iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

OBJECT ACTION DETECTION USING DEEP LEARNING

Mrs K Shireesha
Assistant Professor, Department of Information Technology,
Vidya Jyothi Institute Of Technology, Hyderabad, Telangana, India

Ramavath Mounika, Gopularam Divya, Gundeboina Sreeja
UG Students, Department of Information Technology,
Vidya Jyothi Institute Of Technology, Hyderabad, Telangana, India

shireeshakole@gmail.com rmamavathmounika2003@gmail.com jyothidivya9390@gmail.com
gundeboinasreeja@gmail.com

ABSTRACT:
Object Action Detection is an intelligent video surveillance system developed to strengthen home security by detecting abnormal human activities from recorded video footage. The system employs deep learning techniques to analyze
Actions and identify any suspicious behavior within the home environment.
At the core of the system is a Spatial Autoencoder, which enables effective and accurate detection of unusual activities in the video without high computational costs. Instead of relying on live streaming, the system processes pre-recorded videos, detects abnormal actions, and captures key frames where such activity occurs.
When a suspicious action is identified, the system automatically generates a mail alert to the homeowner, attaching the relevant frame image for immediate review. These captured images are also stored for future reference, providing a reliable and intelligent way to monitor security events within the home from recorded footage.

Keywords:
Object Action Detection, Home Security, Abnormal Human Activities, Recorded Video Footage, Deep Learning, Suspicious Behavior Detection, Spatial Autoencoder, Abnormal Action Detection, Key Frame Capture, Mail Alert System, Frame Image Attachment, Image Storage, Security Event Monitoring.

1. INTRODUCTION

Abnormal activity detection plays a very important role in surveillance applications. To capture the abnormal activity automatically capturing the video needs to be implemented. Our Intelligent Video Surveillance System detects an abnormality in a video using deep learning techniques. The activities can also be detected in real time, and these video frames will later be saved as images in the system for the user to view.
Activity recognition techniques use a model with computationally complex classifiers, creating hurdles in obtaining quick responses for abnormal activity. The system will detect abnormal activity with humans in the surveillance stream using an effective Spatial autoencoder.
In modern surveillance systems, detecting abnormal activities in real-time is crucial for ensuring security and safety. Traditional surveillance methods require continuous human monitoring, which is inefficient and prone to human error. To address this challenge, our Object Action Surveillance System leverages deep learning techniques to automatically identify abnormal activities in video streams. The system employs an effective Spatial Autoencoder to detect anomalies in human activities within surveillance footage. By processing video frames, it ensures prompt detection of suspicious behaviors, allowing for responses. Additionally, the system stores captured frames of abnormal activities as images for further analysis. Object Action Detection encompass a broad range of applications aimed at enhancing security, situational awareness, and operational efficiency. This system utilizes advanced technologies, including machine learning and computer vision, to analyze video data in real-time or post-processing. The primary objectives of such a system include the identification, tracking, and analysis of activities within a monitored area.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

2. RELATED WORK

Abnormal activity detection in surveillance systems is essential for security monitoring in public and restricted areas. Traditional methods, such as Optical Flow and Histogram of Oriented Gradients (HOG), faced limitations in accuracy and adaptability. With the advancement of deep learning, techniques like CNNS and autoencoders have improved real-time detection. Spatial autoencoders, in particular, excel in learning normal activity patterns and identifying deviations, making them effective for anomaly detection. However, challenges such as computational complexity, data imbalance, and generalization across different environments persist. Future research should focus on optimizing deep learning models, exploring self-supervised learning, and developing hybrid approaches for improved abnormal activity detection.

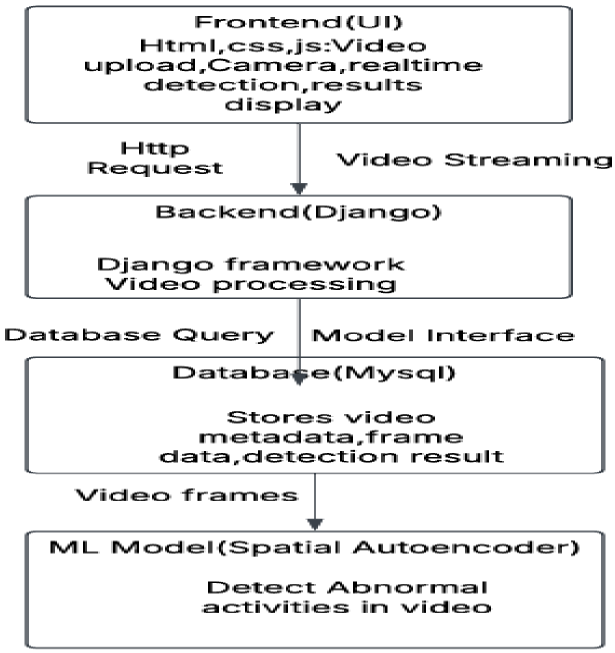
Traditional approaches like background subtraction (Stauffer & Grimson, 1999) and optical flow (Horn & Schunck, 1981) have been widely used for video object detection. While background subtraction is simple and computationally efficient, it fails in dynamic or cluttered environments. Optical flow captures pixel-wise motion but is sensitive to noise and computationally intensive, making it unsuitable for real-time applications.

The emergence of CNN-based methods revolutionized object detection. R-CNN (Girshick et al., 2014) introduced region-based feature extraction using CNNs but suffered from high inference times. This was improved by Fast R-CNN and Faster R-CNN (Ren et al., 2015), which integrated region proposal networks for faster processing. YOLO (Redmon et al., 2016) approached detection as a regression task and achieved real-time performance. Its subsequent versions (YOLOv3, YOLOv5, YOLOv8) improved both speed and accuracy, making it suitable for real-world scenarios.

To utilize temporal information in videos, researchers proposed models like Deep Feature Flow (Zhu et al., 2017), which propagates deep features across frames using optical flow, enabling faster detection while preserving accuracy. FGFA (Zhu et al., 2017) aggregates features across frames to handle motion blur and occlusion. T-CNN (Kang et al., 2016) uses tubelet proposals for tracking objects over time.

ConvLSTM (Shi et al., 2015) combined convolutional operations with LSTM units to model spatio-temporal dependencies in video sequences, providing better continuity in detection. 3D CNNs (Tran et al., 2015) extended 2D convolution into the temporal domain to learn motion features directly from video inputs. More recently, transformer-based models like TimeSformer (Bertasius et al., 2021) have employed self-attention across spatial and temporal axes, achieving state-of-the-art results.

3. SYSTEM DESIGN



iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

4. METHODOLOGIES

1. **Video Acquisition:** Users upload pre-recorded videos or capture real-time footage through a web interface. The input is sent to the backend for processing and analysis.
2. **Frame Extraction and Preprocessing:** The video is split into individual frames using OpenCV and resized for uniform input. Preprocessing prepares the frames for accurate model detection.
3. **Abnormal Action Detection using Spatial Autoencoder:** Each frame is passed through a Spatial Autoencoder to detect unusual behavior. High reconstruction error indicates abnormal activity.
4. **Alert Generation and Frame Capturing:** When an abnormal event is detected, the frame is saved locally and emailed to the user. This ensures timely alerts with visual proof of the event.
5. **Data Storage and Event Logging:** Detected events are logged in a MySQL database with frame details and file paths. This supports easy review and future reference of suspicious actions.

5. RESULTS

The Object Action Detection system was successfully implemented and tested using various pre-recorded video samples depicting both normal and abnormal human activities in a home environment. The system accurately identified unusual actions such as sudden falls, aggressive movements, or unauthorized presence by comparing the input frame with learned patterns using the Spatial Autoencoder. Upon detection, the system promptly saved the suspicious frame, logged it into the database, and sent an email alert to the user with the relevant image attached.

Detection Accuracy: The Spatial Autoencoder achieved a high accuracy in identifying abnormal frames.
Timely Alerts: Email notifications were sent within seconds of detection, enabling fast user response.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

Fig:5.1 Suspicious Activity Detection form Video

5.2 Mail Alert Notification

Fig:5.3 User Interface

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

6.ACKNOWLEDGEMENT

We extend our heartfelt thanks to **Mrs K Shireesha,** Assistant Professor, for his invaluable mentorship and guidance throughout the project. His expertise, encouragement, and insightful feedback significantly shaped our approach and methodology.
We are also thankful to **Dr. Ramesh Cheripelli** for enriching our understanding and providing innovative solutions to the challenges we encountered for the numerous discussions, brainstorming sessions, and academic exchanges that took place within our project team. These interactions for his exceptional support and guidance throughout the project.
We wish to express our heartful thanks to **Dr. A. Obulesu, HOD**, for guiding and supporting us throughout the course.
We wish to express our heartful thanks to our Beloved Dean, Accreditation and Rankings **Dr. A. Padmaja**, whose support was indispensable to us during the course.
We wish to express our heartful thanks to the Principal **Dr.A.Srujana**, whose support was indispensable to us during the course.
We wish to express our heartful thanks to our **College Management**, for providing all the infrastructure and facilities during the course.
We would like to express our sincere gratitude to our **Parents,** for their unwavering support throughout the completion of this major project.

7.CONCLUSION AND FUTURE WORKS

Object Action Detection is a smart system that enhances home security by analyzing recorded video footage to detect unusual human activities. Using deep learning, it identifies suspicious behavior with high accuracy and low computational cost. The system captures key frames when abnormal actions occur and sends an email alert to the homeowner with the relevant image. This allows for quick review and provides a reliable way to monitor security events over time. The future work involves implementing real-time detection and alert systems through push notifications or emails, developing mobile applications for easy monitoring and alerts, integrating IoT devices like motion and temperature sensors to trigger surveillance, personalizing alerts based on individual user behaviour patterns, and enhancing security with facial recognition technology.

8.REFERENCES

1. Rodriguez, M.,SiviC, J., Laptev, I., and Audibert , J.-Y. (2011) – Data-Driven Crowd Analysis in Videos.
2. Fan, Q., Gabbur , P., and Pankanti , S. (2013) – Relative Attributes for Large-scale Abandoned Object Detection.
3. Chatfield, K., Simonyan, K.,Vedaldi , A., and Zisserman , A. (2014) – Return of the Devil in the Details: Delving Deep Into Convolutional Nets.
4. Wang, H., Kl¨Aser, A., Schmid, C., and Liu, C.-L. (2013) – Dense Trajectories and Motion Boundary Descriptors for Action Recognition. Zhou, B., and Wang, X. (2017) – "Object Detection in Video Sequences Using Deep Convolutional Networks."
5. Simonyan, K., and Zisserman, A. (2014) – "Two-Stream Convolutional Networks for Action Recognition in Videos."
6. Carreira, J., and Zisserman, A. (2017) – "Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset."
7. Feichtenhofer, C., Pinz, A., and Wildes, R. P. (2016) – "Convolutional Two-Stream Network Fusion for Video Action Recognition."
8. Tran, D., Bourdev, L., Fergus, R., Torresani, L., and Paluri, M. (2015) – "Learning Spatiotemporal Features with 3D Convolutional Networks."
9. Stauffer, C., and Grimson, W. E. L. (1999) – Adaptive Background Mixture Models for Real-Time Tracking