B.Tech Mini Project Report

Title: Image Caption Generator using Deep Learning

Student Name: Sachin Kumar Singh

Branch: Computer Science and Engineering

Semester: 8th

Project Type: Mini Project

Submitted to: [Instructor’s Name]

Academic Year: 2024–2025

# Abstract

This project aims to develop an Image Caption Generator using deep learning techniques. It involves automatically generating textual descriptions for images by integrating Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). This system bridges the gap between computer vision and natural language processing, allowing machines to "see" and "describe" images like humans.

# Objectives

* To create a model that can generate meaningful captions for input images.
* To explore deep learning frameworks such as TensorFlow and Keras.
* To apply CNNs for feature extraction and LSTM (a type of RNN) for sequence modeling.
* To understand the practical implementation of encoder-decoder architectures.

# Technologies Used

|  |  |
| --- | --- |
| Component | Description |
| Python | Programming Language |
| TensorFlow/Keras | Deep Learning Libraries |
| NumPy, Pandas | Data Handling & Preprocessing |
| Matplotlib | Visualization |
| NLTK | Text preprocessing |
| Jupyter Notebook | Development Environment |

# System Architecture

1. Image Feature Extraction (Encoder): A pre-trained CNN model (like InceptionV3) is used to extract high-level features from input images.

2. Text Sequence Generation (Decoder): An LSTM model processes the image features along with partial captions to predict the next word.

3. Training: The model is trained using paired image-caption datasets (e.g., Flickr8k/Flickr30k/MS-COCO). Tokenization, padding, and vocabulary construction are carried out.

4. Prediction: During inference, the model generates a caption word-by-word based on learned sequences.

# Results

Successfully trained a caption generator capable of describing unseen images.

Example Output:

🖼️ Image: A dog is playing with a ball in the grass.

📝 Caption: "A dog playing with a ball in a grassy field."

# Applications

* Assisting visually impaired individuals with scene understanding.
* Automated image tagging and classification.
* Enhancing content accessibility on the web and social media platforms.

# Learnings

* Hands-on experience with deep learning models and data preprocessing.
* Understanding the integration of vision and language models.
* Worked with real-world datasets and learned to fine-tune pre-trained models.

# Conclusion

This project provided in-depth exposure to deep learning and its capabilities in automating complex human tasks like image captioning. It also enhanced my programming, model tuning, and problem-solving skills. The Image Caption Generator project serves as a strong foundation for future explorations in AI.