Image Captioning Project with Deep Learning

Objectives

The primary goal of this project is to develop an image captioning model that can generate descriptive captions for a given image. This involves combining computer vision techniques for image feature extraction with natural language processing (NLP) techniques for generating coherent and contextually relevant captions.

Methodology

Data Collection: I obtained a dataset containing images and corresponding captions.

Feature Extraction: I used a pre-trained ResNet-50 model to extract image features.

Caption Preprocessing: Captions were preprocessed, tokenized, and used to create a vocabulary.

Model Architecture: The model consists of an image embedding layer, an LSTM-based language model, and an output layer with softmax activation.

Training: The model was trained using paired image features and tokenized captions.

Evaluation: The quality of generated captions was evaluated using various metrics such as BLEU score.

Results
The model successfully learned to generate captions for images.
Caption quality was evaluated using metrics such as BLEU score.
Example images with generated captions were visually inspected for quality.

Fine-tuning the model with additional data.
Exploring different pre-trained models for image feature extraction.
Enhancing the quality of generated captions through advanced NLP techniques.