International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST

Color Detection and Image Caption Generator using Machine Learning

***[1]Minhaj Begum,[2]V Nithya, [3]P Pallavi**

*[1]Assistant Professor, Department of Information Technology, Bhoj Reddy Engineering College for Women, Hyderabad, India

[2,3]Student, Department of Information Technology, Bhoj Reddy Engineering College for Women, Hyderabad, India

ARTICLEINFO

Article History:

Accepted: 05 April 2023
Published: 04 May 2023

Publication Issue

Volume 10, Issue 3
May-June-2023

Page Number

09-12

ABSTRACT

Image captioning aimsto automatically generate a sentence description for an image. Our project model will take an image as input and generate an English sentence as output, describing the contents of the image. It has attracted much research attention in cognitive computing in the recent years. The task is rather complex, as the concepts of both computer vision and natural language processing domains are combined together. We have developed a model using the concepts of a Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM) model and build a workingmodel of Image caption generator by implementing CNN with LSTM. The CNN works as encoder to extract features from images and LSTM works as decoder to generates words describing image. After the caption generation phase, we use BLEU Scores to evaluate the efficiency of our model. Thus, our system helps the user to get descriptive caption for the given input image.

Keywords : Convolutional Neural Network, Long Short Term Memory, Computer Vision, Natural Language Processing.

I. INTRODUCTION

To develop a system for users, which can automatically generate the description of an image with the use of CNN along with LSTM. Automatically describing the content ofimages using natural language is a fundamental and challenging task. With the advancement in computing power along with the availability of huge datasets, building models that can generate captions for an image has become possible. On the other hand, humans are able to easily describe the environments they are in. Given a picture, it's natural for a person to explain an immense amount of details about this image with a fast glance. Although great development has been made in computer vision, tasks such as recognizing an object, action classification, image classification, attribute classification and scene recognition are possible but it is a relatively new task to let a computer describe an image thatis forwarded to it in the form of a human-like sentence. For this goal of image captioning, based on semantics of images should be captured here and expressed in the desired form of natural languages. It has a great impact in the real world, for instance by helping visually impaired people better understand the content of images on the web. So, to make our image caption generator model, we will be merging CNN-RNN architectures. Feature extraction from images is done using CNN. We have used the pre-trained model Exception. The information received from CNNis then used by LSTM for generating a description of the image. However, sentences that are generated using these approaches are usually generic descriptions of the visual content and background information is ignored. Such generic descriptions do not satisfy in emergent situations as they, essentially replicate the information present in the images and detailed descriptions regarding events and entities present in the images are not provided, which is imperative to understanding emergent situations.

The objective of our project is to develop a web based interface for users to get the description of the image and to make a classification system in order to differentiate images as per their description. It can also make the task of SEO easier which is complicated as they have to maintain and explore enormous amounts of data.

II. RELATED WORK

In Literature review, various references of the existing projects are taken into consideration which are similar to this current project.

In this paper one of the most popular deep neural networks is the Convolutional Neural Network (CNN) is explained. There are multiple layersin CNN; such as convolutional layer, & nonlinearity layer, &pooling layer and fully- connected layer as well. The CNNhas an excellent performance in machine learning problems and one of the most common algorithms.

In this paper Sepp Hochreiter explain about the deep neural network algorithm long short team Memory (LSTM). LSTM is local in both space aswell as in time; the computational complexity is per time of step and also the weight pattern representation. In comparison to other algorithm LSTM leads to many more successful runs, and learn much faster. It's even solve complex, artificial long time lag tasks that have never been solved by previous recurrent network.

The fundamental problem in artificial intelligence that connects computer vision and Natural language processing is automatically describing the content of an image.In this paper,A.L systematically analyze a deep neural networks based image caption generation method. Here an image is given as the input, and the method as outputin the form of sentence in English describing the content of the image. They analyze three components of the method: convolutional neural network (CNN), recurrent neural network (RNN) and sentence generation. This model analyze image and generate more trival and relevant words for images.

Current image captioning approaches generate descriptions which lack specific information, such as named entities that are involved in the images. HereDi Lu, Spencer Whitehead had proposed a very new task which generates descriptive image captions, given images as input. Asimple solution to this problem that we are proposing is that we will train a CNN-LSTM model so that it can generate a caption based on the image.

Automatically describing the content of an image using properly arranged English sentences is a tough challenging task, but it could is something very necessary for helping visually impaired people. Modern smart phones are able to take the photographs, which can help in taking surrounding images for visually impaired peoples. Here images as input can generate captions that can be loud enough so that visually impaired can hear, so that they can get a better sense of things present in there surrounding. Here Christoper Elamri uses a CNN model to extract features of an image. These features are then fed into a RNN or a LSTM model to generate a description of the image in grammatically correct English sentences describing the surroundings.

III. PROPOSED SYSTEM

By providing appropriate, expressive, and fluid subtitles, Deep Neural Networks can tackle the problems that emerge in both versions. Accelerate the creation of subtitles. Users of social media will no longer have to waste hours searching for subtitles on Google with the system we offer. Our technology provides an easy-to-use platform for social network users to upload selected photographs. Uploading photographs does not require users to manually input captions. The proposed framework is capable of resolving the picture search issue. Color and black-and-white photos of any size can be uploaded and also can read the caption out in English. Tensor flows and algorithms can be used by neural networks to solve any problem and provide appropriate, expressive, and fluent subtitles. It is feasible to calculate automatic metrics efficiently. You won't have to waste time searching for captions because they'll be generated automatically.

A. Task

The goal is to create a system that accepts an image in the form of a dimensional array, characterizes it, and provides syntactically and grammatically accurate statements as an output.

B. Corpus

As a corpus, I used the Flickr 8K dataset. The collection contains 8000 photos, each with five captions. A single image with five descriptions can help you grasp all conceivable circumstances. A training dataset Flickr 8k.trainImages.txt (6,000 photos), a development dataset Flickr 8k.devImages.txt (1,000 images), and a test dataset Flickr 8k.testImages.txt are all included in the dataset (1000 images).

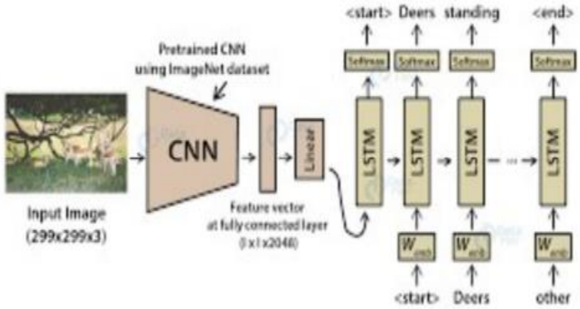

Fig 1 : system architecture

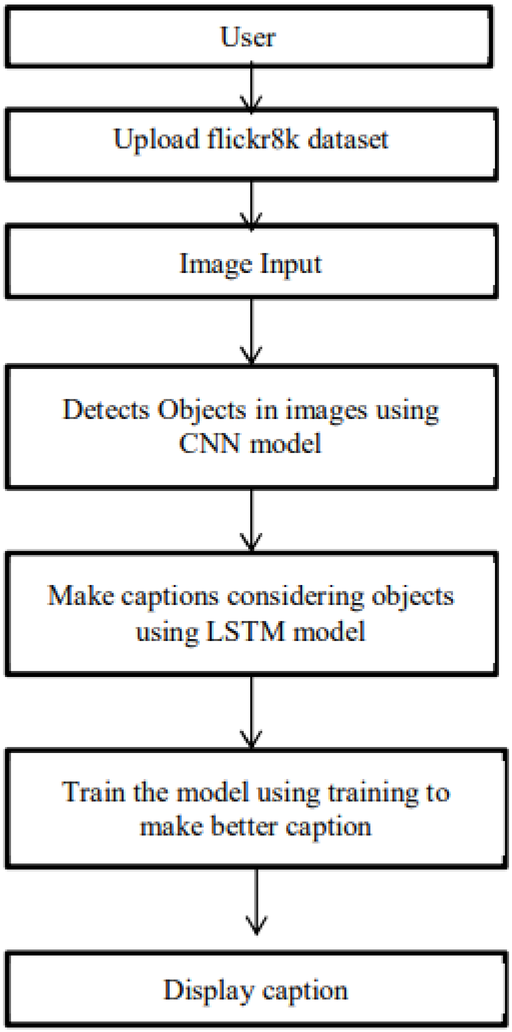
User
Upload flickr8k dataset
Image Input
Detects Objects in images using CNN model
Make captions considering objects using LSTM model
Train the model using training to make better caption
Display caption

Fig 2: Flow Chart

IV. CONCLUSION

We looked at deep learning-Image Captioning approaches in this article. It demonstrated how to categories image annotation approaches, displayed a generic block diagram of the main groupings of, and highlighted the advantages and disadvantages of. We've broken down the benefits and drawbacks of each of the metrics and datasets. There's also a quick rundown of the experiment's findings. We briefly discussed the various research options that could be pursued in this area. Although deep learning based image labeling systems have made significant progress in recent years, robust image labeling approaches that can create high quality labels for practically every image have yet to be achieved. With the introduction of new deep learning network designs, automated captioning will remain a hot topic of research for some time. It makes use of the Flickr 8k dataset, which contains around 8000 photographs, as well as the captions, which are kept in a text file. Although deep learning-based image labeling systems have made significant progress in recent years, robust image labeling approaches that can create high-quality labels for practically every image have yet to be achieved. With the introduction of new deep learning network designs, automated captioning will remain a hot topic for a long time. The number of people using social media is growing every day, and the majority of them submit images, therefore the supply of captions will grow in the future. As a result, this project will be beneficial to them.

V. REFERENCES

[1]. William Fedus, Ian Goodfellow, and Andrew M Dai. Maskgan: Better text generation. arXiv preprint arXiv:1801.07736, 47, 2018.

[2]. Girish Kulkarni, Visruth Premraj, Sagnik Dhar, Siming Li, Yejin Choi, Alexander C Berg, and Tamara L Berg. Baby talk: Understanding and generating image descriptions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35:2891–2903, June 2013.

[3]. Yunchao Gong, Liwei Wang, Micah Hodosh, Julia Hockenmaier, and Svetlana Lazebnik. Improving image-sentence embeddings us- ing large weakly annotated photo collections. European Conference on Computer Vision. Springer, pages 529–545, 2014.

[4]. Peter Young Micah Hodosh and Julia Hockenmaier. Framing image description as a ranking task: Data, models and evaluation metrics. Journal of Artificial Intelligence Research, 47:853–899, 2013.

[5]. Ryan Kiros, Ruslan Salakhutdinov, and Richard S Zemel. Unifying visual-semantic embeddings with multimodal neural language models. Workshop on Neural Information Processing Systems (NIPS), 2014.

[6]. Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visual attention. International Conference on Machine Learning, 2048- 2057, 2015.

[7]. Ting Yao, Yingwei Pan, Yehao Li, Zhaofan Qiu, and Tao Mei. Boosting image captioning with attributes. IEEE International Conference on Computer Vision (ICCV), pages 4904–4912, 2017.

[8]. Quanzeng You, Hailin Jin, Zhaowen Wang, Chen Fang, and Jiebo Luo. Image captioning with semantic attention. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4651–4659, 2016.

Cite this article as :

Minhaj Begum, V Nithya, P Pallavi, "Color Detection and Image Caption Generator using Machine Learning ", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 10 Issue 3, pp. 09-12, May-June 2023.

Journal URL : https://ijsrst.com/IJSRST5231033