Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 7-Issue 1, January 2024
Paper: 2

Image Description Generator Using Deep Learning

Dr.D.Rasi, Associate Professor
Department of CSE
Sri Krishna College of Engineering and Technology
rasid@skcet.ac.in

Ms.A.Adlin, Assistant Professor
Department of CSE
Sri Krishna College of Engineering and Technology
adlina@skcet.ac.in

Abiraj R
Department of CSE
Sri Krishna College of Engineering and Technology
19eucs001@skcet.ac.in

Aparna K
Department of CSE
Sri Krishna College of Engineering and Technology
19eucs012@skcet.ac.in

Darwesh Fazil A
Department of CSE
Sri Krishna College of Engineering and Technology
19eucs026@skcet.ac.in

Grace Ebenezer R
Department of CSE
Sri Krishna College of Engineering and Technology
19eucs036@skcet.ac.in

Abstract- **In order to provide a cogent and comprehensive caption, the image description generator must first comprehend the content and context of the image. This paper describes a deep learning-based method for creating picture descriptions. The system's primary difficulty is producing appropriate captions for the supplied image. This model was developed using a sizable collection of photos and the textual descriptions that go with them. This collection of images and their description were taken from the Flickr8k dataset. The system employs an LSTM to learn the text features, a convolutional neural network (CNN) architectural model that has already been trained to learn the picture features, and the combined output of the two to provide a description for the image. It comprises a decoder that creates the caption and an encoder that collects information from the input image. The model is tested against many benchmark datasets and demonstrates that it performs better in terms of description quality and variety than current methods. The proposed model exhibits strong generalization capabilities and can provide precise and evocative captions for a variety of photos including emotive analysis of the image.**

Overall, this model is capable of producing visual descriptions and has the potential to be used in a lot of real-world scenarios. It has a significant effect in the real world, for example, by assisting those with visual impairments in comprehending the content of images from various sources.

Keywords – Image description, Deep Learning, CNN, LSTM, Sentiment analysis, NLP, RNN

I. INTRODUCTION

A long-standing challenge in artificial intelligence is teaching a computer system to recognize items and describe them using Natural Language Processing (NLP). Until recently, academics studying computer vision regarded this as an impossibility. Models are frequently constructed that will generate captions for an image with the rapid breakthroughs in deep learning techniques, the availability of enormous datasets, and processing capacity. Using deep learning and computer vision, image description generators identify a picture's context and annotate it with pertinent details. The labeling of a picture with English keywords using datasets from model training is part of it. Upon training of the CNN model, which is in charge of extracting image characteristics, these features are passed to the LSTM model, which creates the picture description. The emotive analysis of the image is performed using Convolution Neural Network (CNN), a binary classifier based on deep learning framework. The interplay of visual characteristics with particular expressions of live things is better captured by the model.

II. LITERATURE REVIEW

The interpretation from various literature surveys based on various CNN, RNN models have been considered, of which [1] uses an LSTM (Long-Short Term Memory) based RNN model to build captions and uses Flicker8k dataset to train the

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 7-Issue 1, January 2024
Paper: 2

model. An image caption generator system was proposed in [2] which involves object detection, feature extraction, Convolution Neural Network (CNN) for feature extraction and for scene classification, Recurrent Neural Network (RNN) for human and objects attributes, RNN encoder and a fixed length RNN decoder system.[4] proposes a method to build an image caption generator by implementing the convolutional neural network with long short-term memory.[5] devises an effectual hybrid optimization-based Deep Learning (DL) technique for color image segmentation and classification.[6] proposes a Flower Henry Gas Solubility Optimization-based Deep Convolution Neural Network (FHGSO-based Deep CNN) for image classification.

III. METHODOLOGY

i. Convolutional Neural Network (CNN)

CNN is a kind of deep learning that uses specialized deep neural networks to recognize and categorize pictures. It is used to process data that is displayed as pictures that resemble 2D matrices. It can handle images that have been resized, translated, and rotated. By scanning the visual picture from top to bottom and left to right, it analyzes it and pulls out the pertinent information. Lastly, it aggregates all the image categorization characteristics.

ii. Long short-term memory (LSTM)

Long Short-Term Memory (LSTM), a form of RNN (Recurrent Neural Network), can solve the sequence Prediction issues. It is mostly used to forecast the following word. The essential information is carried out and the irrelevant information is discarded when inputs are processed using LSTM

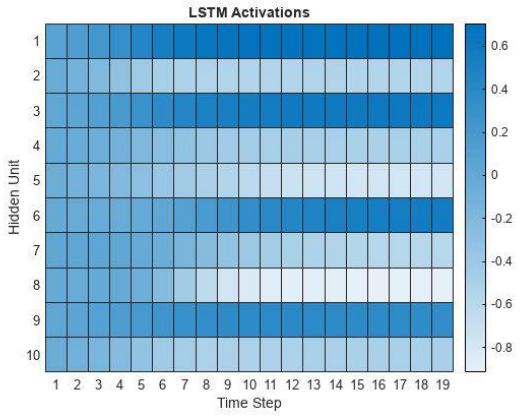

Figure 1. The Heat map for LSTM Activators Hidden Unit is plotted against the Time Step is given above

iii. CNN LSTM Architecture

The CNN-LSTM architecture combines LSTMs to Facilitate sequence prediction with CNN layers for Feature extraction on input data. This approach is Especially made for issues involving the sequence prediction of spatial inputs, such as photos or videos.

They are frequently utilized in activities including Activity recognition, image and video description, and many others.

iv. Performance Evaluation

A powerful algorithm is developed that can take an image input in the form of a dimensional array and produce an output consisting of a sentence that accurately describes the picture and follows grammar and syntactic rules. Using Convolution Neural Networks (CNN), the image is subjected to a sentimental analysis.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 7-Issue 1, January 2024
Paper: 2

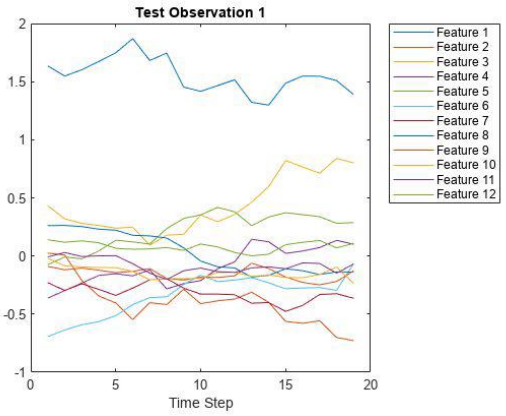

Figure 2. The Test Observation for various features over a varied Time Step is given above

v. Performance Metrics

The proposed system achieved a reasonable accuracy on the test data. The tweaked model is capable of achieving an average BLEU score of 0.545880 with a maximum BLEU score of 0.675894 and a minimum BLEU score of 0.315244. However, more features must be taken into consideration to enhance CNN. The process comprises of employing different optimizers to adjust the parameters, enhancing the data, training deeper networks, and overfitting.

IV. FUTURE WORKS

Video inputs are decoded using an open CV library into a sequence of images. Next, in order to produce descriptions, these images are input into the model to predict suitable descriptions. The created description might be used to explain the motion graphical data. As this model has been developed as an API it can serve various requests from applications across multiple platforms to transcribe the motion graphical data.

V. CONCLUSION

In this paper, an optimized Convolution Neural Network (CNN) architecture is proposed for creating picture descriptions that users can request based on an image. The language-based model converts the features and objects that the image-based model has extracted into a natural language phrase. The image-based model retrieves features from a picture. The ultimate goal of an image description generator is to enhance social networking platforms, picture indexing, and accessibility for those with visual impairments through automatically generated captions or descriptions.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 7-Issue 1, January 2024
Paper: 2

REFERENCES

[1] A. Verma, H. Saxena, M. Jaiswal and P. Tanwar, "Intelligence Embedded Image Caption Generator using LSTM based RNN Model," 2021 6th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2021, pp.963-967, ISBN:978-1-6654-3587-1, doi: 10.1109/ ICCES51350.2021.9489253.

[2] N. K. Kumar, D. Vigneswari, A. Mohan, K. Laxman and J. Yuvaraj, "Detection and Recognition of Objects in Image Caption Generator System: A Deep Learning Approach," 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Coimbatore, India, 2019, pp. 107-109, ISSN: 2575-7288 doi: 10.1109/ICACCS.2019.8728516.

[3] Yeshasvi, Mogula & Thankaraj, Subetha. (2022). "Image Caption Generator Using Machine Learning and Deep Neural Networks". 10.1007/ISBN: 978-981-19-0825-5_14.

[4] Deepa, S.N., Rasi, D. FHGSO: Flower Henry gas solubility optimization integrated deep convolutional neural network for image classification. *Appl Intell* 53,7278–7297(2023)ISSN:0924669X,15737497. https://doi.org/10.1007/s10489-022-03834-4,

[5] Deepa, S.N., Rasi, D. Hybrid optimization enabled deep learning model for colour image segmentation and classification, International Journal of Neural Computing and Applications, Springer London. ISSN: 09410643,14333058. https://link.springer.com/article/10.1007/s00521-022-07614-6,