Video Caption Website with ChatGPT Help!!!

Hao Lu

lh2002@buaa.edu.cn

Introduction

We have built a website for video captioning, where users can upload a video and receive a caption in return. All code design and backend processing are handled by a large language model.

Methodology

Step 1: We instructed ChatGPT to use HTML and JavaScript for the frontend, and Flask for the backend, to build a Python-based web framework.

Step 2: We provided the HTML webpage to ChatGPT and asked it to enhance the appearance and user interface of the frontend (beautify the webpage).

Step 3: We instructed ChatGPT to provide a backend endpoint to handle the uploaded video file, allowing the server to process the video accordingly.

Step 4: We asked ChatGPT to write a function based on **Qwen's API call template**, which invokes the **qwen-vl-max** model to generate a caption for the uploaded video.

Step 5: Finally, we asked ChatGPT to generate a **structured and detailed prompt** to be provided to **Qwen-VL-Max**, enabling it to accurately analyze and caption the uploaded video content.

Table 1: Workflow for Video Captioning Web Framework with Qwen-VL-Max

Website Overview

In this section, we demonstrate the full functionality of our website. Upon uploading a video file, the system returns a JSONL file. This platform can be used for distillation of training data for video large models.

Structured Prompt by ChatGPT

请你以结构化方式总结该视频的主要内容，包括以下几个方面：

1. **视频类型**（例如：教学 / 讲座 / 访谈 / 会议 / 生活记录 / 其他）
2. **主要人物 / 说话人**（列出姓名或角色）
3. **核心内容概要**（用简洁段落概括视频讲了什么）
4. **关键信息要点**
   - 关键点 1
   - 关键点 2
   - 关键点 3
5. **情绪或氛围**（例如：轻松 / 严肃 / 紧张 / 欢快）
6. **适合受众**（该视频适合哪些人观看）

Figure 1: Website design

Figure 2: Upload a video

Figure 3: Success

Figure 4: Return output