NATURAL LANGUAGE PROCESSING

PROJECT PROPOSAL

Team Members:

Ashwin Muthuraman G46745441 ashwin.muthuraman@gwu.edu

Bhoomika Nanjaraja G41609848 bhoomika.nanjaraja@gwu.edu

Bhanu Sai Praneeth Sarva G46159306 bhanusaipraneeth.sarva@gwu.edu

Project Title: Image Caption Generator

Introduction: This project is designed to develop an image caption generator leveraging the VGG16 convolutional neural network for feature extraction from images, paired with Long Short-Term Memory (LSTM) networks for generating natural language captions. We will use TensorFlow for implementing and fine-tuning the model, integrating tokenizers and embeddings to improve caption quality. To assess the effectiveness of the generated captions, we will use the BLEU score from the NLTK corpus toolkit, which will help us measure the grammatical and semantic accuracy of the captions in comparison to standard references. The final model will be integrated into a user-friendly Streamlit application, allowing users to upload images and receive captions automatically.

Technology Stack:

- **Deep Learning Framework**: TensorFlow
- **Model**: Fine-tuned VGG16 coupled with LSTM
- **Frontend**: Streamlit
- **Natural Language Processing**: Tokenizers and Embeddings for text processing
- **Evaluation**: NLTK corpus for BLEU score assessments

Dataset: The primary dataset for this project will be the Flickr8K dataset, which provides a wide variety of images suitable for training our model. It can be accessed here: https://www.kaggle.com/datasets/adityajn105/flickr8k

Alternate Datasets: Depending on available computational resources, we may also consider larger datasets for more extensive training. Potential alternatives include:

Flickr30K Dataset: https://www.kaggle.com/datasets/eeshawn/flickr30k

MSCOCO Dataset: https://www.kaggle.com/datasets/clkmuhammed/microsoft-coco-2017-common-objects-in-context