International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 11 | November - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Automatic Intelligence Caption Generator

Trushna Kapadnis[1], Anuja Modhave[2], Akanksha Narwade[3], Umesh Wagh[4] , Prof. Dr.Deepali Sale[5]

[1] *Department of Computer Engineering, Dr. D. Y. Patil College of Engineering and Innovation, Varale.*
[2] *Department of Computer Engineering, Dr. D. Y. Patil College of Engineering and Innovation, Varale.*
[3] *Department of Computer Engineering, Dr. D. Y. Patil College of Engineering and Innovation, Varale.*
[4] *Department of Computer Engineering, Dr. D. Y. Patil College of Engineering and Innovation, Varale.*

Abstract – An Image Caption Generator is a sophisticated AI system that combines computer vision and natural language processing to automatically create descriptive textual captions for images. This technology utilizes deep learning, particularly Convolutional Neural Networks (CNNs), to analyze and extract meaningful visual features from the input image. These features capture details about the objects, scenes, and elements within the image. Subsequently, a natural language processing model, often built on Recurrent Neural Networks (RNNs) or Transformers, processes these visual features and generates coherent, contextually relevant captions. Post-processing steps may be applied to enhance the quality of the generated text. The primary aim of Image Caption Generators is to facilitate image understanding, improve accessibility, and enhance content search ability by providing human-readable descriptions for visual content. This technology is instrumental in various fields, including content tagging, accessibility tools for the visually impaired, and enhancing user experiences in multimedia content management systems, ultimately bridging the gap between visual and textual information for a more comprehensive and human-like interpretation of image.

Key Words:Image Recognition, Internet, Image-To-Caption, Contextual Understanding,Image Captioning.

1. INTRODUCTION

Image caption generator is a process of recognizing the context of an image and annotating it with relevant captions using deep learning, and computer vision. It includes the labeling of an image with English keywords with the help of datasets provided during model training.Generating complete and natural image descriptions automatically has large potential effects, such as titles attached to news images, descriptions associated with medical images, text-based image retrieval, information accessed for blind users, humanrobot interaction. These applications in image captioning have important theoretical and practical research value.Image captioning is a more complicated but meaningful task in the age of artificial intelligence. Given a new image, an image captioning algorithm should output a description about this image at a semantic level. In this an Image caption generator, basis on our provided or uploaded image file It will generate the caption from a trained model which is trained using algorithms and on a large dataset. The main idea behind this is that users will get automated captions when we use or implement it on social media or on any applications.

2. OBJECTIVE

1. To automatically generate accurate and contextually relevant textual descriptions for a wide range of images.

2. To provide a more inclusive experience for visually impaired individuals by offering detailed image captions.

3. To assist in organizing and retrieving images in multimedia databases and digital repositories.

4. To create more intuitive and natural ways for users to interact with machines, enhancing human-computer communication in the context of visual data.

5. To facilitate a comprehensive interpretation of visual content, bridging the gap between the visual and textual domains

2.1 METHODOLOGY

The development of machine learning tools like picture captioning is beneficial for persons who are blind and cannot understand sights. The descriptions of images can be read out to visually impaired people using an AI-powered image caption generator, helping them to better understand their surroundings.The model proposed takes an image I as input and is trained to maximize the probability of p(S|I) where S is the sequence of words generated from the model and each word St Is generated from a dictionary built from the training dataset. The input image I is fed into a deep vision Convolutional Neural Network (CNN) which helps in detecting the objects present in the image. The image encodings are passed on to the Language Generating Recurrent Neural Network (RNN) which helps in generating a meaningful sentence for the image as shown in the fig. 13. An analogy to the model can be given with a language translation RNN model where we try to maximize the p (T|S) where T is the translation to the sentence S. However, in our model the encoder RNN which helps in transforming an input sentence to a fixed length vector is replaced by a CNN encoder. Recent research has shown that the CNN can easily transform an input image to a vector.

3. SYSTEM REQUIREMENTS:

Hardware Requirements

- RAM minimum required is 5 GB
- Hard Disk : 500 GB
- Processor : Intel i5 Processor
- Operating System : Windows 10
- Key Board: Standard Windows Keyboard
- Speed :2.80 GHz

Software Requirements:

- Operating System : Windows 10
- IDE : Spyder
- Programming Language : Python

4. SYSTEM ARCHITECTURE

The architecture of an image caption generator involves a multi-stage process aimed at seamlessly blending computer vision and natural language processing. Initially, a Convolutional Neural Network (CNN) is employed to extract intricate features from the input image. This results in a condensed feature vector that encapsulates essential visual information. Subsequently, this vector is fed into a Caption Generator, typically implemented as a Recurrent Neural Network (RNN) or Transformer. The Caption Generator, equipped with an optional attention mechanism, sequentially processes the feature vector, producing a coherent sequence of words forming the image caption. To facilitate semantic understanding, an embedding layer converts words into continuous vector representations. During training, the model refines its parameters by minimizing the disparity between predicted and ground truth captions using a suitable loss function. Decoding strategies, like beam search, enhance the caption's quality and diversity. Evaluation metrics such as BLEU and METEOR assess the generated captions against reference captions. Once trained, the model can be deployed to generate descriptive captions for new, unseen images, exemplifying a comprehensive fusion of computer vision and language generation technologies.

In the system architecture of an image caption generator, the initial phase involves feeding an input image through a pre-trained Convolutional Neural Network (CNN). This CNN extracts intricate features, transforming the image into a compact feature vector that encapsulates crucial visual information. This vector is then handed over to the Caption Generator, often implemented as a Recurrent Neural Network (RNN) or Transformer, which takes on the task of generating a coherent and contextually relevant caption. The decoding process incorporates strategies like beam search to enhance the fluency and diversity of the generated captions.

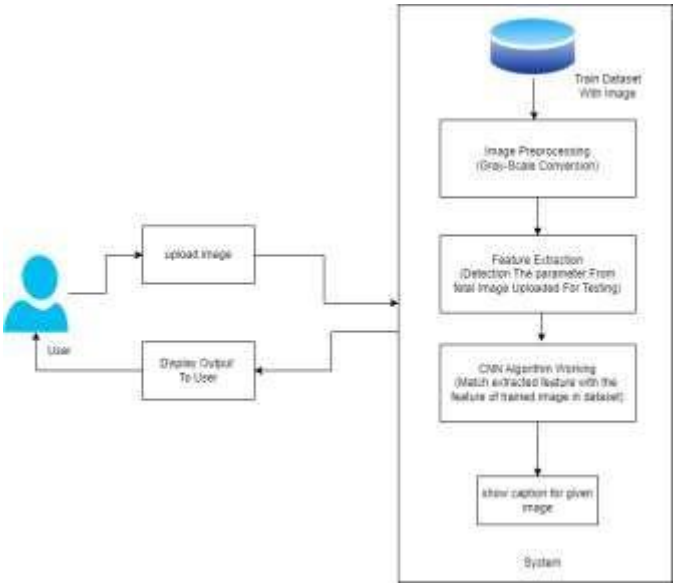

Fig. System Architecture

4.1. DATA FLOW

In Data Flow Diagram,we Show that flow of data in our system in DFD0 we show that base DFD in which rectangle present input as well as output and circle show our system,In DFD1 we show actual input and actual output of system input of our system is text or image and output is rumor detected like wise in DFD 2 we present operation of user as well as admin.

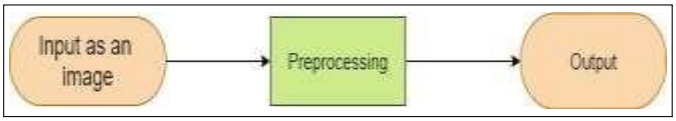

Fig.4.1.Data Flow

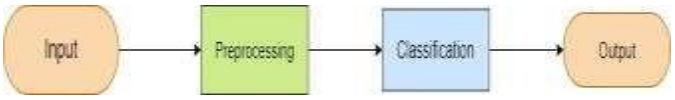

Fig.4.2.Data Flow

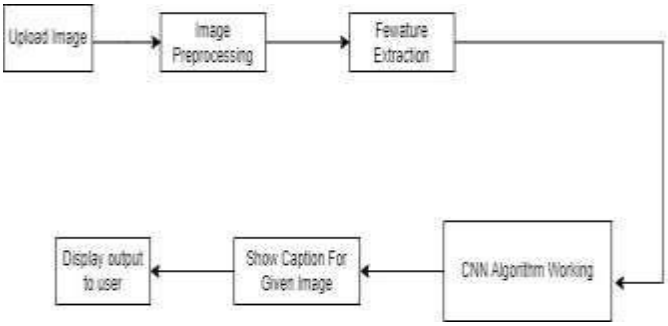

Fig.4.3.Data Flow

5. CONCLUSIONS

In conclusion, the Image Caption Generator using CNN algorithms is a transformative application of artificial intelligence that bridges the visual-textual divide. By automatically generating accurate and contextually relevant captions for images, it enhances accessibility, content management, and user experiences across various domains. While it offers substantial advantages, it's not without limitations, including accuracy challenges and potential biases. However, its versatility and potential for customization make it a powerful tool with wide-ranging applications, from improving content tagging and accessibility to advancing research and education, and enriching human-computer interaction.

6. FUTURE SCOPE

The future scope of Image Caption Generator technology is exceptionally promising, with several exciting developments on the horizon. Advancements in artificial intelligence and deep learning techniques are expected to significantly enhance the accuracy and precision of image captions, making them even more contextually relevant and reducing ambiguities. Multilingual capabilities are likely to become more widespread, enabling these generators to serve a global audience. Furthermore, the incorporation of emotion and sentiment analysis may enable the generation of captions that reflect the emotional content within images, opening up new avenues for creative and empathetic applications. Customization is another area with potential for growth, offering users more flexibility and user-friendly interfaces to tailor captions to their specific preferences. Additionally, as processing power continues to increase, we can anticipate real-time image captioning, making this technology even more responsive and versatile across numerous domains and industries.

ACKNOWLEDGEMENT

We are extremely grateful to our mentor and our Project Coodinator , Prof.Dr Deepali Sale, for his keen interest in our project work and assistance in refining our application. They are always the first to motivate and support us in accomplishing this project adequately. We express our gratitude to the entire teaching and non-teaching staff in our department.

REFERENCES

1. HaoranWang , Yue Zhang, and Xiaosheng Yu, "An Overview of Image Caption Generation Methods", (CIN-2020)
2. B.Krishnakumar, K.Kousalya, S.Gokul, R.Karthikeyan, and D.Kaviyarasu, "IMAGE CAPTION GENERATOR USING DEEP LEARNING", (international Journal of Advanced Science and Technology- 2020)
3. MD. Zakir Hossain, Ferdous Sohel, Mohd Fairuz Shiratuddin, and Hamid Laga, "A Comprehensive Survey of Deep Learning for Image Captioning",(ACM-2019)
4. Rehab Alahmadi, Chung Hyuk Park, and James Hahn, "Sequence-tosequence image caption generator", (ICMV-2018)
5. Oriol Vinyals, Alexander Toshev, SamyBengio, and Dumitru Erhan, "Show and Tell: A Neural Image Caption Generator",(CVPR 1, 2- 2015)
6. Priyanka Kalena, Nishi Malde, Aromal Nair, Saurabh Parkar, and Grishma Sharma, "Visual Image Caption Generator Using Deep Learning", (ICAST2019)
7. Pranay Mathur, Aman Gill, Aayush Yadav, Anurag Mishra, and Nand Kumar Bansode,"Camera2Caption: A Real-Time Image Caption Generator", International Conference on Computational Intelligence in Data Science(ICCIDS) – 2017
8. D. Bahdanau, K. Cho, and Y. Bengio. "Neural machine translation by jointly learning to align and translate.arXiv:1409.0473", 2014.