SDAIA Academy T5C04 Bootcamps: Data Science

Deep Learning Module Project Proposal

Image Captioning: Generate A Description of A Video

Date: November 17, 2021

Prepared for: Nawras Boufaied

Prepared by: Nada Rambu

Objective

The idea of this project is to create an abundant source of video data with the help of deep learning techniques. This data source may assist researchers examining visual media and quickly obtaining a description of their content, saving their time and effort.

The goal of this project is to generate a textual description of a given video. The description will include the activity displayed in the scene along with the emotion expressed by the people in the scene. To do this, a neural network will be implemented that will learns to recognize human activities and emotions from images, then generating a text accordingly.

Data Description

The Flickr30k dataset has become a standard benchmark for sentence-based image description (Find the dataset on the following link: https://www.kaggle.com/hsankesara/flickr-image-dataset). This dataset will be used to generate description of the video frames along with another dataset for the emotion recognition. The other dataset contains 48x48 pixel gray scale images that are categorized based on the emotion shown in the facial expressions, which are: happiness, neutral, sadness, anger, surprise, disgust, and fear (Find the dataset on the following link: https://www.kaggle.com/ananthu017/emotion-detection-fer)

Tools

The following tools will be used to carry out the project:

1. **Keras** To implement the neural networks.
2. **NLTK** toolkit to perform common NLP tasks.
3. **Pandas** library will be used to create data frames for easier data manipulation.
4. **Matplotlib** library will be used to visualize and discuss the results of the analysis.