Automated Image Caption Generator with CNN and LSTM

Aryan Shrestha

Introduction

- Image Caption Generator requires both methods from computer vision to understand the content of the image and a language model from the field of natural language processing to turn the understanding of the image into words in the right order

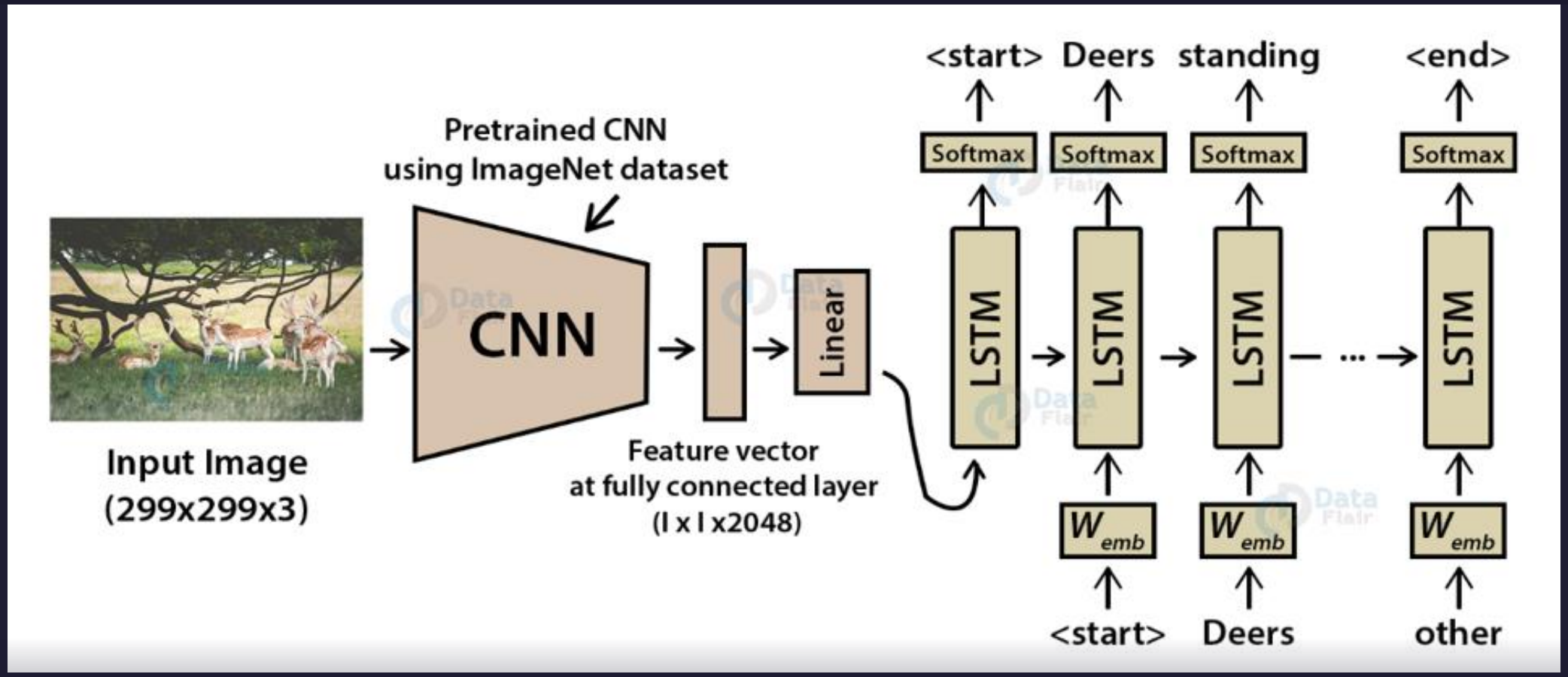
- The model used here consists of two main parts: CNN and LSTM

Introduction

- Convolutional Neural networks are specialized deep neural networks which can process the data that has input shape like a 2D matrix. Images are easily represented as a 2D matrix. CNN is very useful in working with images.

- LSTM stands for Long short term memory. They are a type of RNN which is well suited for sequence prediction problems. LSTM can carry out relevant information throughout the processing of inputs and with a forget gate, it discards non-relevant information.

Image Caption Generator Model



Dataset

8000

Total

6000

Train

7577

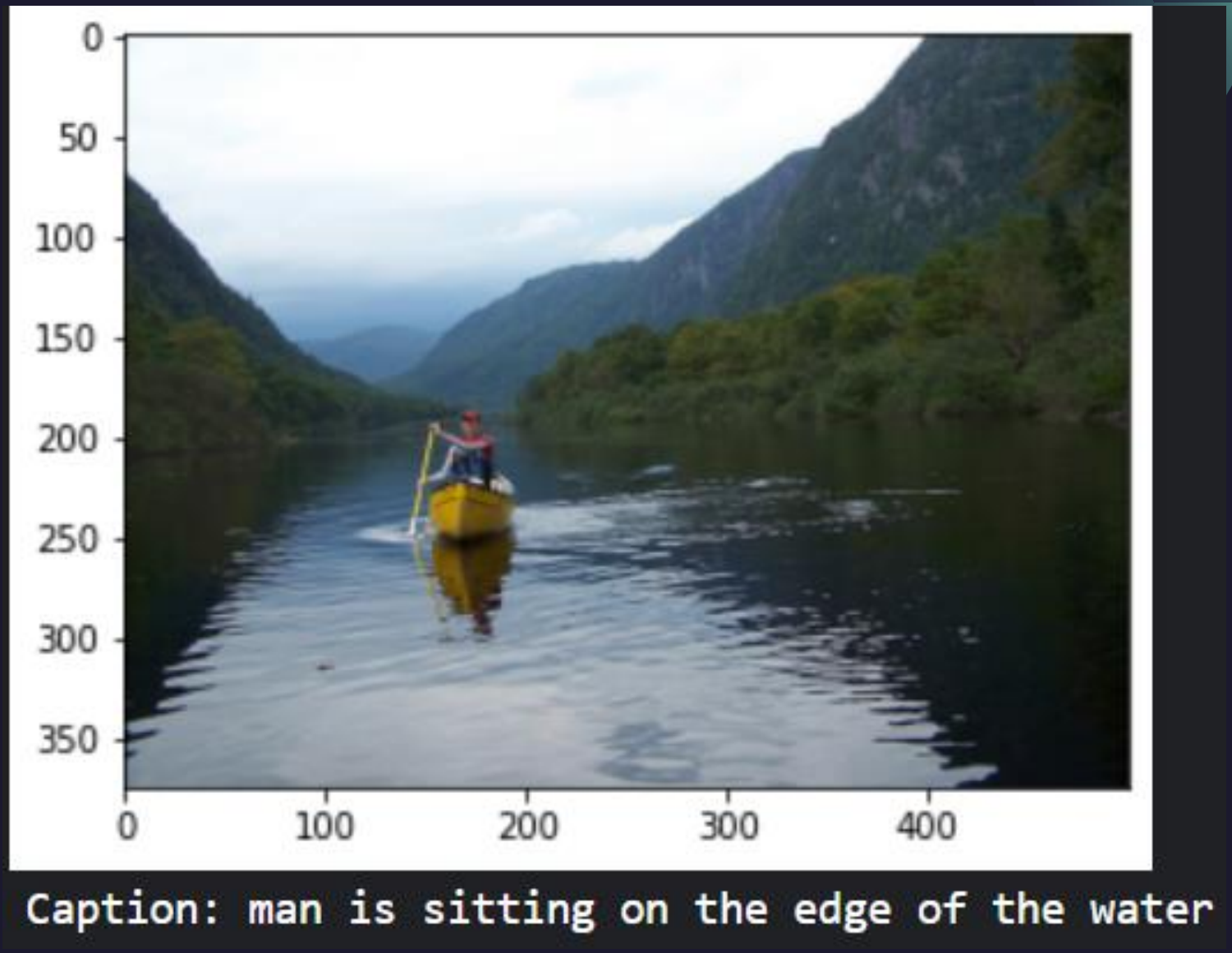
Vocabulary size

Evaluation

- BLUE stands for Bilingual Evaluation Understudy. It compares the machine-generated captions to one or many human-written captions and computes a similarity score based on N-gram precision and a penalty for too-short system translations.

- Beam Search expands all possible next steps and keeps the k most likely, where k is the beam size. At each step, all the successors of all k states are generated. If any one is a goal, the algorithm halts. Otherwise, it selects the k best successors from the complete list and repeats.

Sample Output



Thank You