ABSTRACT

Generally when humans see an image, they are able to pick out individual elements of the image such as objects, colors, people or points of interests in a simple and elegant manner. This, however, is very complicated and difficult for a computer to do. The computer will have to analyze each individual pixel, with the help of various image preprocessing methods, extract features that are usable in the final output. Even after that, the model may lack the vocabulary required to caption the image with appropriate words. This project aims to create a sophisticated image caption model capable of generating captions for provided images by analyzing the various elements of the image by making use of Convolutional Neural Networks (CNNs) for image analysis and using a highly trained text engine to predict the captions. This project can be utilized further in many different fields provided the project has a dataset sufficient enough to train a model and develop a robust enough vocabulary of words to assign to each image appropriately. This project could see utilization of the model in the medical industry, traffic safety, archiving, social information networks, and even cyber security. The scope of this project aims to showcase a general fully implemented model to understand and utilize image caption generation to show the potential of the model and to further improve upon existing models through optimization and larger datasets. This project will employ the concepts and practices of image preprocessing, image analysis, and text analysis to deliver a full implementation of Image Caption Generation.

Keywords: Image, Caption, Generation, Convolutional Neural Networks (CNNs), text analysis