

BTech Final Year Project Jan-May 2025

Phase II Proposal Document

Group Number: B7

Guide

Dr. Aishwarya S Kumar

Group Members

V Rajendra Kumar – AM.EN.U4AIE21164

Surya Vamsi Vema - AM.EN.U4AIE21162

V Rithin Chand - AM.EN.U4AIE21172

Shinoy Yandra- AM.EN.U4AIE21168

1. Phase I Review Comments

Review Comments & Suggestions

Review Comment	How It Will Be Addressed
Support multi class classification	Use Video Masked Auto Encoders for training the multi class classification model
Support detection in live video	Implement live detection using open cv
Deployment Optimization (Scalable deployment)	Optimize the model using quantization and pruning for low-resource environments.
Real world implementation with live detection.	Develop a real time Web application for model accessing.

2. Phase II Proposal

Problem Statement

Phase 1 successfully developed an Anomaly detection in CCTV videos (detect if there is any unusual activity going) using Video Vision Transformers. Phase 2 will focus on enabling type of anomaly detection(detecting the type of anomaly like fights, attacks, assaults, robbery etc...), deploying the model with scalable infrastructure, and integrating the model in a real-time application.

Modules Completed in Phase 1

1. CCTV Video Anomaly Classification Model: Developed and trained multiple deep learning models, achieving 85% accuracy with Video Vision Transformer.

Modules to be Completed in Phase 2

1. Multi-class Classification: Train the multi class classifier model to detect the type of anomaly in the CCTV camera videos.

2. API & Backend Development: Build a Flask-based or Fast-API based API and microservices architecture for scalable deployment

3. Web Application Development: Create a cross-platform web application with live video interface

4. Model Optimization for Edge Deployment: Apply quantization and pruning for lightweight AI models on mobile devices.

5. Improved Deployment & Testing: Conduct real-world field tests and optimize models for edge devices like Raspberry Pi.

Dataset

The types of human behaviors include: Intrusion, Fights, Crime, Violent activities so in order to achieve that we selected UCF - Crime dataset. It is a very large dataset of 120 hours of video of 1900 long and untrimmed real world surveillance videos. This consist of 13 realistic anomalies including Abuse, Arrest, Arson, Assault, Road Accident, Burglary, Explosion, Fighting, Robbery, Shooting, Stealing, Shoplifting, and Vandalism. These anomalies will be filtered according to the requirements mentioned above..

Hypothesis

Using the whole dataset to train a multi-class classification model using the masked autoencoder will allow us to detect the type of anomaly. Coupled with a web application allowing to send the live video to the model.

Title of the work	Journal details/product details	Contributions of the paper/product	Limitations
VideoMAE: Masked Autoencoders are Data-Efficient Learners for Self-Supervised Video Pre-Training	NeurIPS 2022	The paper discusses VideoMAE (Video Masked Autoencoders) for self-supervised video pre-training. Key findings include:	VideoMAE only leverages the RGB video stream without using additional audio or text stream
Big Data Mining of Energy Time Series for Behavioral Analytics and Energy Consumption Forecasting	MDPI Energies, 2018	They developed a Bayesian network for behavioral analytics and energy consumption forecasting using time series data.	It only Limited to energy datasets, lacks real-time adaptation and explainability and also it depend on structured data which reduces flexibility
Real-world Anomaly Detection in Surveillance Videos	CVPR, 2019	They Proposed a weakly supervised anomaly detection framework using Multiple Instance	It is limited robust to diverse environment which affects reliability and also anomaly scores lack clear interpretability

		Learning (MIL) for video surveillance.	
Dynamic Segmentation for Real-Time Human Activity Recognition	MDPI Sensors, 2022	They Proposed a temporal correlation-based segmentation for real-time activity recognition, enabling efficient sensor data processing in HAR systems.	They focused mainly on sensor data, which lacks integration with video datasets
Advances in Temporal Modeling for Video Anomalies	MDPI, 2022	They Introduced temporal modeling for scalable video anomaly detection, enhancing accuracy through temporal consistency analysis	This approach lacks benchmark testing on datasets like UCF-Crime, relies heavily on temporal consistency, reducing flexibility in complex scenarios, and has limited consideration of environmental noise and variability.

Research/Product Contribution

This research focuses on the training and quantization of VideoMAE for anomaly detection using the UCF-Crime dataset, while simultaneously developing a real-world product for automated surveillance. The goal is to optimize VideoMAE by improving its computational efficiency and reducing resource overhead, making it suitable for deployment in practical security applications. By leveraging self-supervised learning and quantization techniques, we aim to enhance model inference speed while maintaining high accuracy in detecting anomalies such as theft, assault, and vandalism. This work bridges the gap between cutting-edge research and real-world implementation, ensuring that anomaly detection systems become more scalable and accessible for security monitoring.

Novelty

1. First-time utilization of Video Vision Transformers and Video Masked AutoEncoder on the UCF-Crime Anomaly detection .
2. Edge AI deployment with quantized and optimized models for mobile devices.
3. Live Anomaly detection from a web application in real time.

References

1. Minah Jung et al., :Masked Autoencoders are Data-Efficient Learners for Self-Supervised Video Pre-Training,' NeurIPS 2022.
2. [Anurag Arnab](#) , 'C ViViT: A Video Vision Transformer, arxiv 2021