PROJECT - 2

HUMAN ACTION ACTION RECOGNITION USING NVIDIA VLM WORKFLOW

TEAM

SUSRITH REDDY B - SE21 URI 164
PRABHAS T - SE21UARI 176
VARSHITH MOHAN G - SE21UARI 185
ABHINAV - SE21UARI 174

Objective

objective of this project is to develop an interface and evaluation system for comparing real and synthetic human action videos, leveraging NVIDIA's Vision Language Models (VLM) to analyze the effectiveness of recognition tasks on these datasets

INTRODUCTION

Human activity recognition is crucial in diverse applications such as surveillance, healthcare, sports analysis, and gaming. This project investigates the use of NVIDIA's Vision-Language Models (VLMs) for recognizing human activities in real and synthetic videos.

The primary goal is to evaluate the consistency and reliability of action detection algorithms across these two types of datasets. Synthetic data, being a scalable and cost-effective alternative to real data, holds great promise for training machine learning models. However, understanding the performance differences between detecting actions in synthetic and real videos is essential to ensure robust model generalization.

This project aims to provide a comprehensive tool for researchers to compare human activity detection rates across real and synthetic video datasets. Leveraging NVIDIA's NEVA API, the project includes a user-friendly interface for video uploading, trimming, and analysis. Through Gradio, it offers an intuitive platform for evaluating activity detection performance.
The insights derived from this project will help guide the development of more efficient and reliable action recognition models, enhancing their applicability across a range of domains.

1. Setup Instructions Prerequisites:

a.Install Python (version 3.8 or later is recommended).
b. Install the necessary Python libraries by running the following command in your terminal: pip install -r requirements.txt c.

Ensure that ffmpeg is installed for video processing. You can install it using the following commands based on your operating system:
For Linux: sudo apt update sudo apt install ffmpeg
For macOS: brew install ffmpeg

2. Setup Instructions
a. Clone the repository from GitHub:
https://github.com/Theerthankasusrith/NVIDIA-VLM

b. Launch the application:
python app.py

We can also use another method
This is using the Colab notebook for running the code
The Colab link:
https://colab.research.google.com/drive/1ii-prkkXkpNdvUeVVm-YnxeM7v3vAXl2?usp=sharing

How to use:

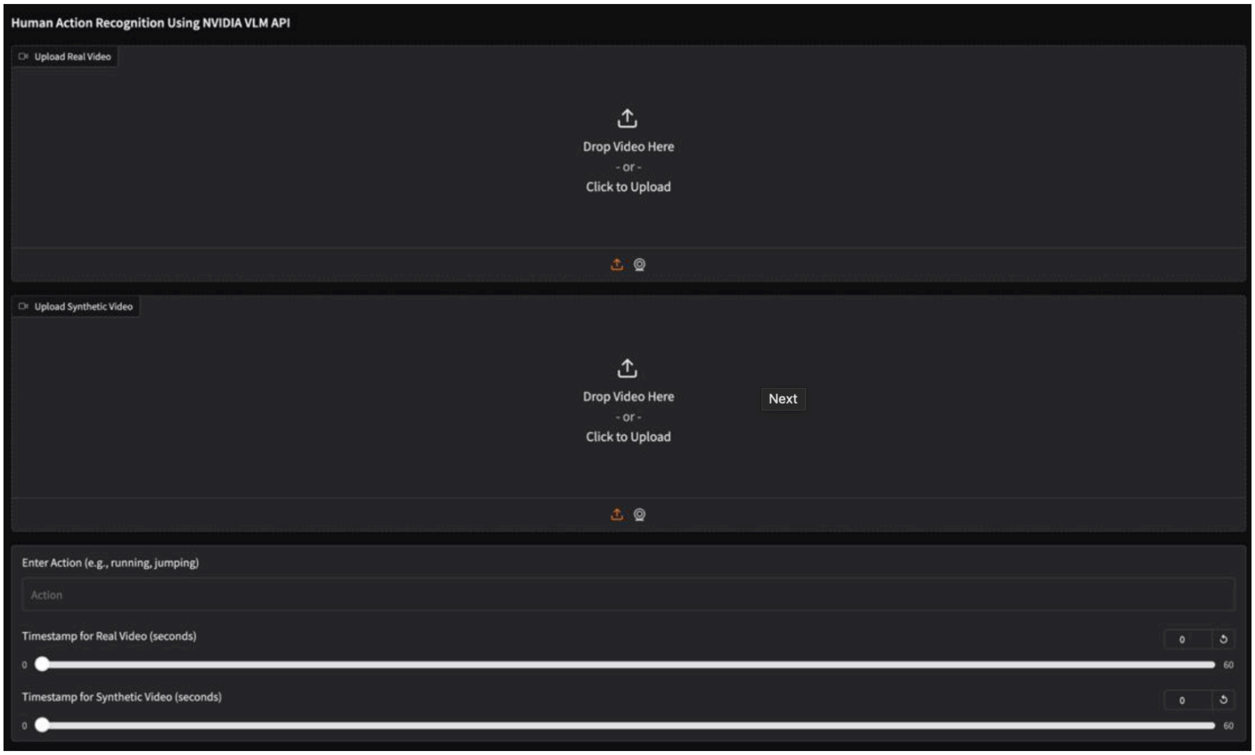
1. Launching the Interface: a. After running app.py, a local Gradio interface will open in your browser.
2. Uploading Videos: a. Upload one synthetic video (Video A) and one real video (Video B).
b. Supported formats: .mp4, .avi, .mov. 3.

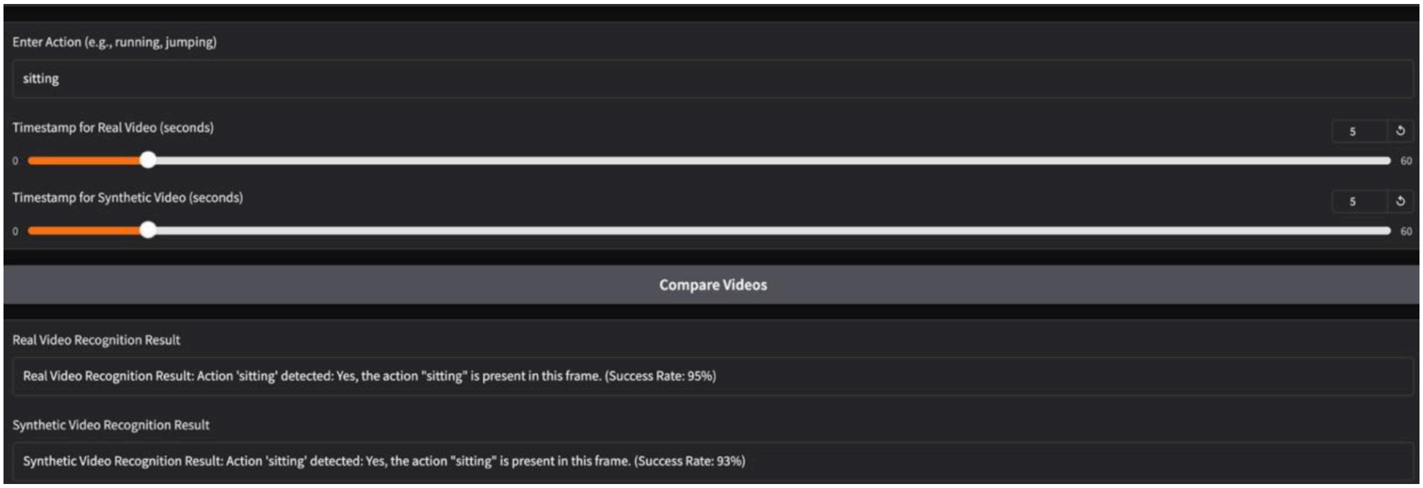
3.Specify an Activity: Enter the activity you want to detect
(e.g., walking, jumping,running).
4. Trimming (Optional):
Enter a trim length (in seconds). If left blank, videos will be trimmed to the duration of the shorter video.
5. Results: Click “Submit” to get the success rates for both videos. b. View results in percentage format.

Outputs

Sample Output:
Synthetic Video - sitting Success Rate: 93.0%
Real Video - sitting Success Rate: 95.00%

RESULTS -


Human Action Recognition Using NVIDIA VLM API
Upload Real Video
Drop Video Here
- or -
Click to Upload
Upload Synthetic Video
Drop Video Here
- or -
Click to Upload
Next
Enter Action (e.g., running, jumping)
Action
Timestamp for Real Video (seconds)
0
0
60
Timestamp for Synthetic Video (seconds)
0
0
60



Enter Action (e.g., running, jumping)

sitting

Timestamp for Real Video (seconds)

5

Timestamp for Synthetic Video (seconds)

5

Real Video Recognition Result

Real Video Recognition Result: Action 'sitting' detected: Yes, the action "sitting" is present in this frame. (Success Rate: 95%)

Synthetic Video Recognition Result

Synthetic Video Recognition Result: Action 'sitting' detected: Yes, the action "sitting" is present in this frame. (Success Rate: 93%)