Ninad Parikh(A20427382)   Harsh Vora(A20445400)

Auto Image Captioning

Research Paper Used:

Paper: Show and Tell: Lessons learned from the 2015 MSCOCO Image Captioning Challenge
Authors: Oriol Vinyals, Alexander Toshev, Samy Bengio and Dumitru Erhan
Year of Publication: 2016
Name of Publication: IEEE
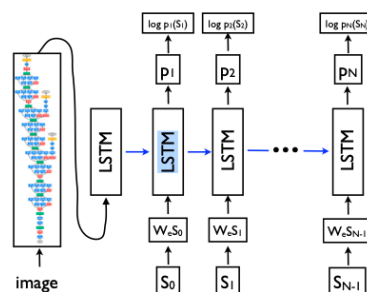Link: https://arxiv.org/pdf/1609.06647v1.pdf

Abstract:

Automatically describing the content of an image is a fundamental problem in artificial intelligence that connects computer vision and natural language processing. For doing that, the content of an image using properly formed English sentences is a very challenging task, but it could have great impact, for instance by helping visually impaired people better understand the content of images on the web. In this paper, a generative model is presented based on a deep recurrent architecture that combines recent advances in computer vision and machine translation and that can be used to generate natural sentences describing an image. The model is trained to maximize the likelihood of the target description sentence given the training image.

Problem Statement:

To create an end-to-end system for the Neural Image Captioning problem. We do this by using a CNN as an image 'encoder' by first pre-training it for an image classification task and using the last hidden layer as an input to the RNN decoder that generates sentences.

Approach:

The paper proposes a neural and probabilistic framework to generate descriptions from images. The model makes use of a recurrent neural network (RNN) which encodes the variable length input into a fixed dimensional vector and uses this representation to "decode" it to the desired output sentence. We then maximize the probability of the correct description given the image. To make the RNN more concrete we use A Long-Short Term Memory (LSTM) net. In this the convolutional neural network (CNN) will be used to inform the LSTM of the image content.

Ninad Parikh(A20427382)   Harsh Vora(A20445400)

Datasets:

MSCOCO Dataset

Flicker8k

Link: https://cocodataset.org/#download

Link: https://www.kaggle.com/hsankesara/flickr-image-dataset

References:

https://arxiv.org/pdf/1609.06647v1.pdf

https://towardsdatascience.com/automatic-image-captioning-with-cnn-rnn-aae3cd442d83

https://medium.com/analytics-vidhya/cnn-lstm-architecture-and-image-captioning-2351fc18e8d7

Team-Member Responsibilities:

Ninad/Harsh: Loading and Preprocessing the dataset and fit the data

Harsh/Ninad: Analyze the output and perform hyper-parameter tuning to optimize the model.