DeepLens: Integrating Deep Learning for Image Captioning and Hashtag Generation

MR. G. NUTAN KUMAR

Assistant Professor, Dept. of Information Technology, Sreenidhi Institute of Science and Technology

DR. K. KRANTHI KUMAR

Associate Professor, Dept. of Information Technology, Sreenidhi Institute of Science and Technology

K. PAVAN

B.Tech Student, Dept. of Information Technology, Sreenidhi Institute of Science and Technology

20311a12a5@sreenidhi.edu.in

V. VIHAR

B.Tech Student, Dept. of Information Technology, Sreenidhi Institute of Science and Technology

20311a12b6@sreenidhi.edu.in

K. KARTHIKEYA

B.Tech Student, Dept. of Information Technology, Sreenidhi Institute of Science and Technology

20311a12a7@sreenidhi.edu.in

Abstract - In this research, we introduce a groundbreaking methodology that leverages deep learning techniques to revolutionize the process of generating descriptive captions and hashtags for images, effectively bridging the gap between computer vision and natural language understanding. Traditional approaches to image caption generation have often relied on rudimentary techniques such as handcrafted features and rule-based systems, which inherently struggle to capture the intricate semantics of images and adapt to diverse datasets. Recognizing these limitations, our novel framework integrates state-of-the-art convolutional neural networks (CNNs) for precise image feature extraction and recurrent neural networks (RNNs), specifically employing the ResNet-50 architecture, for seamless sequence generation. Furthermore, in addition to traditional evaluation metrics such as BLEU scores and human assessments, our system, DeepLens, introduces a groundbreaking feature: automatic hashtag generation. By meticulously analyzing both the content and context of images, DeepLens autonomously generates hashtags that enrich social media content sharing and engagement, presenting a novel paradigm in the realm of image captioning. In addition to its advancements in captioning accuracy and user experience enhancement, DeepLens offers scalability and adaptability to various domains. Its robust architecture allows for seamless integration with different datasets and environments, making it versatile for a wide range of applications. Through this comprehensive approach, we aim to not only enhance the accuracy and relevance of generated captions but also elevate the overall user experience in navigating and interacting with visual content across various platforms and applications.

1. INTRODUCTION

In the realm of computer vision, there persists a notable gap between visual perception and natural language understanding, posing a significant challenge to effective human-machine interactions. While machines excel at recognizing images, their capacity to articulate descriptions in natural language remains limited, hindering seamless communication with users. Existing methods of image captioning often rely on manual inputs or simplistic algorithms, resulting in captions that may lack precision and relevance. Moreover, the absence of automated hashtag generation further restricts machines' ability to optimize social media content sharing and engagement.

To address these challenges, there arises a pressing need for an innovative approach leveraging deep learning methodologies to autonomously generate descriptive captions and hashtags for images. Such an approach would bridge the divide between computer vision and natural language understanding, empowering machines not only to interpret visual content but also to articulate it effectively in natural language. By tackling these challenges head-on, our aim is to enhance machine capabilities in comprehending and communicating visual content, thereby fostering more seamless human-machine interactions and enhancing user experiences across various applications, including social media sharing and content recommendation systems. This comprehensive study focuses on harnessing deep learning techniques to develop a cutting-edge image caption generation model tailored to the renowned Flickr dataset. Leveraging the prowess of neural networks and natural language processing, this research endeavors to push the frontiers of computer vision, equipping machines with the capability to produce precise and contextually relevant image descriptions. Additionally, the exploration extends to automated hashtag generation utilizing generative AI, amplifying the scope of social media content optimization and engagement strategies.

PROJECT OVERVIEW

DeepLens introduces an innovative framework that utilizes advanced deep learning techniques, prominently employing convolutional neural networks (CNNs) for precise image feature extraction and recurrent neural networks (RNNs) like the ResNet-50 architecture for efficient sequence generation. By leveraging these state-of-the-art methodologies, DeepLens aims to surpass the limitations of conventional systems, thereby enhancing the accuracy and relevance of generated captions. Moreover, the project incorporates the evaluation metric BLEU scores to quantitatively assess caption quality. With the integration of hashtag functionality, DeepLens extends its capabilities, promising to revolutionize image captioning. This pioneering approach holds potential for practical applications across diverse domains, including content accessibility, human-computer interaction, and social media marketing.

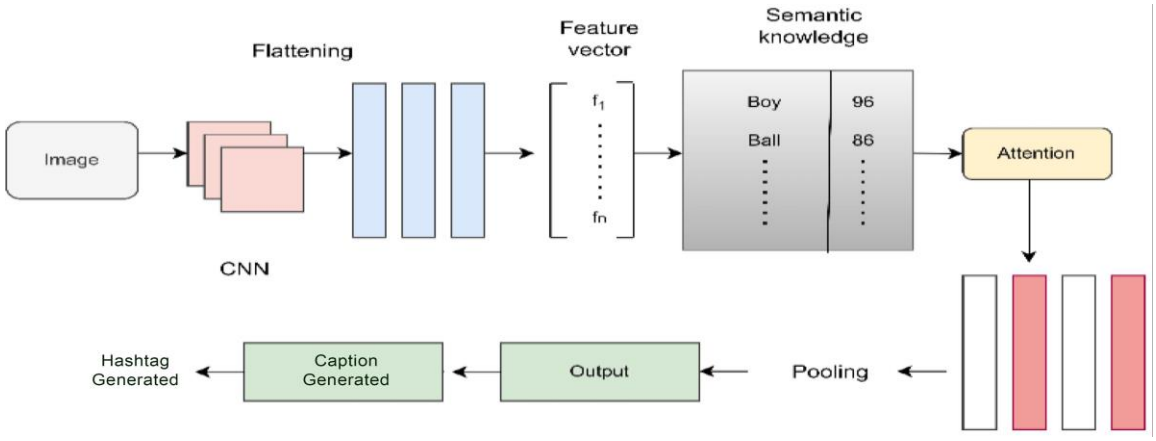
LITERATURE SURVEY

Image caption generation and hashtag generation are pivotal tasks in computer vision and natural language processing, with deep learning methodologies at the forefront of research efforts. Studies by Vinyals et al. (2015) and Xu et al. (2015) have pioneered the use of convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to automatically generate descriptive captions for images, while attention mechanisms and reinforcement learning techniques have further improved

caption accuracy and relevance. Concurrently, research on hashtag generation has explored the application of generative adversarial networks (GANs) and variational autoencoders (VAEs) to produce hashtags that enhance social media engagement. Notably, Zhang et al. (2018) introduced a GAN-based approach for generating hashtags based on image content, demonstrating promising results. Overall, the literature highlights the central role of deep learning in advancing both image captioning and hashtag generation tasks, paving the way for more effective content sharing and user interaction on social media platforms.

2. METHODOLOGY

4.1. MODEL ARCHITECHTURE:



4.2 ResNet-50: ResNet-50 is a convolutional neural network architecture renowned for its depth and performance in image recognition tasks. It comprises 50 layers and utilizes skip connections to address the vanishing gradient problem, enabling the training of deeper networks. ResNet-50 has been widely adopted in various computer vision applications, achieving state-of-the-art results on benchmark datasets like ImageNet.

4.3 RNN: Recurrent Neural Networks (RNNs) are integral to both image caption generation and hashtag generation tasks. In image captioning, RNNs facilitate sequential generation of words based on encoded image features, enabling the creation of descriptive captions. Similarly, in hashtag generation, RNNs are utilized to generate relevant hashtags by analyzing image content and context, thereby enhancing social media content sharing and engagement.

4.4 GAN: To improve prediction accuracy, a machine learning (ML) model known as a generative adversarial network (GAN) pits two neural networks against one another. GANs frequently engage in cooperative zero-sum games unattended and learn new abilities.

4.5 Generative AI: Generative AI, such as generative adversarial networks (GANs) and variational autoencoders (VAEs), plays a pivotal role in image caption and hashtag generation tasks. By providing the generated image captions as input, these AI models can effectively generate relevant hashtags, enhancing social media content engagement. This approach leverages the synergy between caption generation and hashtag prediction, enriching the discoverability and reach of visual content on various platforms.

IMPLEMENTATION

1. Steps involved in Proposed system

MODULES:

Data exploration: We will load data into the system with the help of module.

Processing: We'll train two separate models using the available dataset: one for feature extraction and the other for generating captions. Subsequently, we'll merge these models to train the comprehensive image caption generation model..

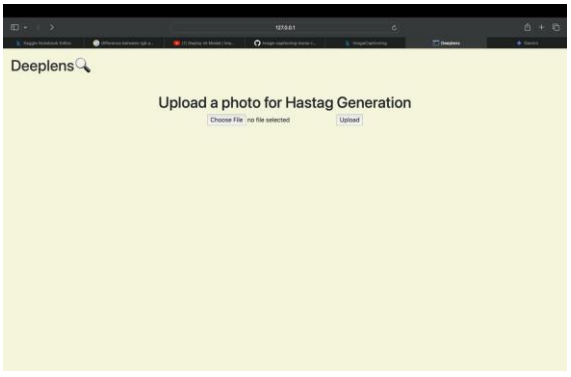
Splitting data into train & test: Data will be separated into train and test groups using this module.

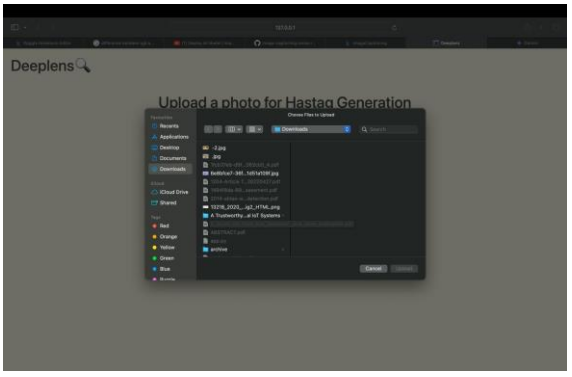
Model generation: Building the model - Feature Fusion ResNet, Feature Fusion with Feed Forward Network Transfer Learning, ResNet50, Keras, GAN.

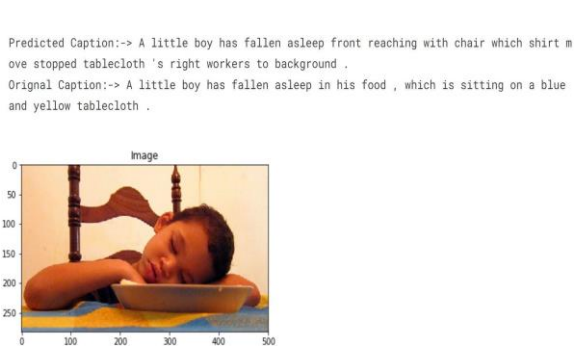
Using above Algorithms accuracy is calculated.

- **User input:** This module will provide data for prediction.

Output:







3. CONCLUSION

In concluding this endeavor, our project stands poised to drive substantial progress at the intersection of computer vision and natural language processing. Through the innovative integration of state-of-the-art deep learning architectures—specifically, the strategic utilization of convolutional neural networks (CNNs) for robust image feature extraction and recurrent neural networks (RNNs) for seamless sequence generation—we aim to forge a harmonious bridge between visual understanding and linguistic expression. Furthermore, by incorporating hashtag generation functionality, our system takes on an added dimension, significantly amplifying its utility in diverse social media contexts and beyond. As we eagerly anticipate the comprehensive evaluation of our model's performance—drawing on metrics like BLEU scores and human assessments—we anticipate gleaning invaluable insights into its efficacy, particularly in ensuring both the accuracy and relevance of generated captions. With a breadth of potential applications ranging from image annotation to the cultivation of enriched accessibility and human-computer interaction experiences, the outcomes of this project promise not only to enrich academic discourse but also to propel tangible advancements in real-world technological landscapes.

REFERENCES

1) Xu Jia et al., Guiding long-short term memory for image caption genaration, 2015.
2) Oriol Vinyals et al., "Show and tell: A neural image caption generator", Computer Vision and Pattern Recognition (CVPR) 2015 IEEE Conference on, 2015
3) Image Caption Generation Using Deep Learning Technique by Chetan Amritkar and Vaishali Jabade.
4) Andrej Karpathy and Fei-Fei Li, "Deep visual-semantic alignments for generating image descriptions", Proceedings of the IEEE conference on computer vision and pattern recognition, 2015.
5) https://www.geeksforgeeks.org/python-introduction-matplotlib/
6) https://www.javatpoint.com/keras
7) https://ieeexplore.ieee.org/document/8276124
8) S. Pasupathy, “Image Caption Creator by using CNN and LSTM”, (IJFMR), E-ISSN: 2582-2160 Volume 5, Issue 2, March-April 2023.
9) Rita Ramos, Desmond Elliott and Bruno Martins, “Retrieval-augmented Image Captioning”, the 17th Conference Chapter pages 3666–3681 May 2-6, 2023.
10) G. Lakshmi Vara Prasad, B. Sri Mounika, P. Vijaybabu, A. Teethes babu and Ch.Srikanth, “Image Caption Generator Via CNN and LSTM”, 123(2022) 78-86 DOI: 10.26524/sajet.2022.12.42.