model and stacked with LSTM layers. The VGG16 model used to extract a rich description of the image content as an image embedding vectors. Finally, through these processes the model is prepare to use and captioning the image accurately.

**Reference:**

1. https://www.semanticscholar.org/paper/Chittron%3A-An-Automatic-Bangla-Image-Captioning-Rahman-Mohammed/e1c4520309fcb3108cf83ac6630222e7956ebe9f