

Project Design Phase-I
Solution Architecture

Date	19 September 2022
Team ID	PNT2022TMID591920
Project Name	Image Caption Generation
Maximum Marks	4 Marks

Image Caption Generator Solution Architecture:

1. User Interface Module:

- **Description:** User interacts with the system through a friendly interface to upload images.
- **Components:**
 - Upload button.
 - User input validation.

2. Preprocessing Module:

- **Description:** Prepare images and captions for feature extraction and training.
- **Components:**
 - Image cleaning and normalization.
 - Caption text processing (lowercasing, punctuation removal, and filtering).

3. Dataset Module:

- **Description:** Utilize the Flickr8K dataset for training and evaluation.
- **Components:**
 - Large collection of images with descriptive captions.
 - Preprocessed dataset for model input.

4. Feature Extraction Module:

- **Description:** Extract meaningful features from input images.
- **Components:**
 - VGG16 deep CNN model.
 - Removal of the last classification layer for fixed-length vector representation.

5. Caption Preprocessing Module:

- **Description:** Process textual captions for LSTM-based caption generation.
- **Components:**
 - Vocabulary creation.
 - Tokenization of words.
 - Word-to-index mapping.
 - One-hot encoding.

6. Model Architecture Module:

- **Description:** Core module for image caption generation.
- **Components:**
 - LSTM network as a sequence model for word-by-word caption generation.
 - Dense layers for mapping extracted image features to captions.

7. Training Module:

- **Description:** Train the model using prepared dataset.
- **Components:**
 - Optimization of model parameters using the Adam optimizer.

- Minimization of categorical cross-entropy loss.
8. **Evaluation Module:**
 - **Description:** Assess model performance using BLEU scores.
 - **Components:**
 - Comparison with reference captions from the Flickr8K dataset.
 9. **Deployment Module:**
 - **Description:** Deploy the trained model for real-time caption generation.
 - **Components:**
 - Integration with IBM Watson for deployment.
 - API endpoints for communication between the user interface and the deployed model.
 10. **User Output Module:**
 - **Description:** Present the generated captions to users.
 - **Components:**
 - User-friendly interface displaying the captions.
 - Option to download captioned images.

Example - Solution Architecture Diagram:

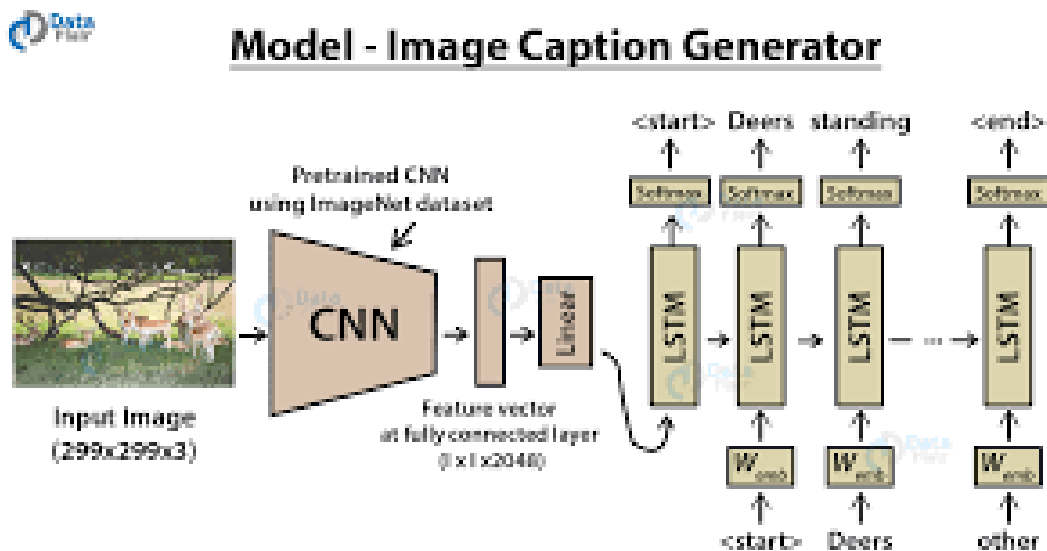


Figure 1: Architecture and data flow of the modal image caption generator