iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

IMAGE CAPTION GENERATOR USING DEEP LEARNING

B. AKSHAY GOUD
UG Student, Department of Artificial Intelligence and Data Science, J.B. Institute of Engineering and Technology, Hyderabad.
P.SAI KUMAR
S. KARTHIK
P. AKSHAY
UG Students, Department of Artificial Intelligence and Data Science, J.B. Institute of Engineering and Technology, Hyderabad.

ABSTRACT
Image caption generation is a significant advancement in Artificial Intelligence, enabling computers to describe images with human-like understanding. This technology has valuable applications in medical imaging, education, and e-learning. The primary goal is to generate accurate English captions for images, which remains a critical challenge.This project introduces an image caption generator that utilizes beam search and greedy search optimization techniques to produce captions. It incorporates a pre-trained Convolutional Neural Network (CNN), specifically the VGG16 model, to extract image features. Additionally, Long Short-Term Memory (LSTM) is used to interpret text features, combining both to generate meaningful captions.
The model's capabilities extend to video captioning by integrating Convolution 3D (C3D). Continuous research, ethical considerations, and responsible data usage are essential to further enhance accuracy and minimize potential drawbacks. This technology has the potential to significantly impact various fields, making AI-driven image interpretation more efficient and accessible.

Keywords:
Image caption generation, deep learning, artificial intelligence, Convolutional Neural Network (CNN), VGG16, Long Short-Term Memory (LSTM), beam search, greedy search, text generation, feature extraction, image processing, video captioning, Convolution 3D (C3D), natural language processing (NLP), machine learning, autonomous captioning, medical imaging, e-learning, dataset training, model optimization, AI ethics, responsible data usage.

INTRODUCTION

Image caption generation is a significant application of Artificial Intelligence, combining deep learning and computer vision to describe images in a human-like manner. This technology has broad applications in social media, medical imaging, education, and e-learning. The project aims to develop an advanced image caption generator using pre-trained Convolutional Neural Networks (CNN), specifically VGG16, alongside Long Short- Term Memory (LSTM) networks. The CNN extracts image features, while the LSTM decodes them to generate meaningful captions. The approach also explores more advanced CNN architectures beyond ResNet, Inception, and EfficientNet for improved feature extraction. Additionally, the project expands into video captioning using Convolution 3D (C3D), enhancing accessibility for individuals with hearing impairments. Despite its benefits, image captioning faces technical and ethical challenges, requiring ongoing research and responsible AI development. The system incorporates a user-centric classification approach, making it adaptable to various domains and platforms, pushing the boundaries of AI-driven image interpretation.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

OBJECTIVES

The primary objectives of Image caption Generator using deep learning are:

1. Accurate Image Interpretation

Develop an AI model that can analyze and understand image content, generating meaningful and contextually relevant captions.

2. Enhanced Feature Extraction

Utilize advanced Convolutional Neural Networks (CNN), such as VGG16, ResNet, and Inception, to extract high-quality image features for improved caption accuracy.

3. Effective Caption Generation

Implement Long Short-Term Memory (LSTM) networks to generate coherent and grammatically correct English captions based on extracted image features.

4. Optimization Techniques

Employ beam search and greedy search strategies to enhance the quality of generated captions.

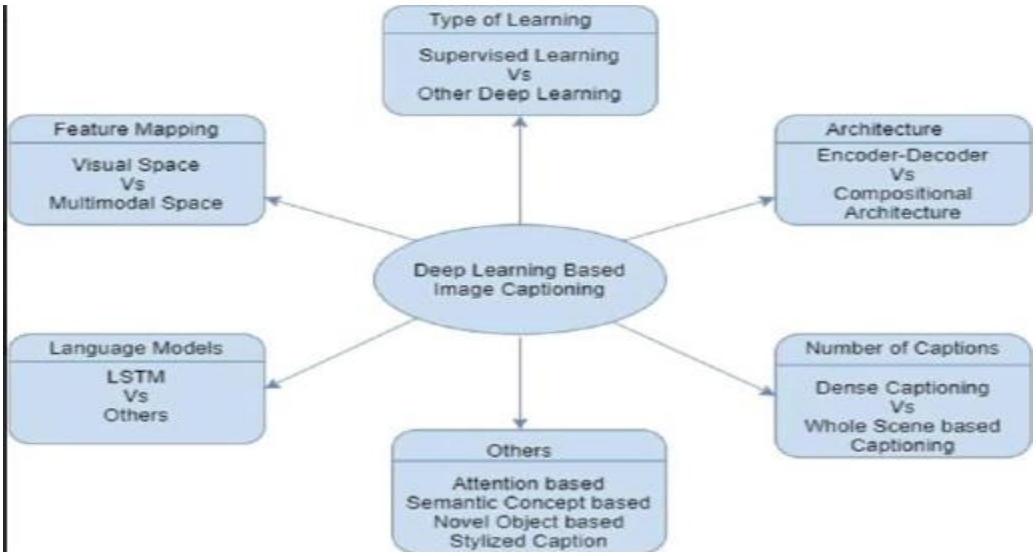
5. Expansion to Video Captioning

Incorporate Convolution 3D (C3D) modules to extend the model's capability to video captioning, making it useful for accessibility applications.

METHODOLOGY

The Flood Prediction using machine learning involves;

1. Data Collection and Preprocessing
2. Feature Extraction Using CNN
3. Caption Generation Using LSTM
4. Optimization Techniques
5. Video Captioning Extension
6. Model Training and Evaluation 7.Deployment and User Adaptability



iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

RESULTS AND DISCUSSION

While transformer-based models dominate many AI applications, CNN-RNN architectures like VGG16 combined with LSTM remain highly effective for image captioning, particularly in resource-constrained environments. These models balance visual feature extraction with sequential text generation, making them a strong choice for this task.The model generates ranked captions for each image, describing key objects, actions, or emotions. The relevance and diversity of captions depend on model settings, with higher temperatures producing more varied but potentially less accurate outputs. CNN-RNN models efficiently combine visual and sequential data, with CNNs extracting image features and RNNs generating word-by-word captions.

ACKNOWLEDGEMENT

We acknowledge with deep appreciation all the efforts made by everyone in successfully carrying out this project on Image Caption Generator using deep Learning. Above all, we wish to thank our faculty members and institution for the insightful guidance, inspiration, and unfailing support received from them throughout this study. Their input and experience have helped mold the focus of our project. Finally, we recognize the open- source research, datasets, and existing studies that were made available to us as the groundwork for our study. Without them, it would have been impossible to implement our model. Our project has been a worthwhile learning experience, and we appreciate being able to make a contribution towards DL-based image caption.

CONCLUSION

In conclusion, our Image Caption Generator framework leverages deep learning techniques to bridge the gap between advanced technology and user accessibility. By integrating CNNs for feature extraction, LSTMs for coherent caption generation, and optimization methods like BeamSearch and GreedySearch, the system transforms raw image data into meaningful and contextually relevant captions. This approach not only enhances the accuracy and relevance of generated captions but also ensures that the framework remains adaptable and user-friendly, even for those with limited technical expertise.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

REFERENCES

1. **Image Caption Generation for Blind Users Of Social Media Websites**
(2023) Authors: Roshan Adhithya SS, Priyadharshini M, Lekshmi Kalinathan
Link:https://www.researchgate.net/publication/370740210_Image_Caption_Generation_For Blind_Users_Of_Social_Media_Websites
2. **Meshed-Memory Transformer for Image Captioning by Cornia et al. (2020)**
Link: [1912.08226] Meshed-Memory Transformer for Image Captioning (arxiv.org)
3. **"Visual to Text: Survey of Image and Video Captioning"**
(2019) Authors: Shuai Li, Zhihao Tao, Kai Li, Yun Fu
Link:
IEEEXplore
4. **Visual Image Caption Generator Using Deep Learning (2019)**
Authors: Grishma Sharma, Priyanka Kalena, Nishi Malde, Aromal Nair, Saurabh Parkar
Link: (PDF) Visual Image Caption Generator Using Deep Learning (researchgate.net)