

Machine Vision Project

Surveillance Threat Detection

Course : CECS - 553

Name : Ankit Raj

CSULB ID : 030824350


Problem Statement

Surveillance systems today can't fully understand both what they see and read, making it hard to warn about different kinds of dangers.

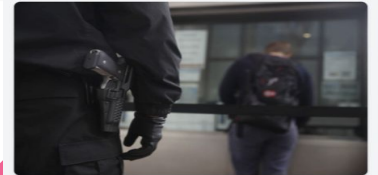
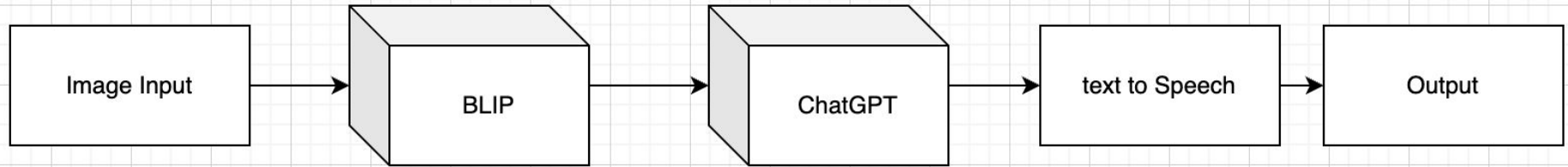
We really need smarter tech that can accurately figure out what's happening in surveillance videos and texts to give better warnings about possible threats right away.



Objective

- The project's objective is to develop an advanced surveillance system that can comprehensively analyze visual and textual data from surveillance camera feeds to detect various types of threatening situations in real-time. This system will utilize state-of-the-art technologies, such as a **Generative Vision-Language Model (BLIP)** for image interpretation and **ChatGPT** for text analysis.
 - By integrating these models, the project aims to provide timely and accurate threat alert announcements, enhancing security measures and enabling proactive responses to potential threats in diverse environments.
- 

System Architecture




****Security Announcement:**** A surveillance image captures a police officer standing outside a building, with a gun in hand. Please note that the officer appears to be in a position of readiness. The situation is being monitored closely for any potential security threats. Thank you for your attention.

Technology Used

Artificial Intelligence Model:

- Generative Vision-Language Model (BLIP) : To understand the image and generate text to brief the image.
- Conversational AI (ChatGPT) : To understand the textual information generated from the image and understand the context to generate alert message and briefing the captured image information.

Web Application Tool:

- NextJs for Frontend creation.
 - Flask for Backend creation.
- 

Project Development

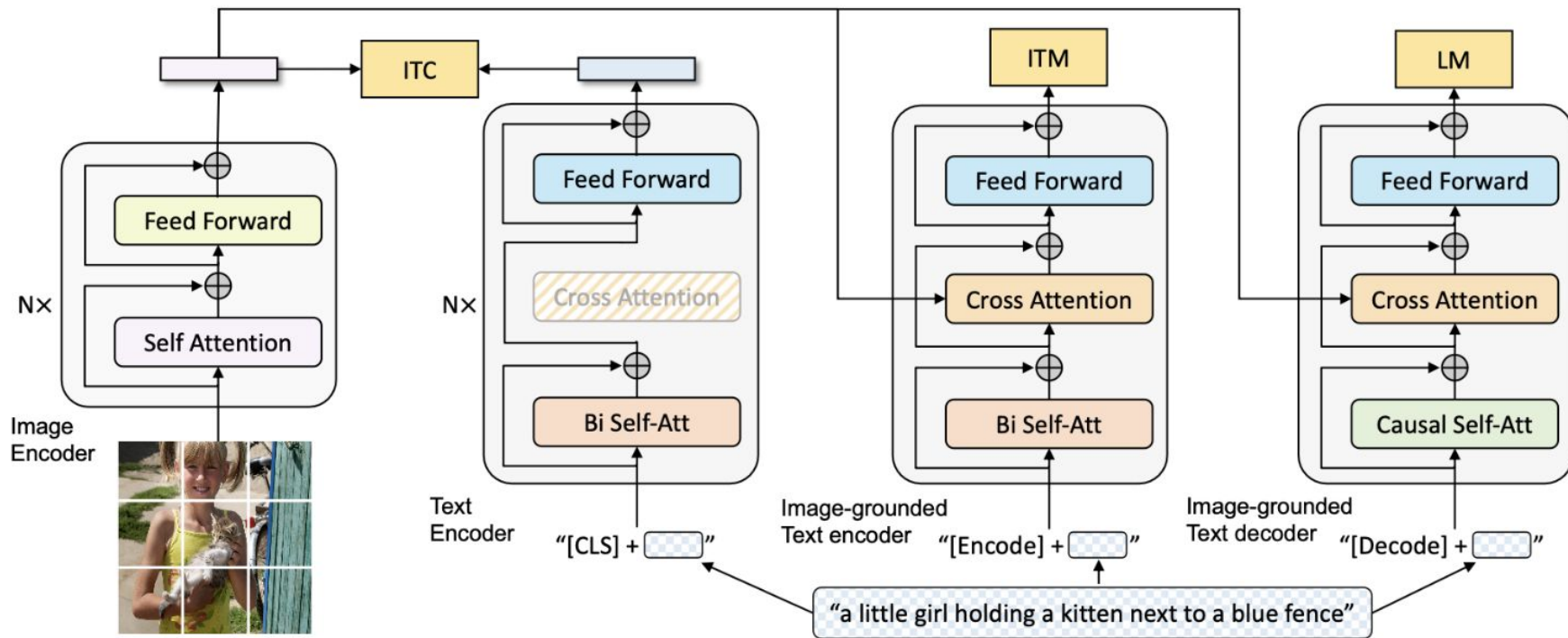
Phase 1: Image Understanding and Generating information in text

- **Using Generative Vision-Language Model (BLIP)**

BLIP, a new VLP framework with state- of-the-art performance on a wide range of downstream vision-language tasks, including understanding-based and generation-based tasks. **Capabilities of BLIP is leverage here to get description of the image and generate textual information.**



BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation



Phase 2: Text Understanding and Analysis

- **Using ChatGPT 3.5**

The generated text from BLIP is passed to ChatGPT, which possesses natural language understanding capabilities. ChatGPT analyzes the text to comprehend the context, identify key information, and extract relevant details about the potential threat.



Phase 3: Text to Speech Conversion

- **Using gTTS python library**

gTTS stands for "Google Text-to-Speech." It's a Python library and CLI tool to interface with Google Translate's text-to-speech API. With gTTS, we converted inference generated by chatGPT to speech using english languages and accents.




Result

BLIP understand the image and generate text.


chatGPT 3.5 Model analyze and understand the text with its context and generate the Security Alert Message.







Upload an Image

Choose File  gunman2.jpeg

Upload

Selected Image Preview:



   SECURITY ALERT    A potentially life-threatening situation has been captured on surveillance cameras. A man wearing a hoodie has been spotted holding a gun in a dark area. Please exercise extreme caution and report any suspicious activity to the authorities immediately. Your safety is our top priority. Stay vigilant and stay safe.

Upload an Image

Choose File  police.jpeg

Upload

Selected Image Preview:



****Security Announcement:**** A surveillance image captures a police officer standing outside a building, with a gun in hand. Please note that the officer appears to be in a position of readiness. The situation is being monitored closely for any potential security threats. Thank you for your attention.

Project Application

- Integrate with Surveillance Camera to make it smart and leverage the analyzing and reasoning power of LLM to detect life-threatening situations and contextual information.
- This approach enhances public place surveillance systems by offering a proactive means of **identifying diverse threats**, thereby enabling prompt responses to mitigate risks and ensure public safety.



Choose File

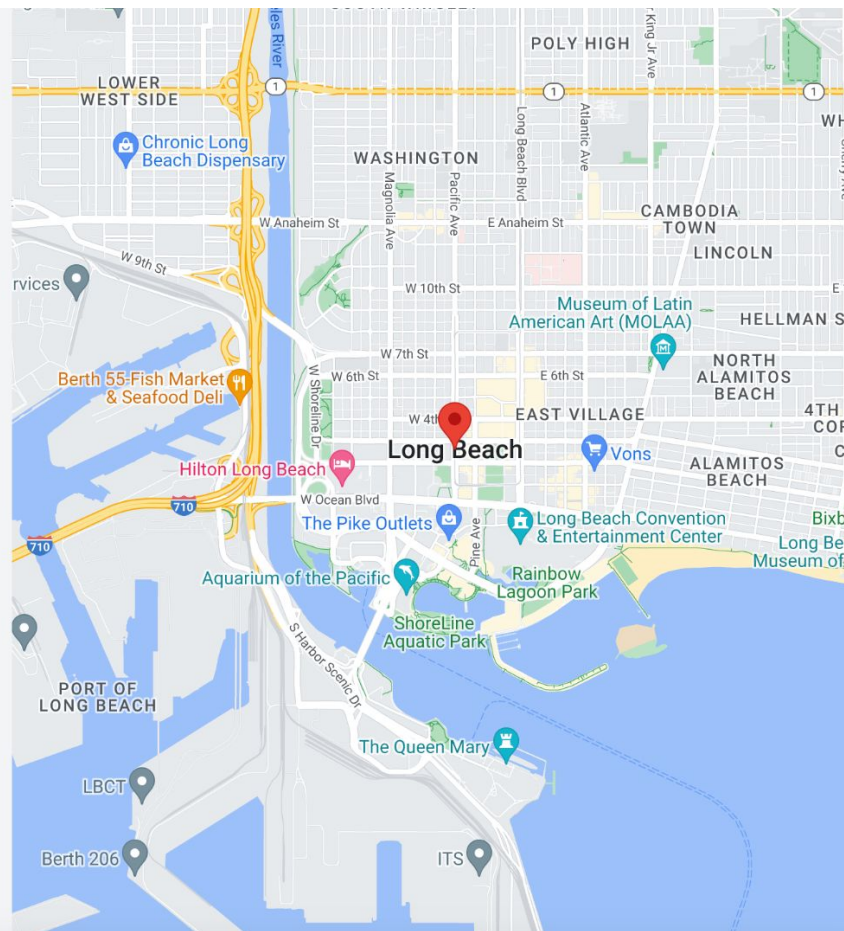
theif3.jpeg

Upload

Selected Image Preview:



SECURITY ALERT A man wearing a black mask was captured on surveillance holding a knife in his hand. This behavior poses a serious threat to public safety and law enforcement authorities have been notified. Please remain vigilant and report any suspicious activity to the authorities immediately. Your safety is our top priority.



Future Work

- Fine Tune the Visual Model to give more relevant information from the image.
- Extend the application of chatGPT to provide additional suggestion to control the situation.



Thank you !!

