PCSE25-49

Suspicious activity recognition from a video using Deep Learning

PROJECT SYNOPSIS

OF MAJOR PROJECT

BACHELOR OF TECHNOLOGY

SUBMITTED BY

Rishabh Kumar Panthri,

Mohan Paliwal and

Abhigyan Tomar

On **November 2023**

guided by

Ms Nishu Gupta

KIET Group of Institutions, Delhi-NCR, Ghaziabad (UP)

Department of Computer Science and Engineering

1.	Name of Student	Rishabh Kumar Panthri Mohan Paliwal Abhigyan Tomar
2.	University Roll Number	2100290100133 2100290100098 2100290130004
3.	Class Roll No.	03 32 04
4.	Branch	CSE
5.	Batch	2021-2025
6.	Proposed Topic	Suspicious activity recognition from a video using Deep Learning
7.	Submitted by	Rishabh Kumar Panthri Mohan Paliwal Abhigyan Tomar

Table of Contents :-

Subject	**Page Number**
Introduction	1
Rationale	2
Objectives	3
Literature Review	4
Feasibility Study	5
Methodology	7
Facilities Required for work	9
Expected Outcomes	10
References	11

Introduction :-

Video surveillance systems has reduced crimes to an enormous extent. The drawbacks of using these surveillance system results in additional cost of manpower working for 24*7 that too with less efficiency. The efficiency wouldn't improve significantly evenif one person focuses on one field of view of Closed Circuit Television (CCTV) camera.Automating video surveillance system is an effective way to maximize the efficiency of surveillance.

Anomalies the video may be considered as events that doesn't pertain to normal behavior of people in the surrounding. Anomalies have spatiotemporal dependence, that is, the location and contextof the video must be known before classifying the suspicious activities and normal activities.

Image Classification models give us the spatial details of each frame in the video while temporal details are given by deep neural networks like LSTM, CTNs or 3D CNNs.

The Deep Learning has proved itself to be very useful for video classification. Deep Neural Networks are efficient in training for video classification but the vanishing gradient problem is occurred during training of DNNs which re, where the gradients of the loss function become very small as they backpropagate through the network. This can make it difficult for the network to learn.

In detail the scope of this project is to keep in consideration the vanishing gradient problem and develop a suspicious activity detection using Big Transfer Learning based on Residual Network architecture for getting the spatial details in the video (for each frame) and then employing the transformer encoder for temporal features. Once the suspicious activity is detected and classified the alarm system conveys about it to the respected authorities.

Rationale:-

Public safety is an asset for the government and the civil servants of the government. A smart video surveillance system is an effective method to achieve the milestone. Automated surveillance systems are of immense use in the public sector as –

- It can achieve efficiency that is unimaginable and unprecedented for humans.
- It is capable of sustained operation (24 * 7) without succumbing to exhaustion.
- It reduces the man power and hence the resultant cost.
- It can be trained to focus on multiple instances of anomaly in single frame which is otherwise difficult to track with human eyes.
- Fast response by the concerned authorities as alarm system alerts all of them at once automatically.
- It guarantees public safety.
- It can serve as a useful tool if criminal investigations are involved by giving additional data or filtering non-useful data.

Objectives :-

- Identify Deep Learning Model for detecting activities from a video.
- Design a Deep Learning Model for recognizing suspicious activity.
- Design an alarm system for alert.

Literature Review:-

Surveillance videos were analyzed by Waqas Sultani et al. demonstrated multiple instance learning and deep multiple instance learning model which can be applied to video once it is divided into segments (bag formation) [1]. The authors employ a C3D network to extract visual features from video frames, followed by a three-layer fully connected neural network. The video is divided into non-overlapping segments treated as bags, with each segment considered an instance. A multiple instance learning (MIL) ranking loss function is used with specified constraints. Training involves randomly selecting positive and negative bags for each mini-batch and optimizing using the Adagrad optimizer with dropout regularization. Evaluation is done using frame-based receiver operating characteristic (ROC) curves and area under the curve (AUC), as opposed to equal error rate, to assess anomaly detection performance.

KarishmaPawar and Vahida Attar described factors that are needed to be considered before designing anomaly detection system [2] also discusses the traditional and state of art approaches available for the task. Real-Time Object Detection with Yolo by Geethapriya. S, N. Duraimurugan, S.P. Chokkalingam. Published in 2019 this paper covers the advantages of YOLO (You only look once) for real time object detectionand classification and its working

Deep Amrutha et al. proposed a system which aims to monitor students' activities on a campus using CCTV footage and notify authorities of any suspicious events [4]. The system architecture includes video capture, pre-processing, feature extraction, classification, and prediction. It classifies videos into three categories: students using mobile phones (suspicious), students fighting or fainting (suspicious), and walking/running (normal). The system uses datasets such as KTH, CAVIAR, and YouTube for training and employs a deep learning approach with a VGG-16 pre-trained CNN model and LSTM for feature extraction and classification. The system processes videos into frames, resizes them, and uses OpenCV for pre-processing. A similar approach was used by Lieyun Dinga and others to detect unsafe behavior at construction sites [6].

Kaiming He et al. introduced the concept of deep residual learning, which addresses the degradation problem in very deep neural networks [7]. The key innovation in the paper is the introduction to residual learning blocks, also known as residual units or residual blocks. These blocks enable the training of very deep networks by introducing skip connections that allow the network to learn residual functions. This means that instead of learning the desired mapping directly, the network learns the residual between the input and the output.

Transformers introduced by Ashish Vaswani et al. [8] in 2017 were employed by Anurag Arnab et al. for video classification [9]. The authors propose several methods of factorising the model along spatial and temporal dimensions to increase efficiency and scalability. They regularize the model during training and leverage pretrained image models. The authors achieve state-of-the-art results on multiple standard video classification benchmarks.

Feasibility Study:-

1. Technical Feasibility:

- **Frame Extraction and Preprocessing**: Frame extraction and preprocessing are well-established and technically feasible using standard video processing tools and libraries like OpenCV and FFmpeg.
- **Frame Embedding**: Using pre-trained CNN models for frame embedding is technicallyfeasible and a common practice in computer vision tasks.
- **BiT (ResNet) Model:** Implementing the BiT model by Google is technically feasible, as it's readily available and has been widely used for image classification.
- **YOLO Object Detection (Optional):** Integrating YOLO for object detection is technically feasible if your project requires this functionality**.**

2. Economic Feasibility:

- **Budget Analysis**: Estimate the costs associated with hardware, software, personnel, and ongoing maintenance. Ensure the project is financially feasible within the allocated budget.
- **Return on Investment (ROI):** Evaluate the potential benefits, such as enhanced securityand safety, and assess whether they justify the project's cost.

3. Operational Feasibility:

- **Resource Availability:** Determine whether you have access to skilled personnel who can handle the technical aspects of the project, including machine learning and deep learning experts.
- **Operational Workflow:** Ensure that the proposed workflow aligns with existing operational procedures and doesn't disrupt regular activities in the public place.

4. Legal and Ethical Feasibility:

- **Privacy and Data Protection:** Assess the legal and ethical considerations related to capturing and analyzing video data in public places. Ensure compliance with privacy laws and regulations.
- **Data Retention and Handling:** Define data retention policies, data handling procedures, and security measures to protect collected data

5. Schedule Feasibility:

- **Timeline:** Create a realistic project timeline with well-defined milestones for each phase, from data collection to deployment.
- **Deadlines:** Ensure that the project aligns with any deadlines or security requirements for public safety.

6. Security Feasibility:

- **System Security:** Assess the security of the system to prevent unauthorized access and ensure the integrity of data.

- **Data Security:** Implement measures to protect the collected data, especially when it comes to sensitive information.

7. Social and Environmental Feasibility:

- **Community Acceptance:** Evaluate the community's acceptance of surveillance systems in public places and address any potential concerns.
- **Environmental Impact:** Assess any environmental impact, such as energy consumption, and aim for sustainability in system design.

Methodology:-

The high level abstraction of the project is as follows:-

Step 1. Problem Definition and Scope:

- This project is based on activity classification and then suspicious activity detection.
- The suspicious activities include chain snatching, physical attacks, pickpocketing and suspicious object detection like unclaimed bags at public place. This model can be further improved with training it with more of such activities.
- Ensure that the system can analyze the video in real time or near real time.

Step 2. Data Acquisition:

- Collect video data from surveillance camera or other sources.
- Ensure that you have access to a continuous and high-quality data stream.

Step 3. Data Preprocessing:

- Convert video data to grayscale format.
- Normalize pixel values to a specific range.

Step 4. Frame Extraction:

- The frame extraction is crucial step for analyzing the video.

Step 5. Image Classification (Spatial nature)

- Image classification is done using the Big Transfer Learning based on Residual Network Architecture.

Step 6. Frame Embedding:

- Utilize YOLO for detecting humans potentially threatening objects like unclaimed luggage and then utilizing it for object localization.

Step 7. Temporal Context Modeling with Transformer Encoder:

- Analyze the sequence of classified activities to recognize and classify larger,context-aware activities or actions.
- Configure the Transformer encoder for the desired sequence length and dimensionality.
- Add positional encodings to capture the temporal order of frames (from step 5).
- Let the Transformer encoder process the frame embeddings to capture temporal dependencies and patterns.
- Apply classification layers on top of the encoder's output to classify activities at each time step.

Step 9. Alerts and Notifications:

- Configure an alert system to notify security personnel or authorities in real-time when potentially threatening objects or behaviors are detected.

- The web development methodology for the application involves creating a dynamic alert system that integrates a machine learning algorithm for anomaly detection.

Facilities required for proposed work:-

Here are tools associated with the technologies mentioned in the context of the project:

1.) Data acquisition:

- Video surveillance camera for video source.
- Depending on the data source, we may need additional libraries for data retrieval like pandas, numpy, etc.

2.) Data Preprocessing and Frame extraction

- Python's OpenCV for gray scale conversion and normalizing pixel values
- Python's Open CV for frame extraction.

4.) Image Classification with Big Transfer Learning (ResNet):

- Python libraries like TensorFlow or PyTorch for implementing ResNet.
- TensorFlow Hub is a repository of pre-trained models including BiT models

5.) Frame Embedding.

- Use OpenCV to implement YOLO algorithm.

6.) Transformer Encoder:

- Python's PyTorch or TensorFlow for implementing the Transformer encoder.
- Python's transformers library for pre-trained Transformer models.
- Python's TensorFlow's libraries for deep learning components.

7.) Alert System:

- The backend, built using Flask (Python) or Django, processes real-time data, communicates with the pre-trained machine learning model, and triggers alerts based on defined criteria.
- The frontend, developed with HTML, CSS, and JavaScript (using frameworks like React or Vue.js), provides a user-friendly interface for administrators to monitor activities and configure alert settings. RESTful APIs facilitate seamless communication between the frontend and backend.
- The system ensures security through authentication mechanisms and employs a notification system for multi-channel alerts. The tech stack comprises Flask/Django, HTML/CSS/JavaScript, React/Vue.js, and RESTful APIs, offering scalability and adaptability to evolving security needs. The application fosters a synergistic integration of machine learning and web development for proactive security monitoring.

Expected Outcomes:-

- Big Transfer Learning based on Residual Network architecture has set remarkable benchmarks on ILSVRC-2012, CIFAR-10, VTAB and many more which is enough to prove its relevance for the task of image classification. It is expected to give an unprecedented accuracy in our task.
- Transformers have emerged as a powerful architecture for video classification, demonstrating promising results in various applications. Their ability to capture long-range dependencies and model temporal relationships makes them well-suited for extracting meaningful information from video sequences.
- Both of these architectures are proved efficient against the vanishing gradient problem so their combination will prove to set an irrefutable benchmark in video classification history.

References:-

1.) Real world anomaly detection in surveillance videos by Waqas Sultani, Chen Chen, Mubarak Shah. (2018)

2.) Deep learning approaches for video-based anomalous activity detection by Karishma Pawar and Vahida Attar. (2019)

3.) Real-Time Object Detection with Yolo by Geethapriya. S, N. Duraimurugan and S.P. Chokkalingam. (2019)

4.) Deep Learning Approach for Suspicious Activity Detection from Surveillance Video by Deep Amrutha C.V, C. Jyotsna andAmudha J. (2020)

5.) Video Processing Using Deep Learning Techniques: A Systematic Literature Review by Vijeta Sharma, Manjari Gupta, Ajai Kumar and Deepti Mishra. (2021)

6.) A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory by Lieyun Dinga, Weili Fang, Hanbin Luo, Peter E.D. Lovec, Botao Zhong and Xi Ouyang. (2022)

7.) Deep Residual Learning for Image Recognition (2016) by Kaiming He Xiangyu Zhang Shaoqing Ren Jian Sun.

8.) Attention Is All You Need (2017) by Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit , Llion Jones, Aidan N. Gomez, Łukasz Kaise, Illia Polosukhin**.**

9.) ViViT: A Video Vision Transformer Anurag Arnab, Mostafa Dehghani, Georg Heigold, Chen Sun, Mario Luˇci´c and Cordelia Schmid.