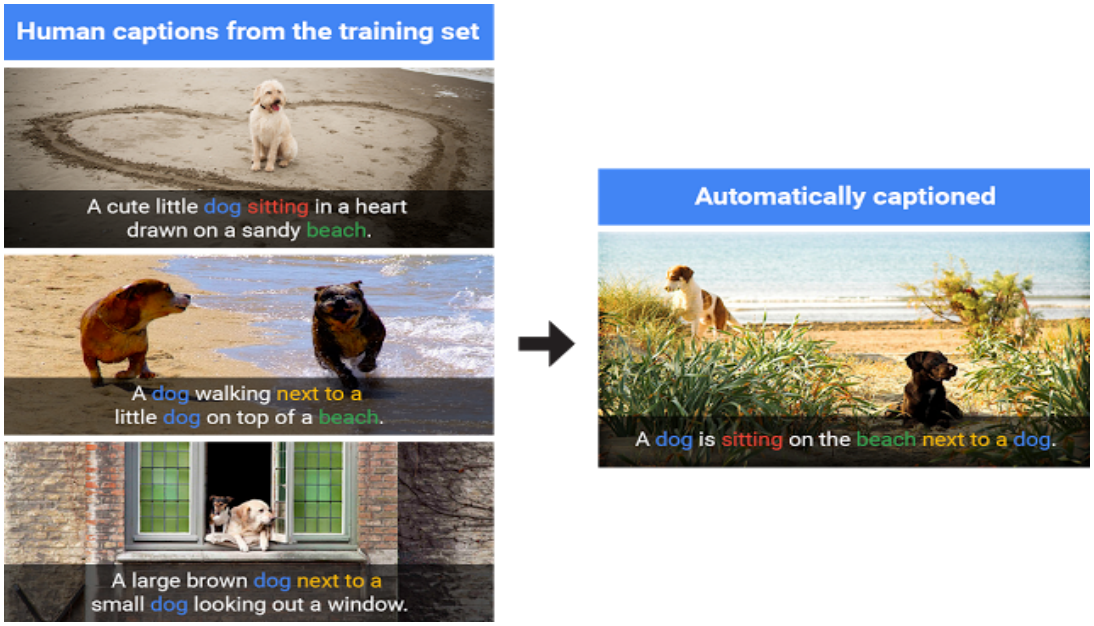
INFO7390 Advanced Data Sciences and Architecture Project Proposal

Topic : Image summary generation or caption generation for images using CNN-LSTM and GRU network



Ankana Asit Baran Samanta 001007431

Kshitij Zutshi 001021288

Parth Shah 001006181

Abstract

Image caption generation/ image summarization is a task that involves generating a semantic description of an image in natural language and is currently accomplished by techniques that use a combination of computer vision (CV), natural language processing (NLP), and machine learning methods. The inspiration for such an application can be inferred from Social media platforms like Facebook(Now Meta), that summarize the image posted by the user and infer details like - where you are, what you wear etc. This application also has a profound use in assisting visually impared individuals in comprehending the images of the real world. In this task, we work on a model that generates natural language description of an image. We intend to use a combination of convolutional neural networks to extract features and then use recurrent neural networks to generate text from these features. We incorporated the attention mechanism while generating captions. We evaluated the model on the MSCOCO database.

Problem Statement

Given an image, we want to obtain a sentence that describes what the image consists of.

Dataset

The dataset that we are going to be using is from MSCOCO (Common objects in context) dataset consisting of images and their annotations

https://cocodataset.org/#download

Approach

For image caption generation, we will be primarily using two models :

1. **LSTM (Long Short Term Memory)**
2. **GRU (Gated Recurrent Unit)**

For better accuracy we will be extending our model through Hyperparameter Tuning and provide visualization for the outputs and the results of the models.

References

https://cs224d.stanford.edu/reports/msoh.pdf

https://iopscience.iop.org/article/10.1088/1742-6596/1712/1/012015/pdf

https://arxiv.org/pdf/1412.3555.pdf

https://www.freecodecamp.org/news/building-an-image-caption-generator-with-deep-learning-in-tensorflow-a142722e9b1f

https://vision.cornell.edu/se3/wp-content/uploads/2018/03/1501.pdf