**COMS4995 Deep Learning for Computer Vision**

**Project Proposal**

**Project Title**

SeasonScope: A Comprehensive System for Season Recognition and Captioning from Outdoor Images

**Participants**

1: Jennifer Duan (jd3794)

2: Haowen Xu (hx2364)

3: Shiying Chen (sc5299)

**Objectives**

The primary goal of this project is to develop a comprehensive system, "SeasonScope," capable of identifying seasons (spring, summer, autumn, and winter) and generating contextually accurate captions from outdoor images. We aim to achieve an accuracy rate of above 80% on our test dataset for season identification, and the captions should be descriptive and relevant.

This project, to the best of our knowledge, brings a new perspective to the area of image-based season classification by combining a diverse, globally-sourced dataset with advanced deep learning techniques.

**Division of Tasks**

1. Data collection: Haowen Xu, Jennifer Duan
2. Classification: Haowen Xu, Shiying Chen,
3. Image Captioning: Jennifer Duan, Shiying Chen

**Plan for Completion**

1. Data Collection and Preprocessing
   1. Classification

We plan to curate our dataset by collecting outdoor images from various open-source platforms like Flickr and Unsplash, ensuring a diverse representation of all four seasons. The collected images will be manually labeled with the correct season. For preprocessing, we'll resize the images to a standard size, normalize the pixel values, and split the data into training, validation, and test sets.

* 1. Image Captioning

We plan to use the existing Flickr-8k Dataset for training and validation, and curate our own dataset for testing by collecting outdoor images from various open-source platforms like Flickr and Unsplash.

1. Deep Learning Framework

We will use the PyTorch framework for this project. It provides efficient GPU utilization and has extensive support for both image classification (Convolutional Neural Networks) and sequence generation (Long Short-Term Memory networks), which are crucial for the captioning part of our project.

1. Network Architecture

The network architecture for our project, "SeasonScope," will be a fusion of a Convolutional Neural Network (CNN) for image feature extraction and a Long Short-Term Memory (LSTM) network for caption generation.

* 1. Classification

The CNN component will consist of an input layer that holds raw pixel values, followed by convolutional layers to extract features from the images, and pooling layers to reduce the dimensions and introduce scale and orientation invariance. Fully connected layers will then perform high-level reasoning, with the final layer using an activation function like softmax to classify the season depicted in the image. We aim to use a pre-trained model like ResNet or AlexNet for transfer learning, fine-tuning it to our specific task. We will also experiment with data augmentation techniques to enhance the diversity of our training data and prevent overfitting.

* 1. Image Captioning

The features extracted by the CNN will be fed into the LSTM, which will generate contextually accurate captions based on the identified season. The CNN will also be fine-tuned for season classification, enhancing the accuracy of both the classification and captioning tasks.

1. Originality and Creativity

While image classification is a well-explored area, focusing on season identification brings a unique challenge due to the inherent variability of the seasons across different locations and times. By curating our own dataset and optimizing a deep learning model for this specific task, we aim to contribute a new perspective to the field.

1. Expectations and Results

We aim to have a functioning model by the end of the project timeline. Our goal is to achieve an accuracy rate above 80% on the test dataset for classification and high similarity between generated caption and label caption. However, we acknowledge the challenges associated with this task and will consider the project successful if we can demonstrate clear progress and learning, even if the final accuracy is below our initial target.

**Conclusion**

This project aims to provide end-to-end fluency in deep learning, from data collection and preprocessing to model development, training, and testing. It offers a unique challenge in the field of image classification and has the potential for practical applications in various domains.