Video Summarization and Image Caption Generator

SYNOPSIS SUBMITTED TO ASIAN SCHOOL OF MEDIA STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

AWARD OF DEGREE OF

M.Sc.

in

Data Science

By

Rahul Pathak

(University Enroll. No: 000000000000)

Under the Supervision of

Prof. Gaurav Kumar

ASIAN SCHOOL OF MEDIA STUDIES

NOIDA

2023

Problem Statement

The increasing volume of multimedia content, particularly videos and images, makes it challenging to extract key information efficiently. A system that can summarize videos and generate captions for images would aid in quick understanding and accessibility.

Motivation

The rapid increase in digital content necessitates efficient ways to summarize and describe multimedia files. Manual annotation is time-consuming and inefficient, highlighting the need for automated solutions. This project will address challenges in video content consumption and image accessibility, benefiting industries like media, entertainment, and education.

Problem Formulation/Objectives

- Develop an efficient method for extracting key frames from videos.
- Implement an automated image captioning model using deep learning.
- Improve accuracy and relevance of generated captions and summaries.
- Optimize computational efficiency for real-time applications.

Methodology/Planning of Work

1. Data collection and preprocessing (image and video datasets)
2. Implementation of deep learning models (CNNs, RNNs, Transformers)
3. Training and fine-tuning of models
4. Evaluation using standard metrics
5. Optimization for real-time performance
6. Testing and validation
7. Deployment and user testing

Facilities Required for Proposed Work

- Software: Python, TensorFlow/PyTorch, OpenCV
- Hardware: High-performance for training deep learning models

-

Project Outcomes

- A functional model for automatic video summarization
- An image caption generator capable of describing images accurately
- Enhanced accessibility for multimedia content
- Potential applications in various domains such as media, education, and assistive technology

Abstract

This project aims to develop a system for video summarization and image caption generation using deep learning techniques. Video summarization will extract the most relevant frames to provide a concise version of the original video, while image caption generation will involve automatic description of images. The implementation will utilize machine learning models such as CNNs, RNNs, and transformer-based architectures. The project will enhance content accessibility, improve searchability, and support applications in media, education, and assistive technologies.

Thank You