Image Captioning Using Deep Learning

An overview of concepts, methodology, and implementation

INTRODUCTION

- Image captioning is the process of generating textual descriptions for images.

- It combines techniques from computer vision and natural language processing.

- This presentation explores the theory and practical implementation of image captioning models.

How Image Captioning Works :

1. A Convolutional Neural Network (CNN) extracts visual features from the image.

2. A Recurrent Neural Network (RNN), typically an LSTM, generates a caption word by word.

3. The system is trained on image-caption datasets to learn context and structure

Overview of Workflow :

- • Load Flicker8K dataset and clean caption txt.

- • Extracted image features using InceptionV3.

- • Prepared input-output sequences for training.

- • Saved processed data for model training.

Data Preparation

- Load image-caption datasets (e.g., Flickr8k).
- Clean and normalize captions (remove punctuation, lowercase, tokenize).
- Extract image features using pre-trained models like InceptionV3.

Model Architecture

- Encoder: CNN (e.g., InceptionV3) extracts image features.
- Decoder: LSTM receives image features and generates caption tokens.
- Combined with embedding layers and dense layers for predictions.

Training the Model

- Input: Image features and partial captions.
- Output: Next word in the caption sequence.
- Loss: Categorical crossentropy.
- Optimizer: Adam or RMSprop.
- Evaluate with BLEU or METEOR scores.

Model Prediction

Caption: while a large little and man in the grass in the grass in the tree man in the in the tree grass by a set by a set by a set by a set by

Image caption Generator Tool

Challenges Faced

- **Data Loading Issues**
Faced problems while loading the dataset and managing file paths.
- **Slow Training (Epochs)**
Training the model took a long time and often slowed down or crashed.
- **Deployment Difficulties**
Struggled to turn the trained model into a working web or app-based tool.
- **Accuracy and Output Quality**
Some generated captions didn't match the image well or made little sense.

Applications

- Assistive technologies for the visually impaired.

- Automated image tagging and content moderation.

- Descriptive image search engines.

- Enhancing accessibility in digital media.

Business Benefits

- **Reduces Manual Work**
  Automatically adds captions to thousands of images, saving time and effort.

- **Improves User Accessibility**
  Makes platforms more inclusive for users with visual impairments, meeting legal standards.

- **Boosts Search Visibility**
  Helps images appear in search results, attracting more users and increasing engagement.

CONCLUSION

1) Image captioning bridges visual and linguistic understanding.

2) Combining CNNs and RNNs enables machines to describe images effectively.

3) Future work involves improving accuracy, reducing bias, and real-time inference.

Thank You

PRESENTATION BY –
Aparna Tiwari
Sufiyan Zamindar
Vignesh Thingalaya
Sonal Vishwakarma