CPSC 8430 Deep Learning – Homework 2 Report

~ Sarthak Nikhal (C12796856)

GitHub repo link - <https://github.com/SarthakNikhal/>

1. Problem Statement

Create a model that generates video captions for an input video based on sequence to sequence architecture. The input is a video. The output will be a sequence of text that describes the video in short. The above is achieved by LSTM network which is based on Recurrent Neural Network(RNN). The stack used is

• Python • CUDA • torch • numpy • scipy • pickle • pandas

1. Data set

The dataset provided as part of the homework is used to train the model. The dataset contains 1450 videos for training & 100 for the purpose of testing. Training and testing label files contain the labels for the model.

1. Method

The model in this homework is sequence to sequence model that takes video frames as input and generates sequence of test. The preprocessing part consists of padding and masking the sequence of text. Masking is used to hide or flag the words that should not be detected. This process improves performance. The data folder contains training and testing data.

Tokenization;

<pad> - padded word

<eos> - end of sentence

<sos> - start of sentence

<unk> word not in dictionary (unknown)

Information about text;

Vocab size – 6496

Average length of captions – 8.49

1. Results

Bleu Score – 65.4

After 100 epochs – 66.7