Business Report: Video Analysis for Violence Detection Using YOLOv8 and BLIP-2 with Question-Based Frame Retrieval

1. Executive Summary

This project, completed within 3 weeks, developed an advanced violence detection system leveraging real-time machine learning models. It integrates live-stream data processing through the YOLOv8 model, which generates predictions of potential violent activities. These predictions are further refined by the BLIP-2 model to generate captions, while the BART model summarizes the results into comprehensive reports. Additionally, the system incorporates a Question-Answering (QA) feature that allows users to input questions related to the video. The system processes these queries using a SentenceTransformer model to generate embeddings for both the captions of the video frames and the user's query. The cosine similarity between the query embedding and the caption embeddings is calculated to retrieve the most relevant frame, enabling efficient analysis and visualization of specific events.

The system is fully integrated with Firebase for real-time data storage and features a Flutter-based mobile application that allows for easy access to reports and user interactions. The project was completed on time and within budget, successfully meeting all initial objectives.

2. Project Objectives

The main objectives of the project were to:

- Detect Violence in Real-time: Implement YOLOv8 for violence detection from live-stream video frames.

- Automate Captioning and Reporting: Utilize BLIP-2 and BART models to automatically generate captions and summarized reports based on detected events.

- Enable Question-based Frame Retrieval: Implement a QA system using a SentenceTransformer model to retrieve the most relevant video frames based on user queries.

- Provide Accessible Data Storage and Query System: Store the data in Firebase for easy querying and access through a mobile Flutter application.

3. Progress Report

The project was completed ahead of schedule within 3 weeks, demonstrating a high level of team efficiency:

- YOLOv8 Integration: Successfully integrated YOLOv8 for real-time analysis of video frames to detect violent activity.

- BLIP-2 and BART Models: Deployed for caption generation and automated reporting of detected events.

- Firebase Setup: Reports, live frames, and queries are stored and managed in Firebase, enabling real-time data access.

- Flutter Application: A mobile app was developed to allow users to interact with the system, retrieve reports, and input queries for frame retrieval.

- QA System: The team implemented a QA system using the SentenceTransformer model (paraphrase-MiniLM-L6-v2) for embedding generation. The model calculates the cosine similarity between the user's query and the frame captions to retrieve the most relevant frames.

4. Budget and Financials

The project was completed within the allocated budget:

- Hardware Costs: Standard laptops and cameras were used for local testing (fixed cost).- Cloud Storage: Firebase costs remained within projected usage.

- Cloud Storage and Services: Firebase's cloud services were utilized within the expected storage and bandwidth limits.

- Development Costs: No additional development costs were incurred due to the efficient completion within the 3-week timeline.

5. Challenges and Risks

Key challenges encountered during the project included:

- Real-time Performance Optimization: One of the significant challenges was ensuring that the YOLOv8, BLIP-2, and BART models worked seamlessly in real-time.

- False Positives: Initially, there were concerns about the accuracy of violence detection, especially with false positives. Tuning the models reduced this risk.

- Data Privacy: Storing sensitive video content and the associated reports in Firebase required rigorous security protocols, which were implemented via encryption.

6. Team and Resources

The team consisted of:

- Data Scientist: Handled model selection, fine-tuning, and training.

- Backend Developer: Integrated Firebase and managed report generation.

- Frontend Developer: Developed the Flutter application.

- DevOps Specialist: Ensured system scalability and cloud integration.

7. Timeline and Milestones

The project was completed within 3 weeks with the following key milestones:

- Week 1: Integrated YOLOv8 for violence detection and initial tests on real-time video streams.

- Week 2: Deployed the BLIP-2 and BART models for captioning and report generation, and integrated Firebase.

- Week 3: Completed the development of the Flutter application and implemented the SentenceTransformer-based QA system for frame retrieval based on queries.

8. Stakeholder Updates

Weekly updates were provided to stakeholders, covering the progress of model integration, report generation, and mobile application development. Feedback from stakeholders regarding the user interface, QA feature, and performance were incorporated during the development process.

9. Next Steps

The next steps for this project include post-deployment testing to monitor real-world performance and ensuring Firebase scales with the increasing data usage. Future versions could expand to include more complex event detection systems.

10. Conclusion and Recommendations

This project successfully delivered a real-time violence detection system with automated report generation and SentenceTransformer-based question-answering frame retrieval. The system is ready for deployment, but ongoing monitoring and updates to the models will be essential for maintaining performance and accuracy. Future improvements may focus on adding new event detection capabilities, optimizing cloud usage, and ensuring continuous security audits to safeguard the sensitive data stored in Firebase.