Mini Project Report

Group ID: 34
Ajay Chhajed - IMT2019006
Vismaya Solanki - IMT2019095
Yeshwant D - IMT2019098

Task 1: Design a CNN-LSTM system that can perform image captioning and test it against Flickr8k dataset, to achieve a good BLEU score on it.

Introduction: We have implemented neural network-based caption generator. We used CNN and encoder and LSTM-based RNN as decoder.

CNN Encoder:

Convolutional layer – This uses three things

1) Input data
2) Feature detector
3) Feature map

The output of this layer is called feature map or activation map.

Pooling layer – It is like convolutional layer, except for the fact that this does not have any weights. There is max pooling and average pooling.

Fully connected layer – This layer performs the task of classification based on features extracted from the previous layers. The output is the probability for every input from 0 to 1 of belonging to a particular class.

LSTM decoder:

LSTM helps RNN to remember the inputs for long duration of time. This is called gated cell. We have 3 gates here

1. Input
2. Forget
3. Output

Here decoder needs to create the word-by-word caption. It will take input as the encoded feature vector from CNN. We use categorical cross-entropy.

Word embeddings:

We must convert the captions for training images into embedded captions. We used glove embeddings.

Beam search:

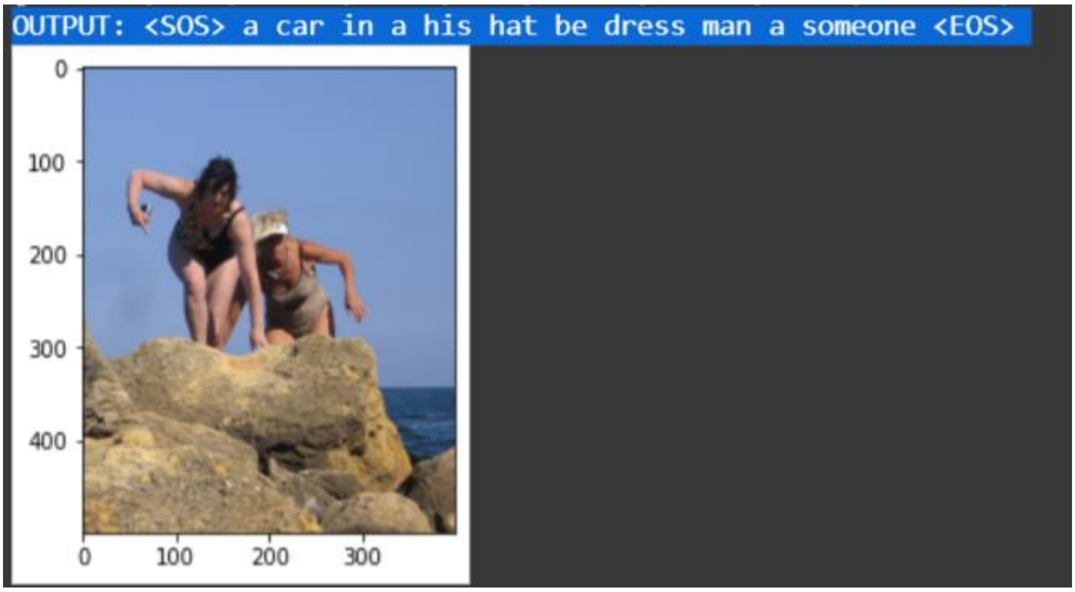
The algorithm is a best-first search algorithm which iteratively considers the set of the k best sentences up to time t as candidates to generate sentences of size t + 1, and keep only the resulting best k of them, because this better approximates the probability of getting the global maximum.

BLEU scores:

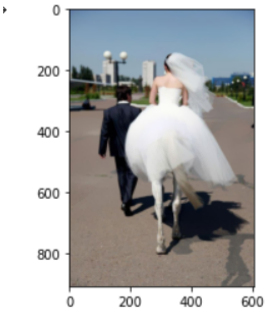
Full form of BLEU is bilingual evaluation understudy. BlEU's output will be a number between 0 and 1. The score indicates how similar is the given text with respect to the reference text with values given closer to 1 representing more similar texts.A perfect score is not possible in practice as a translation would have to match the reference exactly.

Our Scores – 0.43 (approx.)

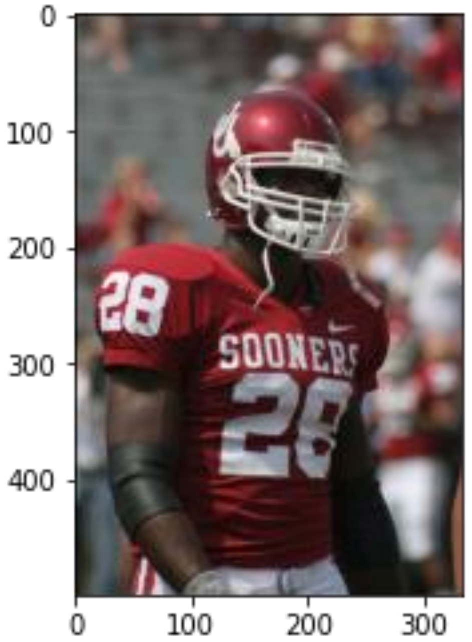
Test images:



```
OUTPUT: <SOS> a car in a his hat be dress man a someone <EOS>
```



```
a little girl in a white dress be run along a dirt road
```



```
Predicted Caption: A football player 25 the Sooner 25 .
```

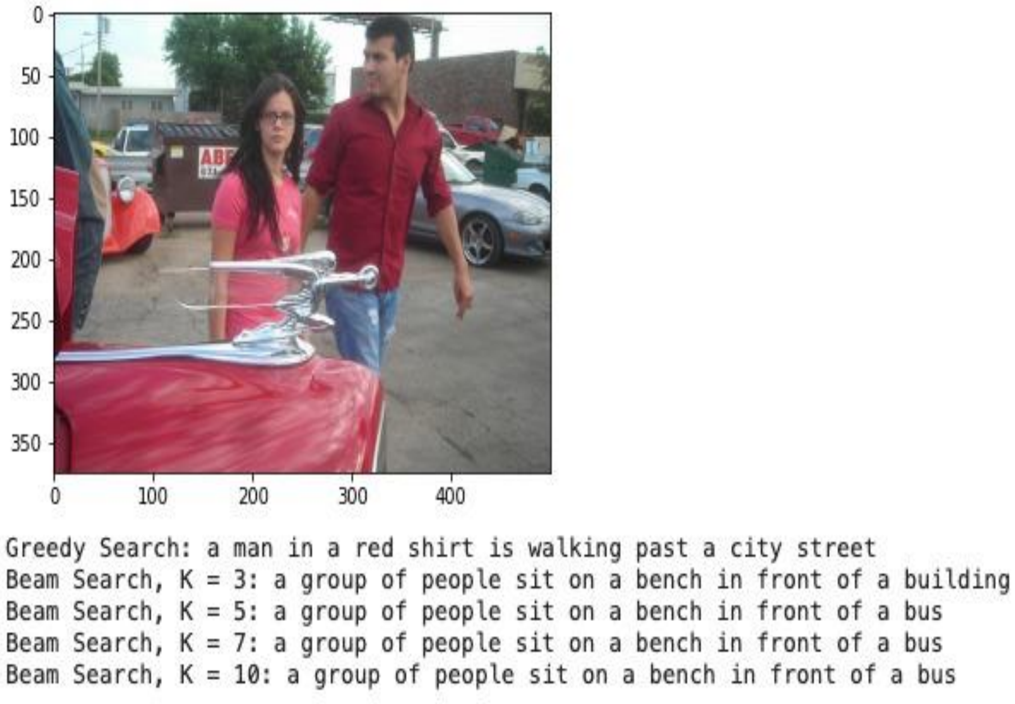
Task 2:

Word embeddings propagate bias to supervised downstream applications, resulting in biased decisions that mirror the data's statistical patterns. These downstream applications handle activities including information retrieval, text production, machine translation, text summarization, and online search, as well as consequential decision-making during resume screening, university admissions automation.

As you can see in the below image, we predicted that “A man in a red shirt is walking past a city street”. We are biased towards the man in this case as both man and woman are present in the scene, still, our model predicted for man. This shows the language-biased case.

Word embeddings learn to correlate concepts with co-occurring attributes when words representing concepts appear frequently with those traits. As man are more frequently detected on outside streets and, in comparison to males, women appear nearer to terms like family and arts, whereas men look nearer to keywords like profession, science, and technology.

According to our findings, stereotypical linkages exist between gender, race, age, and their intersections also comes under language-biased. Minorities suffer when stereotypical associations extend to downstream apps that deliver information on the internet or make significant judgments about people.



```
Greedy Search: a man in a red shirt is walking past a city street
Beam Search, K = 3: a group of people sit on a bench in front of a building
Beam Search, K = 5: a group of people sit on a bench in front of a bus
Beam Search, K = 7: a group of people sit on a bench in front of a bus
Beam Search, K = 10: a group of people sit on a bench in front of a bus
```

References:

https://machinelearningmastery.com/calculate-bleu-score-for-text-python/

https://towardsdatascience.com/image-captioning-in-deep-learning-9cd23fb4d8d2

https://towardsdatascience.com/a-simple-guide-to-the-versions-of-the-inception-network-7fc52b863202

https://machinelearningmastery.com/beam-search-decoder-natural-language-processing/