Image Captioning using Deep Learning

Abhisek Konar (abhisek.konar@mail.mcgill.ca)
Aashima Singh (aashima.singh@mail.mcgill.ca)

Problem Introduction

- Image captioning acts as a bridge between NLP and Computer Vision
- Image captioning involves generation of meaningful descriptions given an image.
- In this project, end-to-end neural image captioning systems are explored.

Model Description

- CNN model for encoding images:

  A pretrained VGG19 model trained on IMAGENET followed by a dense layer to produce a 128 dimensional representation of each image.

- LSTM model for embedding words into vector representations:

  One-hot encoding is used where a word is represented a vector with total size of vocabulary

- LSTM model for caption generator: acts as a ‘decoder’ and generates the target sentences

Model Architecture:

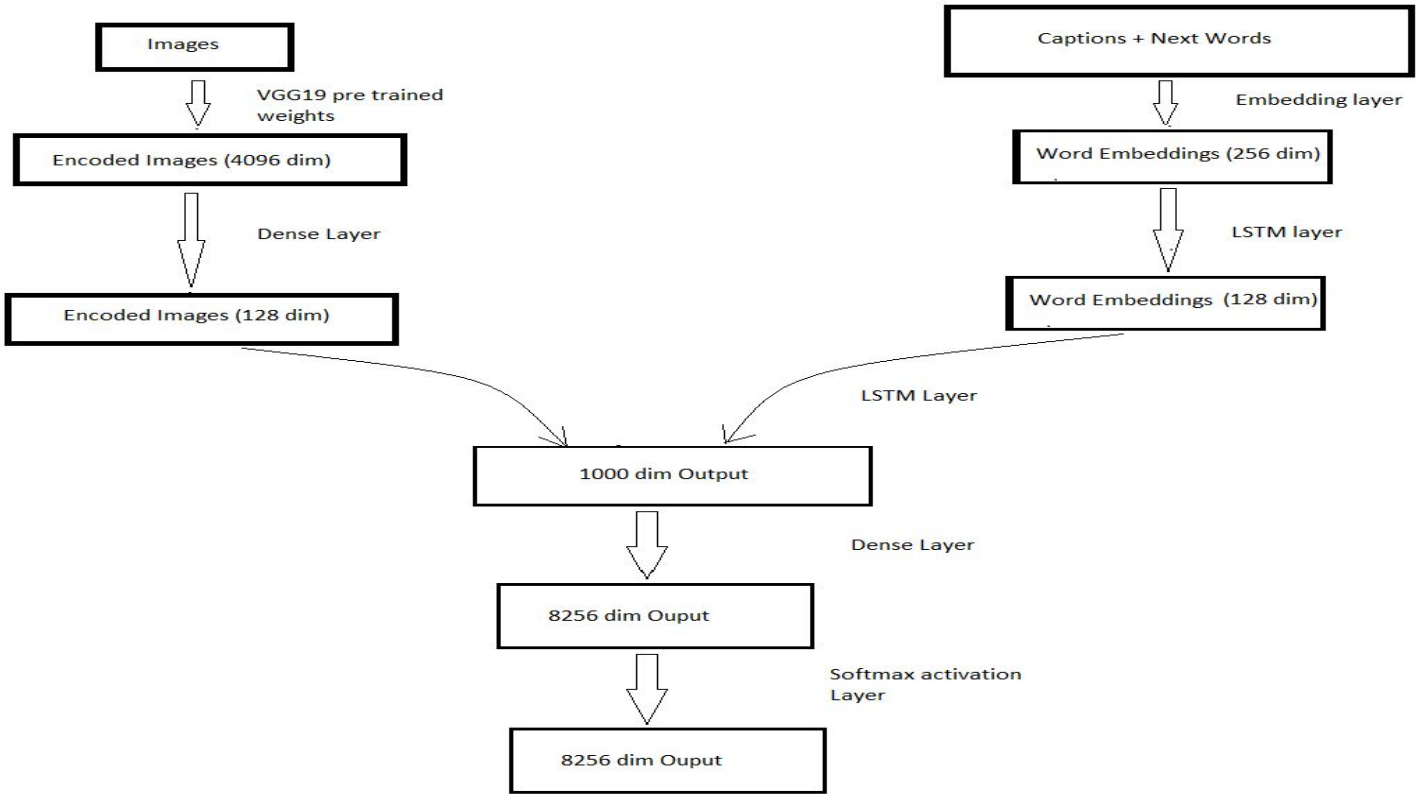


Image Representation

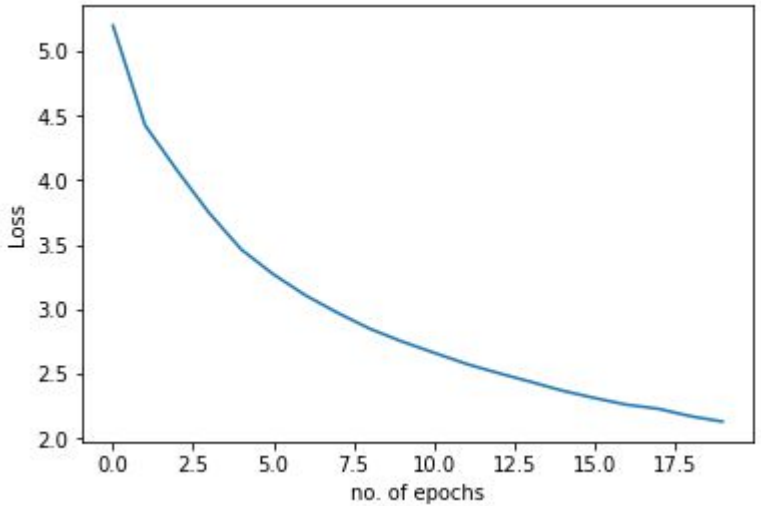
- The Image model starts by encoding the input using the weights from a VGG19 trained on IMAGENET.
- For each image, its representation from the penultimate layer is extracted.
- The resultant is a 4096 dimensional vector as features for each image representation.
- This is then passed through a Dense layer and then used as an input to the decoder LSTM that generates sentences.
- Both the image and the word embedding are mapped to the same space.

Language Model and LSTM decoder

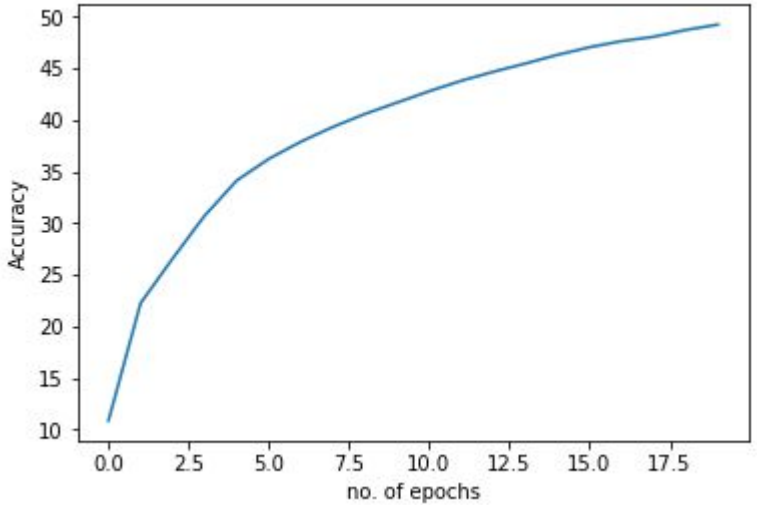
- In the language model,each caption is converted to a vector of size 40 . And a 1 hot encoding is created for the next word for every word in the captions.
- This is then passed through a sequence of Embedding layer and LSTM and a 128-dim representation of each word is obtained.
- This, along with the 128-dim image representations, is passed to the LSTM sentence generator.
- Predicts the next word conditioned on the current image and the previous words defined by

  $P(S(t)|I, S(0),...S(t-1)$

Preliminary results



Loss function



Validation Accuracy

Preliminary Results

A man outside on a street

A brown brown dog running through a pool

A man on a cliff