PROJECT TOPIC

Image Captioning

Abstract:

In a world increasingly saturated with visual content, the ability to automatically generate descriptive textual captions for images has become a pivotal technological frontier. This project is a comprehensive exploration of the intersection between Computer Vision (CV), Natural Language Processing (NLP), Large Language Models (LLM), Transformer architecture, and Generative AI, with the overarching goal of enabling machines to understand and eloquently describe the content of images.

The project unfolds through a series of meticulously orchestrated steps. The initial phase involves the collection and preprocessing of a diverse dataset containing images and their associated captions. Well-established datasets like COCO and Flickr30k provide the foundation for this endeavor, and image preprocessing and caption tokenization facilitate the subsequent modeling stages.

The heart of the project lies in the fusion of Computer Vision and Natural Language Processing. A Convolutional Neural Network (CNN) is trained to extract intricate features from images, which are then combined with the power of recurrent neural networks (RNNs) or Transformer-based models to generate rich textual descriptions. The use of pre-trained word embeddings and attention mechanisms ensures the synergy of visual and textual information.

The project places special emphasis on the integration of Large Language Models (LLM), such as GPT-3, into the captioning process, allowing for enhanced contextual understanding and more expressive captions. Transformer architecture, known for its proficiency in handling sequential data, is also employed to further elevate the model's performance.

As technology continues to shape the way we interact with visual content, this project stands at the forefront, not only offering innovative solutions to accessibility and content indexing but also enhancing the user experience in applications driven by imagery. With continuous improvement and fine-tuning driven by user feedback, this project embodies the ongoing evolution of AI-powered image captioning, promising to make visual content more accessible and understandable to all.

GROUP MEMBERS:

R. MOUNIKESWARI (21KN5A4205)

S.V. MADHAVA REDDY (20KN1A4255)

K. CHANDRA SESKHAR (20KN1A4223)

T. SANDEEP (20KN1A4260)