Chittron: An Automatic Bangla Image Captioning System

Automatic image caption creation strives to automatically create an accurate summary of an image in a natural language. The stark fact is that the bulk of image captioning works have focused almost entirely on English Idioma [1, 2, 3]. In comparison, the related data sets, such as MSCOCO[4], have a strong Western preference. That leads to a two-pronged problem: (1) only English is the language in which captions are produced, and (2) the The data collection is not indicative of cultural characteristics.

A naive approach to creating an image caption is perhaps to infer words from the regions of the image. And hook them up. Back in 1999, one of the first image annotation systems[5] was created. Later on, in 2002, The role of image captioning was re-cast as that of machine translation [6]. It turned out, though, that the procedure For a variety of causes, including inability to perform properly,

Data sets such as MSCOCO, with 5 captions per file, have 200, 000 images. There are extensive BanglaLekha-ImageCaptions Smaller in number, with just 16, 000 photographs, all taken from the web's public domain with comparatively smaller files. Topic matters diversified. Nearly all of the photographs are in some way linked to Bangladesh, with some being significant.To the broader Subcontinental sense of India.

We will address the model used in the proposed framework in this chapter. The debate was divided into two sections.Where one concentrates on the planning for the model and the other reflects on the setting for practicing. This work predicts captions of lengths of up to 10 tokens. Thus, the caption is first for each image-caption pair,Truncated to 10 tokens or expanded. The extension is achieved by padding the undefined token with tokens.

In this section, both quantitative and qualitative studies are discussed. Quantitative outcomes are stated in terms of BLEU score[16], which is a commonly used criterion for machine translation applications evaluation.The model received an average BLEU score of 2.5, which is not remarkable, admittedly. BLEU ratings, however,In cases where several reference sentences are available, they are usually measured.

This paper reports on the creation of "Chittron," a method that uses Deep Neural Networks to automatically produce Bangla image captions and the BanglaLekha-ImageCaptions data set collection consists of 16000 images, with a single excessively detailed caption per image. To train a model using a pre-trained VGG16 model and stacked LSTM layers, this data set is used.