Developing an Image Caption Generator Using a Convolutional Neural Network

Saeed, Zernab
Department of Computer Science
Middle Tennesee State University
Murfreesboro, TN 37128

Sindell, Daniel
Department of Computer Science
Middle Tennesee State University
Murfreesboro, TN 37128

Lopez, Alex
Department of Computer Science
Middle Tennesee State University
Murfreesboro, TN 37128

Karki, Pratab
Department of Computer Science
Middle Tennesee State University
Murfreesboro, TN 37128

Kwarteng, Michael
Department of Computer Science
Middle Tennesee State University
Murfreesboro, TN 37128

Abstract **– The integration of Neural Networks in everyday technology is growing exponentially. Computer Scientists are constantly pushing the limits of creativity and researching, developing and integrating application of Neural Network Models into the real world. With the rapid shift into technology, it is now our turn to put our knowledge to the test. Our objective is to utilize the already available research and leverage the tools we use in class to build a deep learning Neural Network that generates a caption after analyzing an image. This is known as image captioning and is leading to have infinite uses in every field. Image Captioning is still new and challenging; as students we hope to utilize already available data sets to build a convolutional neural network that analyzes images and displays a predicted caption. We hope to test our model and train it to at least 80% accuracy.**

I. INTRODUCTION

Neural Networks are computer systems that model the activity of the human brain. They are composed of interconnected parallel processing elements that work together to solve a problem and make connections by "learning" instead of memorizing. They are trained to recognize new data sets by using what they have already learned to make connections, much like the human brain. The advancement of technology and the expansion of the field of Artificial Intelligence has increased an interest in the area of Image Captioning. This idea of processing and understanding a scene in the image using a deep learning neural network and outputting a caption has many uses, for example in the realization of human-computer interaction.

For the purpose of this project, our team was interested in leveraging the tools we learned in Dr. Phillips' Neural Networks class to build a simple Image Captioning Network. This paper documents our process of building a deep learning neural network by leveraging tools we learned in class, along with research and methods we explored to test our network. We will be documenting our successes and failures to gain a better understanding of deep learning.

Our group started off with researching the core foundations we needed in order to understand the working of an Image Caption Generator. We came across with unfamiliar terminology and concepts that were necessary in building a deep learning network. The application of Image Captioning is extensive, and it combines the knowledge of computer vision and natural language processing.

To start off, we initially planned to validate inputs of a raw data set and train those before feeding it into our network. However, due to time constraints and inability to grasp our head around the complexity of loading raw data, we decided to use a pretrained data set provide by Google. However, in OLA 6, we were introduced to a pretrained data set of images which we found useful and debated on using.

We hope to make this paper as simple as possible so that others wishing to explore more about Image Captioning can utilize this as a resource and use it as a tutorial, while others can extend the length and complexity of our model to increases its efficacy.

II. BACKGROUND

A. Introduction to Deep Learning and Convolutional Neural Network

For the reader's convenience and to document our understanding, we will provide a brief overview of Deep Learning and Convolutional Neural Networks that are the foundation of our Neural Network Model.

B. Tools

C. Others

III. OUR MODEL

A. Data used and overview

B. Convolutional Network and Layers

C. Generating Captions

IV. RESULTS

A. Methods and tools used to test model

V. DISCUSSION

VI. CONCLUSION