



National Textile University NTU

Name	Muhammad Abdullah, Maryam Sameen, Muhammad Kaif
Reg no	22-NTU-CS-1358, 22-NTU-CS-1354, 22-NTU-CS-1364
Class and Section	BS-AI 6 th
Course Code and Title	AIC-3072L & Machine Learning
Submitted To	Miss Kainat Abdullah
Date of Submission	05-06-2025

Image Caption Generator

1. Introduction

This project focuses on generating descriptive captions for images using deep learning techniques. It combines computer vision (CNNs) and natural language processing (LSTMs) to build an end-to-end system that can describe images in human-readable text. A user-friendly Flask-based web interface allows users to interact with the model by uploading images and receiving predicted captions.

2. Problem Statement

Understanding and interpreting the contents of an image is a challenging task for machines. The objective of this project is to create an AI system that can look at an image and describe it accurately using a natural language caption.

3. Dataset Description

The dataset consists of a set of images with corresponding textual descriptions stored in a file named 'image_captions.txt'. Image features were pre-extracted using the InceptionV3 model and saved in 'image_features.pkl'. Tokenization and sequence padding were applied to prepare the textual data.

4. Model Architecture

The model is composed of a CNN encoder and an LSTM decoder. The encoder, based on InceptionV3, extracts feature vectors from images. These vectors are passed to the decoder, which uses an LSTM network to generate corresponding textual descriptions. The architecture also includes embedding layers and sequence padding.

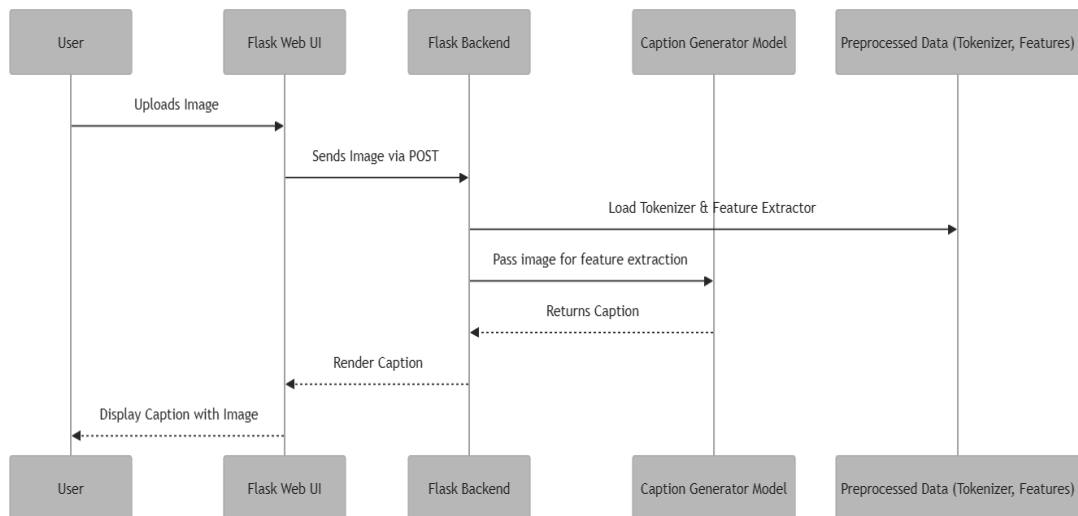
5. Tools and Technologies

- Python
- TensorFlow/Keras
- NumPy and Pillow
- Flask (for web deployment)
- Jupyter Notebooks (for training)

6. Flask Web Application

A web application was built using Flask (`app.py`) to serve the trained model. Users can upload an image via an HTML form, and the server returns a generated caption. CSS is used for styling, and Jinja2 templating dynamically renders predictions and previews.

7. Sequence Diagram



8. User Interface

