

Image Captioning using Neural Networks

Name:

Aditya Gurjar

Pritish Uplavikar

Literature Review

- [1] A. Karpathy and L. Fei-Fei. Deep Visual-Semantic Alignments for Generating Image Descriptions. *CVPR*, 2015
 - Mapped words in the image description to regions in the image using R-CNNs
- [2] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan. Show and Tell: A Neural Image Caption Generator. *arXiv:1411.4555*, 2015
 - Fed encoded image and word vectors as input to a bidirectional LSTM

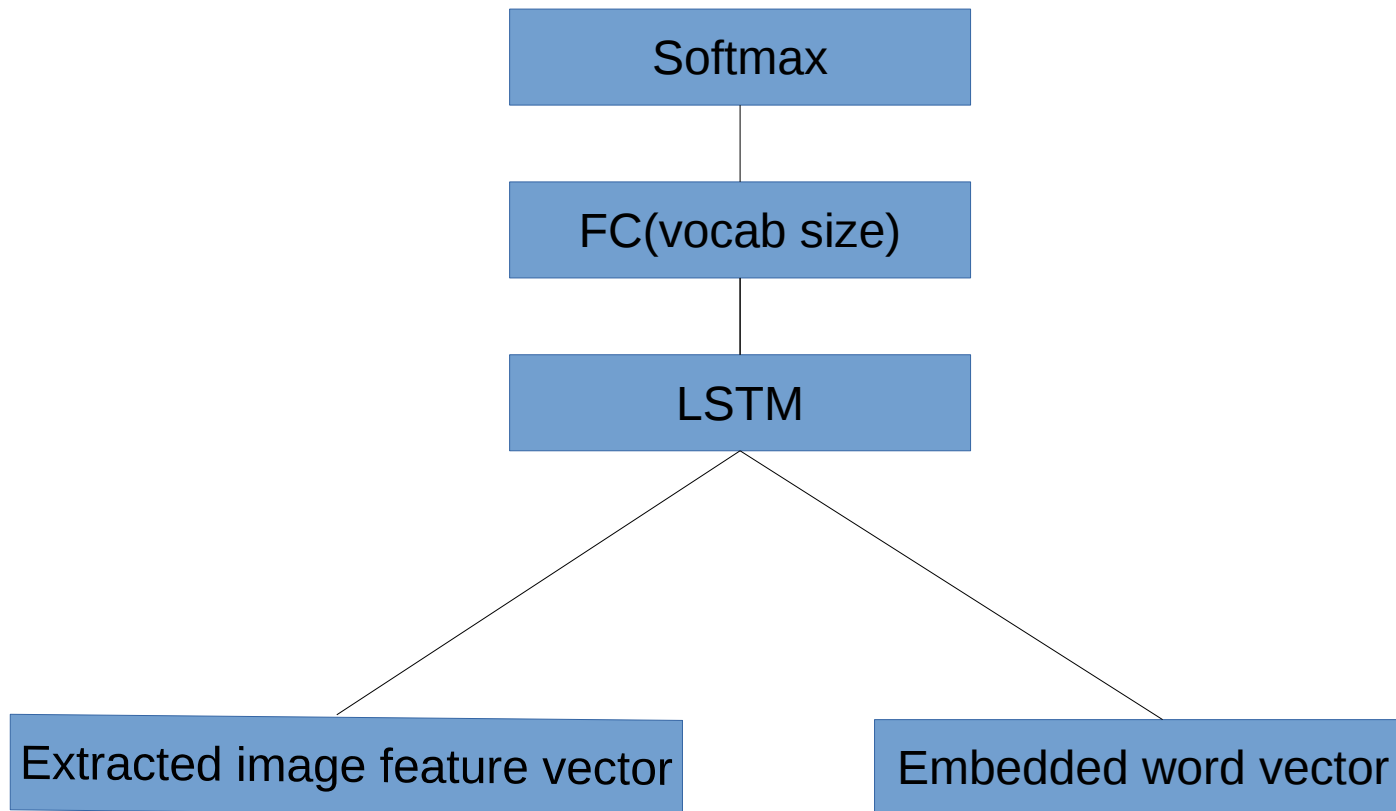
Dataset

- Flickr8k
 - 8000 images
 - 5 human generated captions per image

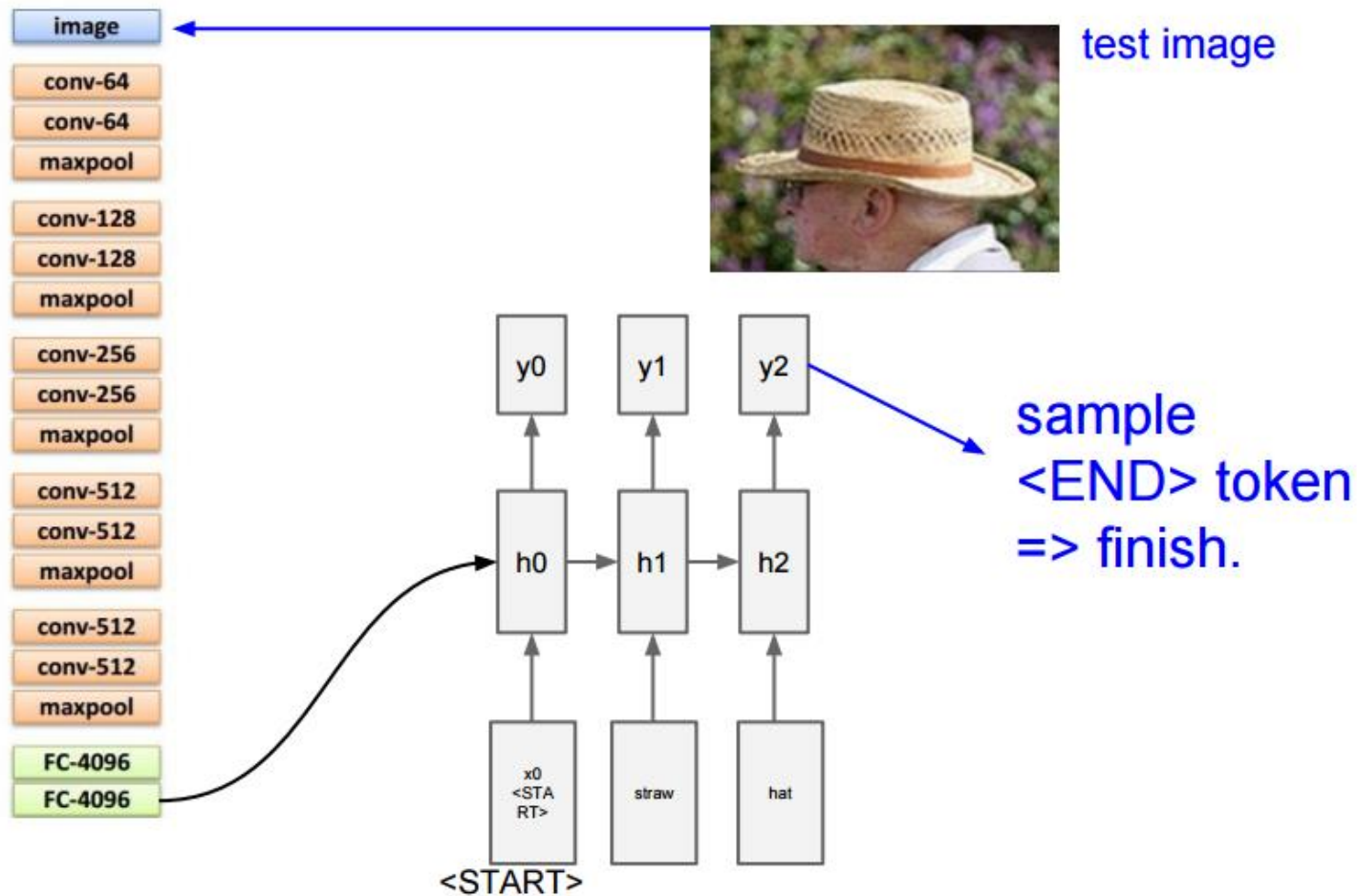
Algorithmic Overview

- Pre-trained CNN model on ImageNet: InceptionV3
 - Extract all the image features
- Vectorizing words in training set captions
- Output of CNN along with word embeddings fed into a RNN for generating captions

Model



How it works?



Results

- Results
- Link

Thank You!

Any Questions?