A20

Efficient Image Captioning with Attention Mechanisms: Leveraging CNNs and Simplified GRUs

Abstract:

Image captioning, the task of automatically generating descriptive sentences for images, is a challenging problem in computer vision and natural language processing with significant applications, including, but not limited to, helping visually impaired people and providing better ways of indexing images, and summarizing image content on social networks. This work is based on deep neural network solution for this problem by using the CNN for the image part and the RNN for the sentence part. In this case, Inception v3 is a more efficient CNN architecture, and in order to determine the best architecture for feature extraction we will test a few other state-of-the-art CNN architectures. Further, we suggest using Gated Recurrent Units (GRU), which is a simpler version of LSTM and contains less number of parameters, which helps to improve the rate of training without any compromise in the performance level. This shall be achieved through the use of attention mechanisms which enables the model to draw attention to important parts of the image during the generation of captions just as the human would. By applying means like Greedy Search in producing sentences, we believe that the anticipated result is going to be an improved image captioning model with balanced accuracy and simplicity in terms of the number of computations to be made. This project intends to help in improving the existing for image captioning solutions, as well as trying to make these as accessible as possible.

Key words: Image Captioning, Attention Mechanisms, Convolutional Neural Network(CNN) Gated Recurring Units(GRUs), Deep Learning

Team Members:

(21241A1256) Sriram Chiliveri

(21241A1246) P. Mahesh Babu

(21241A1265) Yenamala Varshit

Signature of the Guide:

Dr. R. V. S. S. S. Nagini

Associate Professor