Ben Murray
Ventura
CS 474 - Deep Learning
20 Feb 2021

Final Project Proposal

Natural Language Processing and Computer Vision are both areas of study that I have been interested in for a long time. After analyzing several project ideas within both of these fields, I settled on image captioning as my source for a final project. In particular, I'm looking at this paper as a guide, although I will branch off and attempt to improve on their methods as I iterate.

The overall goal of this project, put simply, is to "write and train a deep recurrent neural network capable of autonomously captioning images". Measures of success for this project center around accurate sentence generation. This is usually measured via sentence perplexity, a standard method for evaluating language models. Another important metric is BLEU score, which stands for bilingual evaluation understudy. It's a common algorithm for evaluating the quality of text translated from one language to another. Between the BLEU score and average sentence perplexity, I should be able to adequately compare my model to many others published. If I can achieve the same or similar success as several other models found in the linked paper, I will consider the project successful, but of course my goal is to improve on their model.

To accomplish this, I will use a multi-modal recurrent neural network. It uses a typical deep CNN to perform feature extraction on the image, and then combines this with a recurrent neural network to estimate the probability distribution of different sentences produced. It is fairly complicated, and will be challenging to implement, thus providing a successful scope for this project. As far as data is concerned, there are several good datasets out there to train on. I'll begin initially with the Flickr30k dataset found here and continue finding additional data as needed. There are a lot of available resources out there because it is a common problem, so I should have no issue finding sufficient amounts of training data.

It's easy to imagine the benefits of automated neural network-based image captioning, from accessibility improvements to improved image classification, so there is a clear real-world purpose. It's a great project that blends the best of NLP and CV, and there is a wealth of academic research and data out there to guide me. I'm excited to begin!