Name: Ankit Raj | Csulb ID: 030824350

Surveillance Threat Detection

Machine Vision Project

CECS 553

1. Introduction

This project is designed to enhance surveillance systems by automatically detecting life-threatening situations in camera photos. It leverages cutting-edge technology, primarily a Generative Vision-Language Model (BLIP), to interpret the content of images captured by surveillance cameras. BLIP excels at understanding visual content, allowing it to analyze the photos and identify potential threats or dangerous situations accurately.

In conjunction with BLIP**,** the project integrates ChatGPT, a powerful conversational AI model, to analyze the text generated based on the interpreted images. ChatGPT's natural language processing capabilities enable it to understand and contextualize the information extracted from the images. By combining both models, the system can effectively process visual and textual data to generate threat alert announcements in real-time.

2. Problem Statement

In today's complex security landscape, traditional surveillance systems often struggle to effectively detect and respond to various life-threatening situations beyond the identification of specific weapons like guns or knives. There is a critical need for advanced technologies that can accurately interpret surveillance camera images and understand the context of potential threats in real-time.

Current approaches lack the capability to comprehensively analyze visual and textual data, hindering the ability to provide timely threat alert announcements. As a result, there is a pressing need for a sophisticated solution that leverages cutting-edge Generative Vision-Language Models (such as BLIP) and conversational AI (such as ChatGPT) to enhance threat detection capabilities. By integrating these technologies, the project aims to address these challenges and provide a proactive and efficient means of detecting diverse life-threatening situations in surveillance camera photos, thereby enhancing security measures and ensuring public safety.

Key aspects of Surveillance Threat Detection may include:

- **Monitoring**: Constant observation of surveillance camera feeds to detect unusual or threatening behavior in real-time.
- **Analysis**: Analyzing video footage and sensor data to identify patterns, anomalies, or potential threats.
- **Alerting**: Generating alerts or notifications when suspicious activity is detected, enabling security personnel to respond promptly.
- **Response**: Taking appropriate action in response to identified threats, such as deploying security personnel, notifying law enforcement, or activating emergency protocols.
- **Prevention**: Implementing measures to deter potential threats and enhance security, such as access control systems, perimeter fencing, or security patrols.

3. Objective

This project aims to develop an advanced surveillance system that can detect a wide range of potential threats to human life, going beyond just the identification of specific weapons like guns or knives.

The key aspects of this project are:

- Utilization of a Generative Vision-Language Model (BLIP):The system uses BLIP, a powerful AI model that can interpret and understand the contents of images, to analyze the visual data from surveillance cameras.
- Integration with ChatGPT: The project combines the image analysis capabilities of BLIP with the natural language processing and generation abilities of ChatGPT 3.5. This allows the system to not only identify potential threats in the images but also generate real-time text-based alerts and notifications about the detected threats.
- Proactive Threat Detection: The goal is to create a proactive security system that can identify potential threats to human life, beyond just the presence of specific weapons. This could include detecting suspicious behaviors, unusual objects, or other indicators of potentially dangerous situations.
- Enhanced Security and Responsiveness: By combining advanced computer vision and language AI, the system aims to provide a more efficient and

effective way of identifying and responding to threats in real-time. This could lead to faster response times and improved security measures in various environments, such as public spaces, buildings, or critical infrastructure.

Overall, this project represents a significant advancement in the field of intelligent surveillance systems, leveraging the latest AI technologies to enhance security and protect human life. The integration of BLIP and ChatGPT is a novel approach that could set a new standard for proactive threat detection and response in surveillance

4. Project Development

The modular development approach of this project offers several benefits, primarily flexibility and ease of maintenance. By breaking down the project into separate modules, each module can be tested independently on desired inputs, ensuring its functionality and reliability. Additionally, the loosely coupled nature of the modules allows for easier integration, enabling developers to combine them seamlessly to create the final project.

This modular architecture offers flexibility, as changes or updates to one module can be made without affecting the others. This minimizes the risk of unintended consequences and simplifies maintenance efforts. Moreover, it facilitates scalability, as new modules can be added or existing ones modified as needed to accommodate evolving requirements or incorporate new features.

4.1. Technology used in this project

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 an alert message and briefing the captured image information.

Web Application Tool:

- NextJs for Frontend creation.

- Flask for Backend creation.

4.2 Development Phases

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. We leverage the capabilities of BLIP to get a description of the image and generate textual information.

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.

5. Result Analysis

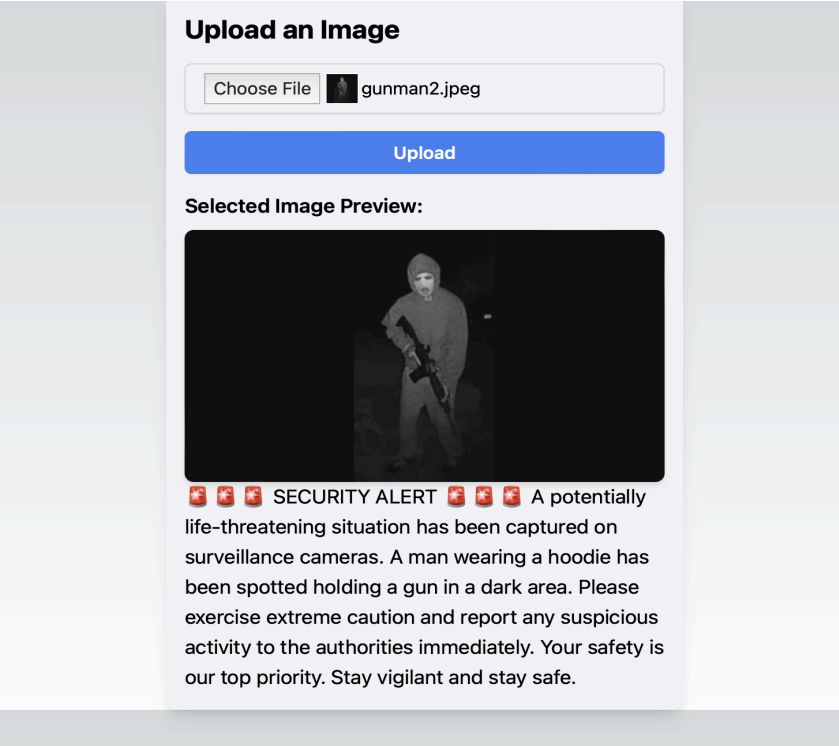

Fig 5.1

Above Fig 5.1 shows that the system generated a Security Alert message and describes the image to identify the suspect. This smart system could play a significant role in monitoring and enhancing public safety.

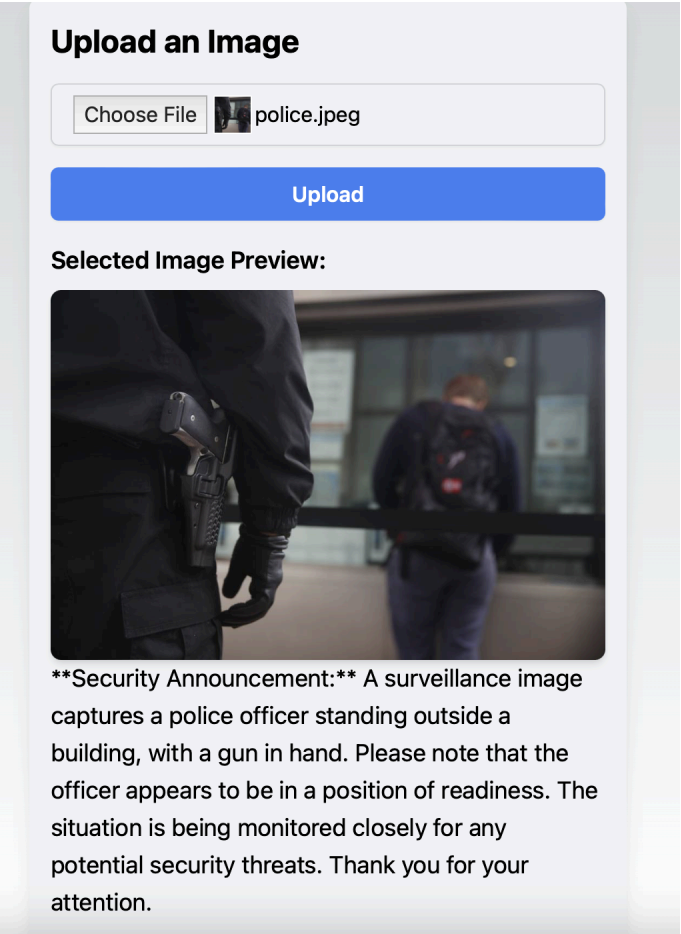

Fig 5.2

Above Fig 5.2 shows that the system creates a security announcement as the surveillance image captures a police officer standing in readiness with a gun. With the understanding capability of LLM Model, the system generates a security announcement , there could be potential threat.

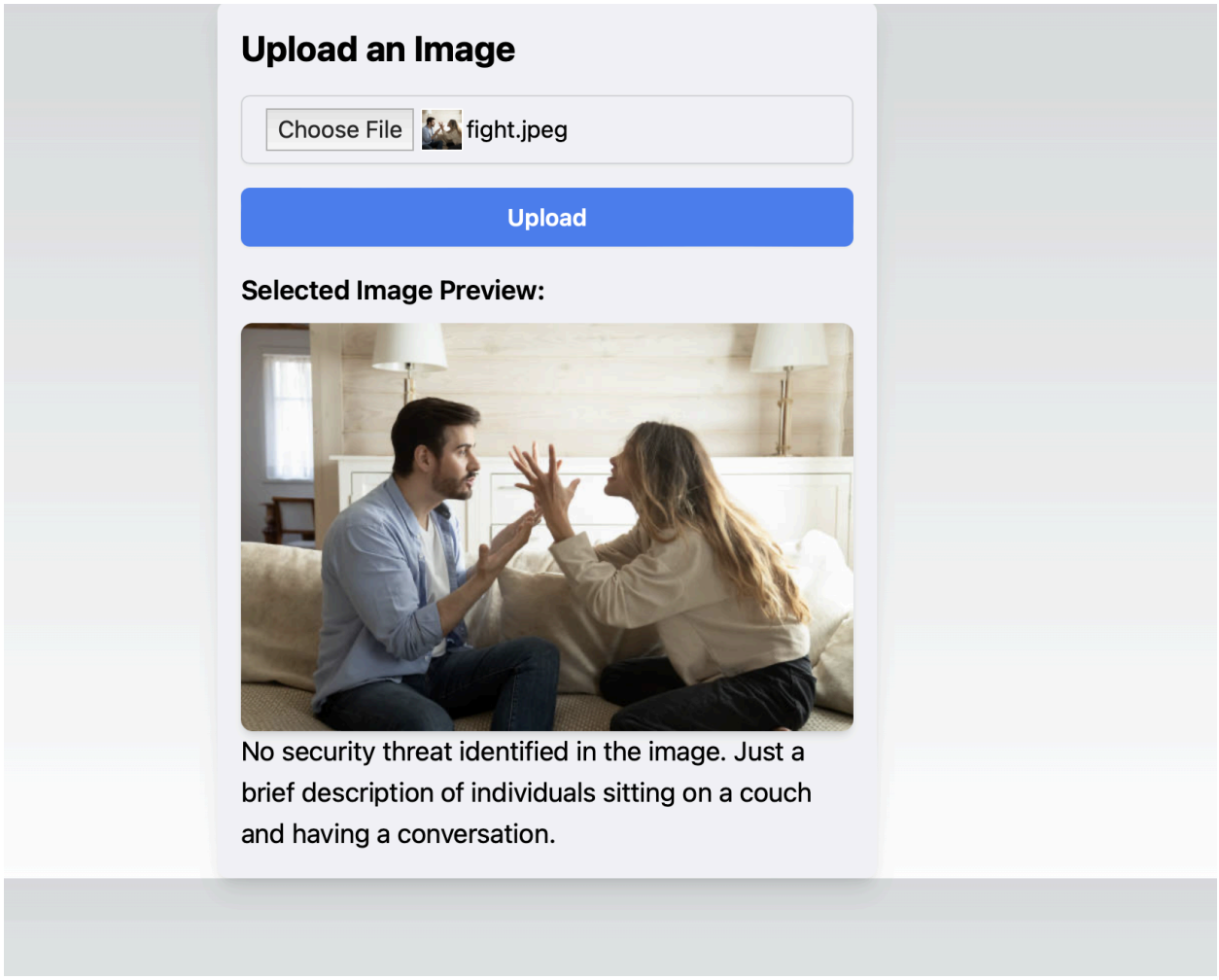

Fig 5.3

In the above figure 5.3, the system is intelligent enough to identify the two individuals having conversation so it wasn't classified as a security threat.

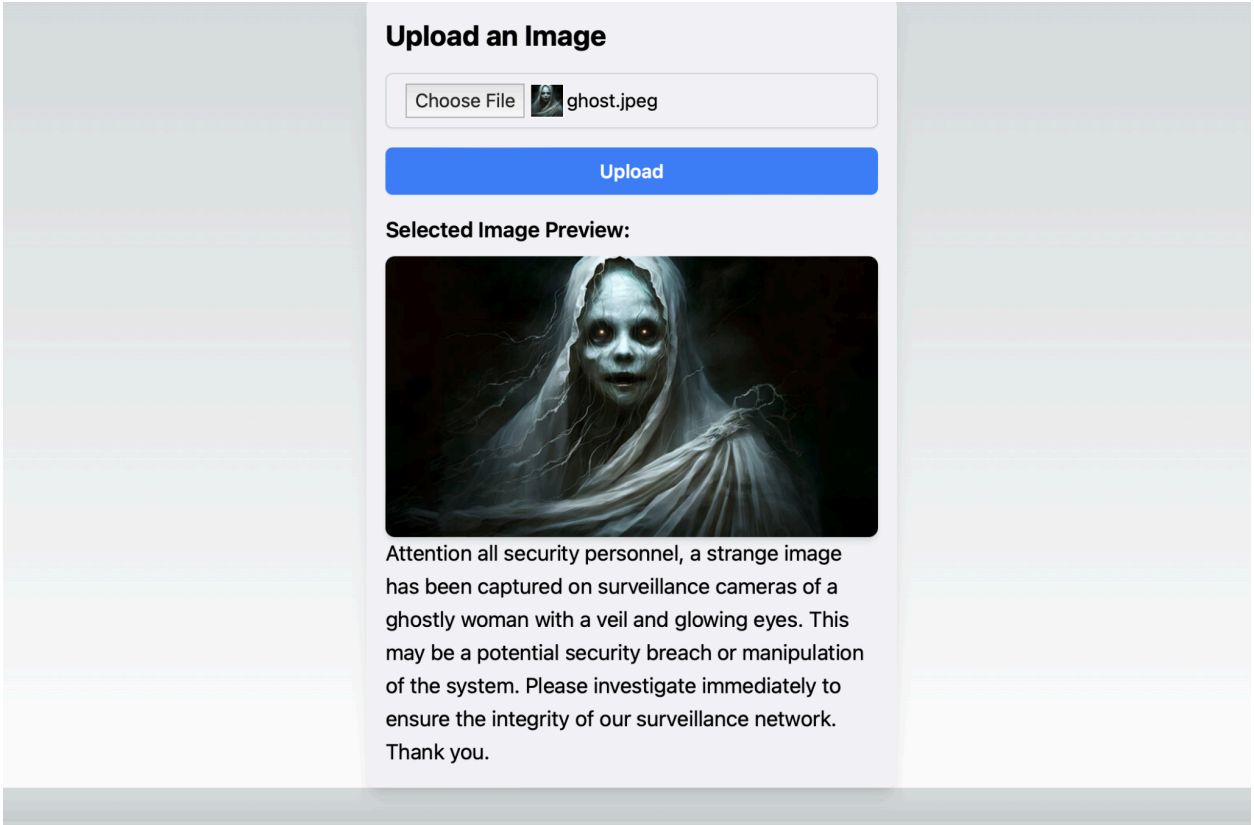

Fig 5.4

In the above fig 5.4, the system recognised it as an unusual behavior and identified it as a potential security breach or surveillance system manipulation.

6. Applications

- **Public Safety**: Deployment in public spaces such as airports, train stations, stadiums, and shopping malls to detect and respond to threats, ensuring the safety of individuals.
- **Law Enforcement**: Integration into law enforcement agencies' surveillance systems to assist in crime prevention, suspect tracking, and emergency response.
- **Commercial Facilities**: Utilization in commercial establishments such as banks, corporate offices, and manufacturing plants to enhance security measures and protect assets.

- **Healthcare Facilities**: Deployment in hospitals and healthcare facilities to detect and respond to security breaches, patient safety incidents, or medical emergencies.

7. Future Work

1. Integrate with Surveillance Camera to detect the human life threat situation:
   - Establish a connection between the system and the surveillance cameras to receive real-time video/image data.
   - Implement a module to continuously process the incoming visual data and analyze it for potential threats.
2. Fine-Tune the Visual Model to give more relevant information from the image:
   - Utilize transfer learning or fine-tuning techniques to further train the Generative Vision-Language Model (BLIP) on a dataset of images depicting various threat scenarios.
   - This will help the model become more accurate and specific in identifying potential threats, beyond just the presence of weapons.
3. Extend the application of ChatGPT to provide additional suggestions to control the situation:
   - Integrate the ChatGPT language model to not only generate alert messages but also provide suggested courses of action to mitigate the detected threats.
   - This could include recommendations for security personnel, emergency responders, or even automated systems to help control and de-escalate the situation.

8. References

- https://huggingface.co/models
- https://nextjs.org/docs?utm_source=create-next-app&utm_medium=appdir-template&utm_campaign=create-next-app
- https://flask.palletsprojects.com/en/3.0.x/
- https://stackoverflow.com/
- https://platform.openai.com/docs/guides/fine-tuning