Image Captioning and Segmentation Project

Page 1: Project Overview

Introduction

This project focuses on two key tasks in computer vision: image captioning and image segmentation. Image captioning generates descriptive captions for input images using deep learning models, while segmentation involves identifying object boundaries and labeling regions in an image.

Abstract

We used CNN-LSTM architecture for image captioning and U-Net for segmentation. MS COCO and Pascal VOC datasets were used for training. The final system integrates both models to process an image and produce its segmented view along with a meaningful caption.

Tools Used

- Python
- TensorFlow / PyTorch
- OpenCV
- NLTK / spaCy
- Flask / Streamlit
- Jupyter Notebook

Image Captioning and Segmentation Project

Page 2: Technical Steps & Results

Steps Involved

1. Load and preprocess datasets (MS COCO / Pascal VOC)
2. Train CNN + LSTM captioning model
3. Train segmentation model (U-Net or Mask R-CNN)
4. Integrate captioning and segmentation into one pipeline
5. Evaluate using BLEU (for captions) and IoU (for masks)
6. Build and test UI using Streamlit or Flask

Results & Metrics

Sample caption:

"A group of people standing on a beach."

Sample segmentation:

Correctly identifies and masks 'person', 'sea', and 'sky' in the input image.

Metrics Achieved:

- BLEU Score: 0.67
- IoU Score: 0.61