IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 8, Issue 10, October 2020

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

|| Volume 8, Issue 10, October 2020 ||

Neural Image Caption Generation Using Deep Learning

Binshad P.B, Priyanga K.K

Department of Computer Science, Christ College(Autonomous), Irinjalakuda, Kerala, India

Assistant Professor, Department of Computer Science, Christ College(Autonomous), Irinjalakuda, Kerala, India

ABSTRACT: Image captioning, which aims to automatically generate a sentence description for an image, has attracted much research attention in cognitive computing. The task is rather challenging, since it requires cognitively combining the techniques from both computer vision and natural language processing domains. Existing CNN-RNN framework-based methods suffer from some problems. The model uses the pre-trained inception-v3 image embedding model stacked with Gated Recurrent Unit (GRU)layer. The proposed model has been trained and validated with the COCO dataset. I assign different weights to the words according to the correlation between wordsand images during the training phase. I additionally maximize the consensus score between thecaptions generated by the captioning model and the reference information from the neighbouring images of the target image, which can reduce the misrecognition problem.

KEYWORDS: Deep Learning, Convolutional Neural Network, Gated Recurrent Unit

I. INTRODUCTION

Image captioning is one of the most challenging tasks connecting the recent research on computer vision and in the field of natural language processing (NLP). Image captioning aims to produce a sentence that describes a given image. It is already had major impacts in various field such as further image analysis (e.g. image search),video tracking, cross-view retrieval, sentiment analysis, childhood education and helping blind people to understand images. It also has the potential to give positive changes in many different areas including human-computer interaction, security,and others.In a single image, it usually consists of several objects that each object has attributes,position, and how the object is related to another. These are described by the caption generated.

This paper is inspired by the concept of encoder-decoder Framework. It is one of the state-of-the-arts in the area of machine translation that has been proven to produce good resultsfor generating proper sentences for image captioning problem. A convolutional neural network(CNN) was used as an encoder that functions as a feature extraction from the given image. The architecture of CNN used here is inception-v3 that has a good performance in the case of image annotation. It also utilize the Gated Recurrent Unit(GRU) as a decoder that serves to produce sentences with input extraction of image features performed by CNN. Gated Recurrent Unitsare used to overcome some of the problems that exist in the Recurrent Neural Network (RNN),such as vanishing gradients. GRU has been proven to be able to train models quickly and shownbetter results compared to LSTM.

II. RELATED WORK

The early efforts on image captioning mainly adopt the template-based methods, which requirerecognizing the various elements, such as objects as well as their attributes and relationships inthe first phase. These elements are then organized into sentences based on either templates orpre-defined language models, which normally end up with rigid and limited descriptions. Despite achieving the state-of-the-art performance, existing CNN-RNN framework-based methods suffer from two main problems,

- Information inadequateness problem : These methods treat different words of a caption equally, which makes distinguishing the important parts of the caption difficult.
- Misrecognition problem : The main subjects or scenes might be misrecognized using the traditional methods.

III. PROPOSED SYSTEM

I evaluate the proposed GRU model on the MS COCO dataset. GRU is commonly used in NLP, especially in the task of machine translation. We take a machine translation examplebecause of encoder-decoder concept that successfully generating very well sentence and we will implement that concept of the model for this research. GRU can overcome vanishing gradient problems and can remember certain relevant features for a long series of sequences.
In the task of image captioning, the concept of encoderdecoderis widely used. The encoder is replaced by an image embedding model such as CNN.The main contributions of thesystem,

- I propose to use the training images as references and design a novel model, named Gated Recurrent Unit(GRU)with attention Mechanism for image captioning.

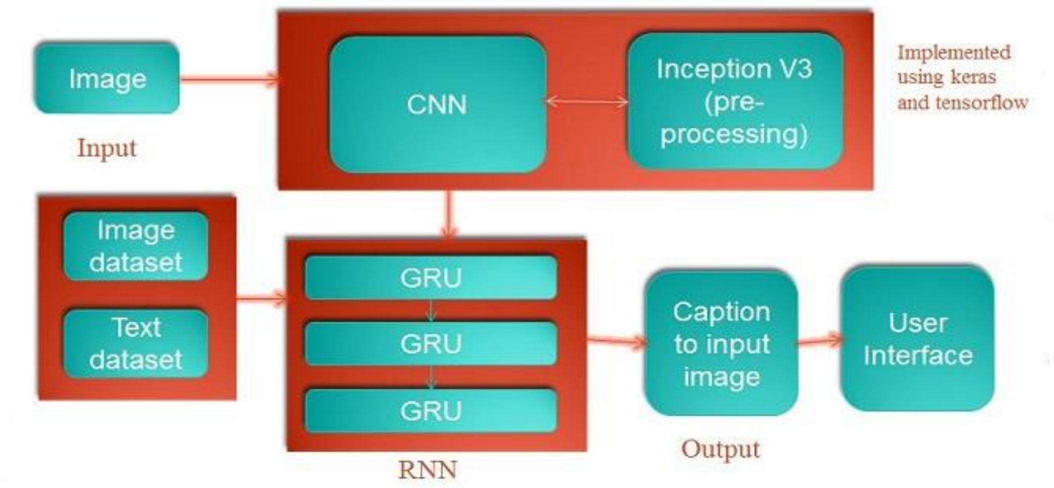

Figure 1:Working Diagram

IV. METHOD

There are 4 main processes in the system designed here namely, preprocessing (image and caption), image feature extraction, caption generation, and model scoring.

The modularity criteria are:

A. DATASET COLLECTION

Dataset Collected From COCO Dataset which gives free and lively images.

B.PREPROCESSING

Preprocessing on images is done to convert image objects into RGB arrays. Then the array is resized to (299, 299, 3).Preprocessing in the caption is performed to make sentences that were previously in the form of the word into a sequence of tokens based on a unique word index in the dictionary.At the training phase, the model has two inputs. The first input is an image feeding into the pre-trained Inception-v3 model with the removed output layer and will outputting extracted images features. The second input is a description that has been done by preprocessing so that it becomes a sequence index of tokens.

C. WEIGHTED TRAINING

The inception V3 model is employed as the encoderto extract CNN features of the targetimage and the training images. During the weighted training stage, the weight attached to each word in the training captions is calculated firstly.Then, the GRU model is trained using the weighted words and CNN features of the training images under the proposedweighted likelihood objective.

In the generation stage, the trained GRU plays as a decoder role, which takes the CNN features of the target image as input and generates the description words one by one.

D. GRU-BASED SENTENCE GENERATION

Gated recurrent units are a gating mechanism in recurrent neural networks that aims toovercome the vanishing gradient problem. GRU is similar to LSTM with the forget gate and update gate, but it has fewer parameters than LSTM without output gate. GRU performance on certain tasks on sequence data (i.e. text, sound) was found to be similar to LSTM. GRU uses both update and reset gates to solve the vanishing gradient problem of a standard RNN. Both of these vectors determine the information that will be continued or omitted in each GRU unit. The information flow in a GRU unit is shown by Fig 1. The update gate aims to determine how much information from the previous units/timestamp that must be forwarded.The reset gate is used by the model to determine how much in-formation from the past units/timestamp to forget.Current memory content used to store the relevant information from the previous units/timestamp.

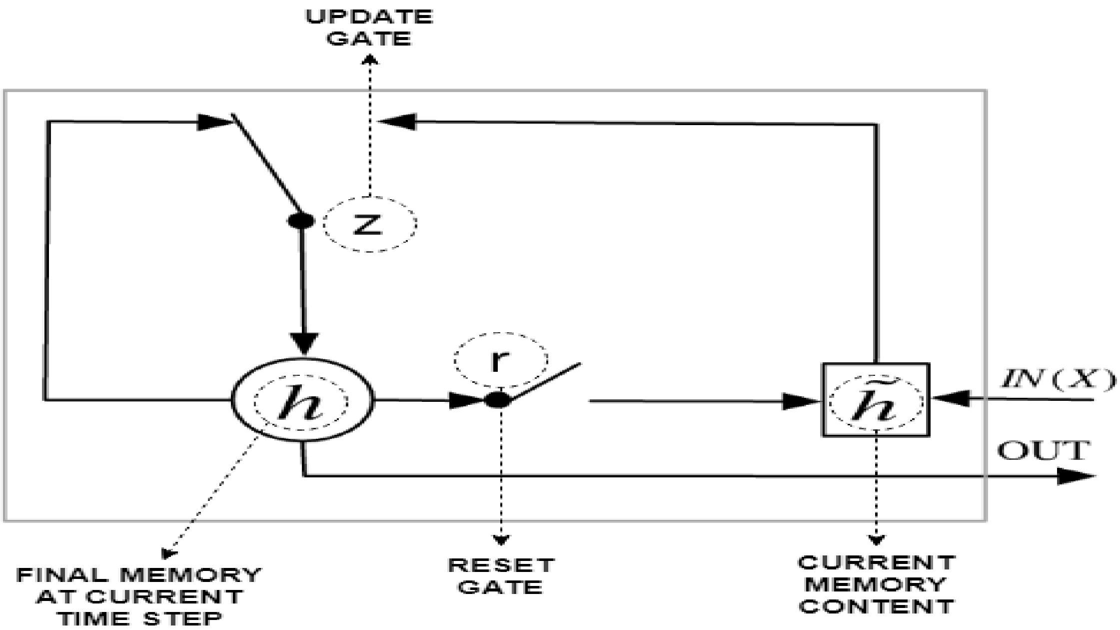

Figure 2:The information flow in a GRU unit

V. IMPLEMENTATION

To demonstrate the presented method, We use the COCO dataset for training the model. The authors implemented their approach using Python programming language(widely used for deep learning). we use random images for testing the model.The following are the summary of results

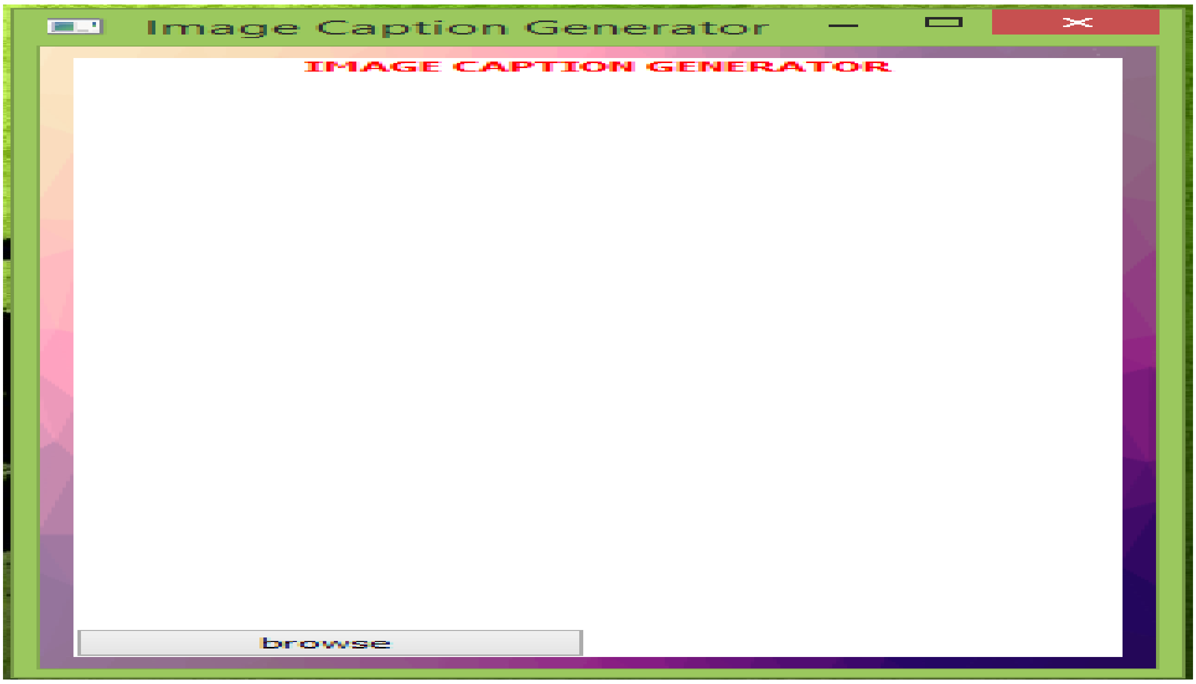

Figure 3:Main Interface

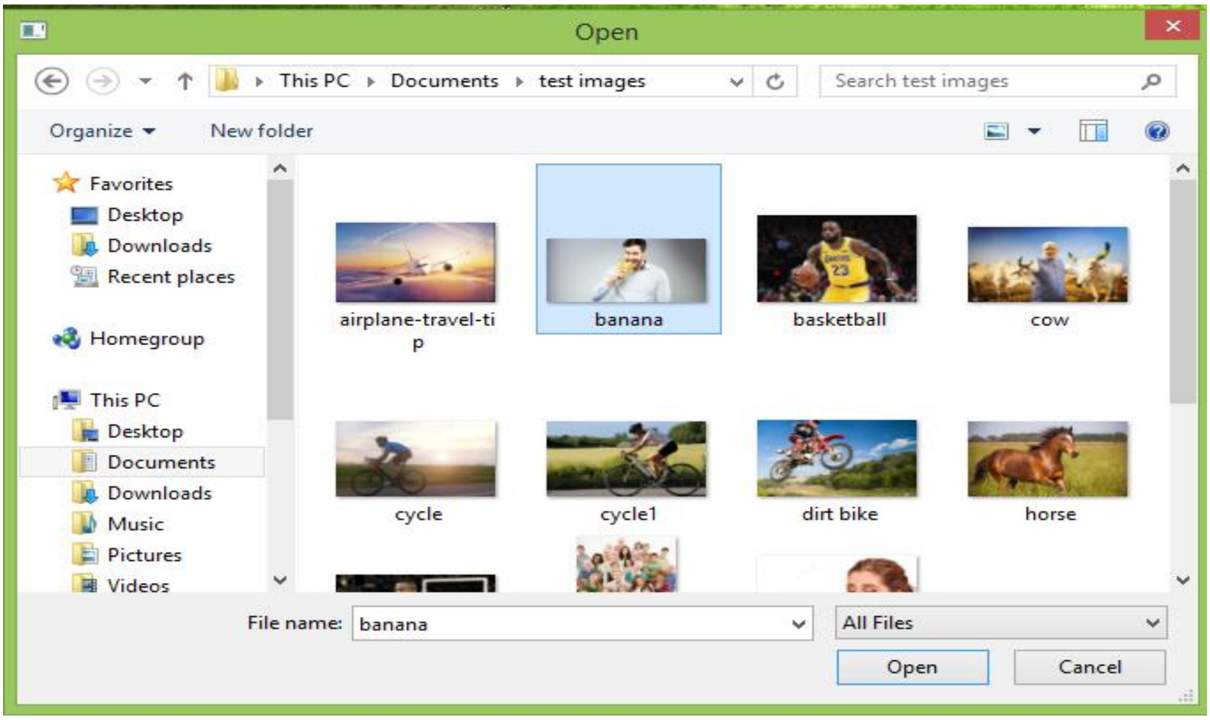

Figure 4: Select a random Image

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

|| Volume 8, Issue 10, October 2020 ||

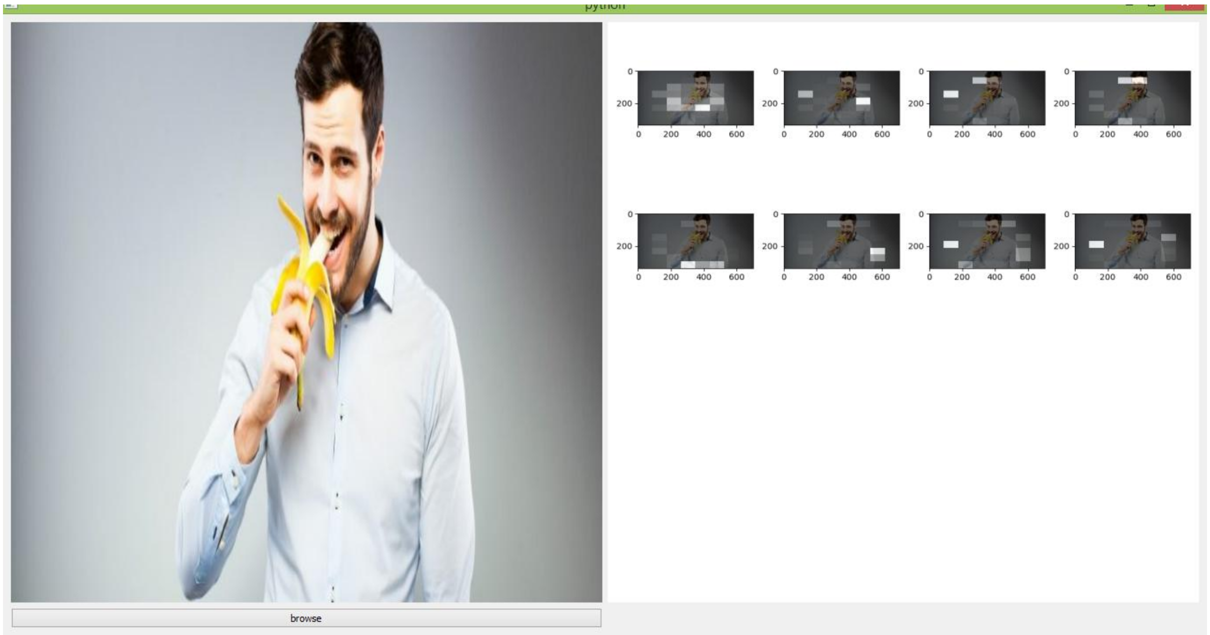

Figure 5: Plotting the objects

Figure 6:Sentence Generation

Figure 7:Another Result of sentence generation

VI. CONCLUSION

we have presented a reference-based GRU model, where the central idea is to use the trainingimages as references to improve the quality of generated captions. In the training phase, the words are weighted in terms of their relevance to the image, including the overall occurrences, part of speech and corresponding synonyms, which drives the model to focus on the key information of the captions. In the generation phase, we proposed a novel evaluation function by combining the likelihood with the consensus score, which could fix misrecognition and make the generated sentences more natural sounding. Extensive experiments conducted on the MS COCO datasets corroborated the superiority of the proposed GRU over the state-of-the-art approaches for image captioning.
In future, We can be enhanced this model for Robot Training , Automatic video subtitling and many more fields by training the model with more dataset.

REFERENCES

[1] Vinyals, Oriol, et al. "Show and tell: A neural image caption generator."Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on.IEEE, 2015.
[2] Gerber, Ralf, and N-H. Nagel."Knowledge representation for the generation of quantified natural language descriptions of vehicle traffic inimage sequences."Image Processing, 1996.Proceedings., InternationalConference on. Vol. 2.IEEE, 1996.
[3] Yao, Benjamin Z., et al. "I2t: Image parsing to text description." Proceedings of the IEEE 98.8 (2010): 1485-1508.
[4] Farhadi, Ali, et al. "Every picture tells a story: Generating sentences from images." Euro-pean conference on computer vision.Springer, Berlin, Heidelberg, 2010.
[5] Yang, Yezhou, et al. "Corpus-guided sentence generation of natural images." Proceedings of the Conference on Empirical Methods in Natural Language Processing.Association for Computational Linguistics, 2011.
[6] Kulkarni, Girish, et al. "Babytalk: Understanding and generating simple image descrip-tions." IEEE Transactions on Pattern Analysis and Machine Intelligence 35.12 (2013): 2891-2903.
[7] Mitchell, Margaret, et al. "Midge: Generating image descriptions from computer vision de-tections." Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics.Association for Computational Linguistics, 2012.
[8]Software Engineering a practitioners approach by Roger S.Pressman
[9] Neural Networks and Deep Learning by Michael Nielsen.
[10]Gulli and Kapoor'sTensorFlow Deep Learning Cookbook.
[11]Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algo-rithms Book by Nicholas Locascio and Nikhil Buduma

INNO SPACE

SJIF Scientific Journal Impact Factor

Impact Factor:
7.488

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details