

Image Captioning in Tamil Language with Merge Architecture

G Rajalingam# and WU Wickramaarachchi

Department of Computing, Rajarata University of Sri Lanka, Mihintale, Sri Lanka

#gobigarajalingam@gmail.com

Image Captioning is the process of describing the content of an image using a natural language. This task that involves computer vision and natural language processing has been attempted on the English language with enormous success, owing to the presence of massive image-caption paired corpora as Flickr and Microsoft Common Objects in Context (MS-COCO). However, such developments in this arena have been a novelty for non-English languages with the exception of a few such as Chinese, Turkish, German and Arabic. In the case of Tamil language, this premise has been barely touched upon, due to the lack of a large, paired corpus. In this work, a paired corpus inspired from Flickr30K dataset has been created in Tamil language for the image captioning purpose. Along with it, this paper includes the experiments with an image captioning model, using a combination of Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) architecture; specifically the Merge model for Tamil language caption generation. This methodology incorporates the image vectors in a layer following the LSTM layer. The results of the research have proven satisfactory in the evaluation with a Bilingual Evaluation Understudy (BLEU) score of 0.37, and this indicates further development with the presence of a more refined and improved dataset.

Keywords: *Tamil caption generation, convolutional neural network, long short-term memory, natural language processing*