IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

VIDEO SURVEILLANCE USING DEEP LEARNING: A SURVEY

[1]Arshiya Dilshad, [2]Kavita Agrawal

[1]Student, [2]Head of the department

[1]Computer Science & Engineering

Integral University, Lucknow, India.

Abstract:- In security and monitoring systems spanning a range of areas, such as public safety, traffic management, and private property protection, video surveillance has emerged as a crucial instrument. Conventional surveillance techniques frequently depend on human error-prone and labour intensive manual monitoring. An overview of deep learning approaches for video surveillance is given in this study, with particular attention to important topics including anomaly detection, activity recognition, and object detection. We have examined how convolutional neural network (CNN), recurrent neural network (RNN), and their different variations process and interpret video input in terms of architecture and functionality. We also go over the difficulties in addressing occlusion, scalability, and real-time processing in video surveillance. This paper demonstrated how the deep learning models may effectively enhance the functionality of video surveillance systems using case studies and experimental outcomes. This research survey the past work on video surveillance on different methods and approaches.

Index term:- Real-time processing, CNN, RNN, object detection, activity recognition, anomaly detection, deep learning, video surveillance, and security.

I. INTRODUCTION

Video surveillance is becoming a crucial part of today's infrastructures for monitoring and security. There is an increasing need for efficient surveillance systems, as seen by the abundance of cameras in public areas, businesses, and homes. Conventional video surveillance systems, on the other hand, mostly depend on human operators to keep an eye on recorded material and live feeds, a laborious procedure that is prone to oversight and human mistake. Video surveillance systems might be completely transformed by deep learning, a type of artificial intelligence (AI) that leverages neural networks to learn from data. Impressive performance has been demonstrated by deep learning models, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), across a range of computer vision applications, such as object detection, activity recognition, and picture recognition. These developments make it possible to create automated surveillance systems that are very accurate and efficient in identifying patterns, detecting abnormalities, and analysing video feeds in real time. The main purpose of this work is to present a review of the use of deep learning techniques in video surveillance. We will look at the basic ideas behind deep learning, investigate the designs and features of CNNs and RNNs, and talk about how they are used to process and analyze video data.

II. LITREATURE REVIEW

The field of video has significantly improved with the application of deep learning. The development of deep learning-based surveillance systems has been aided by the works and methodology reviewed in this part. Video surveillance requires object detection as one of its essential elements. Traditional machine learning techniques were employed in the early methods, which frequently had performance limitations and necessitated the creation of handmade features. Deep learning has significantly increased object identification efficiency and accuracy, especially with the introduction of Convolution neural network (CNN). It included the methods and advantages of each reviewed strategy. A survey of machine learning (ML) methods for visual identification of activities. There are many surveys and reviews were then conducted covering a wide range of topics, including how to handle challenges, crowd estimation, abnormal behaviour and crowd recognition, crowd and violence detection using intelligent video surveillance, and deep learning approaches.

Motion-based object detection is the method of identifying locations or noteworthy things in a video by analysing variations in motion. [1] In intelligent video surveillance, the YOLO object detection approach yielded great outcomes. when the models were trained using the given data. Thanks to its enhanced algorithms, the system functioned more precisely, identified objects and categories, and could detect motion. [7] With unclear predictions, the neural networks and algorithms utilized in [1] functioned well.

The proposed CNN-based methodology is one potential way to count and identify crowds in still images from various scenarios.[10] A CNN model is trained for various computer vision tasks. It is utilized to extract characteristics that enable the accurate depiction of crowd density. Additionally, a strong connection has been seen between the local counts in the surrounding regions, and the accuracy of detection is improved by feature extraction approaches.[2]Using an ensemble deep learning model, more convolutional layers may be added for even higher accuracy.

The implementation of a crowd behaviour analysis technique based on the optical flow magnitude correlation coefficients is shown in [3]. It's been stated that an unusual incident has been observed if the throng disperses quickly. Rapid movements tend to have large magnitude variations and fast motion vectors. In [3]Motion patterns during anomaly times can be distinguished by examining the correlation coefficients of succeeding frames. It is therefore concluded that the optical flow approach combined with correlation estimation is the most efficient way for detecting anomalous crowd behaviour wherever.

By using a machine learning method based on Support Vector Machines (SVM)[20] to assess if a human action in a video is violent or not. An pose estimation technique was applied. It investigated how different combinations of these features affected the ability to identify violent behaviour from video. Overall, the classification obtained by the top-performing SVM model was excellent. It included key points, angles, and contact characteristics that were computed throughout an input range of 60 frames. A computer vision-based system is proposed in[18][9]research to calculate important metrics automatically from video sequences, including detection, pattern recognition, assessment of crowd density, and traffic concentration. We also look at helpful crowd attributes like direction and speed that crowd control personnel would find helpful.

A detection-based counting technique may precisely estimate crowd numbers in low-density areas, it is less dependable in crowded settings. However, a regression-based technique is used to retrieve the total density information in locations with high population density[17]. It may choose the optimal counting technique adaptively based on the real density conditions at each point on the picture. In order to estimate the crowd density, DecideNet first creates unique density maps using detection and regression. Multi-task learning has been studied recently to increase the reliability and effectiveness of surveillance systems. These methods can take use of similar characteristics and lower computational cost by training a single model to handle many related tasks, such activity recognition, anomaly detection, and object detection.

III. METHODOLOGY

A. CONVOLUTION NEURAL NETWORK

A Convolutional Neural Network (CNN) is a type of deep learning algorithm primarily used for image recognition and classification tasks. Convolutional Neural Networks (CNNs) are widely used in video monitoring applications due to their ability to effectively analyse visual data, including detecting and tracking objects, recognizing activities, and understanding scene dynamics. Here's how CNNs are applied in video monitoring.[1][3] CNNs are essential to video surveillance applications because they provide automatic real-time visual information analysis, interpretation, and understanding, improving efficiency, security, and safety across a range of applications. An expansion of the conventional 2D CNN, a 3D CNN is made especially to process data that has a third dimension, including volumetric pictures or video (like CT or MRI scans). A 3D CNN processes 3D spatiotemporal data, whereas a 2D CNN normally works with 2D spatial data. Because 3D data processing involves more complex calculations and has more parameters compared to other types of neural networks, training a 3D CNN usually demands a substantial amount of processing power. [3] Furthermore, because 3D CNNs are prone to overfitting, methods like data augmentation and dropout are frequently used to lessen this problem.

B. LONG SHORT TERM MEMORY

Recurrent neural networks (RNN) of the Long Short-Term Memory (LSTM) type function especially well for sequence prediction tasks. LSTMs are used to interpret and evaluate sequential data from video streams in the context of video surveillance. Since long-term patterns can be recognized by LSTMs, they are perfect for identifying actions and behaviours that occur over a series of frames in video surveillance material.[27][28] Through the acquisition of typical movement and behaviour patterns, LSTMs are able to recognize deviations or anomalies that might point to odd or suspicious activity. Using historical video data, LSTMs can forecast future states or motions, which is useful for proactive surveillance and threat mitigation.

C. YOLO

Typically, "YOLO" stands for "You Only Live Once." This well-known acronym, which is frequently linked to taking chances and grabbing opportunities, expresses the notion that life is short and ought to be lived to the fullest. The term became widely used after becoming popular in social media and music, especially among young people [1][7]. It describes a real-time object identification system that is well-known for its quickness and accuracy in the fields of object recognition and computer vision.

D. SUPPORT VETOR MACHINE

Support vector machines, or SVMs, are useful in crowd monitoring for a number of applications, including anomaly detection, crowd behaviour analysis, crowd density estimate, and crowd counting. [20]. It works especially well for classification problems where the data can be translated into a higher dimensional space where linear separation is achievable, or where the data is separable linearly.

E. HISTOGRAM OF GRADIENT

HOG descriptors are frequently employed as feature vectors for the training of machine learning models for tasks involving object detection and recognition. They are efficient at preserving the gradient information that distinguishes objects in photos, making them resistant to alterations in background clutter and lighting[5][11][19]. It has been used for many different object detection tasks, but is especially well-liked for detecting pedestrians.

Summary:-

S.NO	REFRENCE NO.	YEAR	METHODS	PROPOSED WORK	FINDINGS
1.	[1]	2024	Deep learning using YOLO	Motion-based object detection method for tracking and identifying things	Examining changes in motion in a video
2.	[2]	2024	Swin Transformer-based deep learning (DL) model	Establishing a Dynamic Multipoint Virtual Private Network to provide end-to-end security	substantial effects on crowd control, security, and public safety
3.	[3]	2023	CNN Based Methods	Image Processing and Image Extraction	Utilizing a deep convolutional network called RESNET,
4.	[4]	2023	Optical Flow Magnitude	Understanding human behaviour from footage	The optical flow magnitude correlation coefficients have been put into practice.
5.	[5]	2023	OpenCV based Histogram of Oriented Gradients (HOG)	System for Real-Time Crowd Detection that is completely autonomous	Monitoring to help academic institutions oversee students on campus more effectively.
6.	[6]	2023	Deep Learning and Regression	Density estimation	Crowd Counting
7.	[7]	2022	YOLO	Social Distancing	Person Identification and Positioning
8.	[8]	2021	Big Data and Artificial Intelligent	AI-powered L&T Smart World crowd control system	Minimize the likelihood of dangerous and overcrowded situations
9.	[27]	2021	3D Convolution neural network	Crowd violence detection framework	Aims to detect and estimate the violence level in real time
10.	[28]	2021	CNN and LSTM	Violence Detection	Detecting the presence of any violent act.
11.	[9]	2021	2D Convolution Neural Network	A 2D convolutional neural network (CNN) that has been trained beforehand	Recognizing odd behaviour in crowded scenarios in films
12.	[10]	2021	CNN	An intelligent system for monitoring crowds that can recognize and count people	Detecting and counting people through the use of overhead view video sequences
13.	[11]	2020	Histogram of oriented gradients (HOG)	Detection of pedestrians in forbidden territory	Using SVM to categorize behaviour as violent or non-violent.
14.	[12]	2020	RFID Wi-fi Bluetooth	Survey Using RFID, wi-fi Bluetooth	Monitoring Crowd
15.	[13]	2020	CNN and Deep convolutional neural networks (DCNN)	Scale Driven Convolutional Neural Network (SD-CNN) and DISAM models	Comparing the CNN and DCNN Methods
16.	[14]	2019	Deeper Convolution Neural Network through IOT	The method is centred on creating a density map generator using CNN.	enumerating people and producing superior density maps
17.	[15]	2019	Optical flow method and social force model	Ingenious use of predictive neural networks for crowd anomaly identification	The degree of deviance is assessed using actual frames.
18.	[26]	2019	Convolution neural network	Dissects the frames from the image and classifies it into violence or non-violence.	Detecting abnormal activities
19.	[24]	2020	CNN and LSTM	Monitor the situation through mobile application.	Detection of violence activities.
20.	[25]	2018	Convolution neural network - GoogleNet	Pre-trained CNN architecture	Identifies aggressive human behaviour
21.	[16]	2018	Physical Model Based	Force Model and fluid Dynamics	Crowd Segmentation and motion detection using Physics methods
22.	[17]	2018	Decide-Net	Crowd Counting using (Detection and Density Estimation Network)	Crowd Counting by detection and regression.

23.	[18]	2017	Machine Learning, Longest Common Subsequence	Locate and identify crowded	regions Estimate crowd density
24.	[19]	2015	Histogram of Oriented Gradient (HOG)	Hierarchical crowd breakdown at any particular picture frame	An implementation of motion based crowd modelling makes use of a GMM of DT technique.
25.	[20]	2013	Support Vector Machine (SVM)	RGB images with spatially and temporally encoded characteristics are used in the ResNet architecture.	Identification of violent action

IV. SCOPE AND APPLICATION

There are numerous benefits to implementing CNNs for video surveillance., security, and urban planning. CNNs offer effective analysis of complicated visual data by utilizing the power of deep learning and computer vision. Deep learning-based video surveillance is a major development in security and monitoring that offers improved capabilities and efficiency in a variety of industries. But it also means that privacy, moral ramifications, and technological difficulties must be carefully considered.

A. SCOPE

- Security and Safety:- In order to help prevent crime and improve security, deep learning models may evaluate video feeds in real-time to identify potentially dangerous situations, illegal access, and suspicious activity. [11][9][8][7][5][2]In order to ensure public safety, it is utilized in public areas such as train stations, airports, and shopping canters to keep an eye out for any strange activity or behaviours.

- Traffic Management:- For improved traffic management and planning, deep learning may be used to track traffic flow, identify infractions, and examine pedestrian movement. Real-time video stream analysis to find incidents and notify the appropriate authorities for prompt action.

- Healthcare:- Deep learning may be used in hospitals to monitor patients and make sure they get help quickly in an emergency. It is very helpful in senior care, this technology may identify falls and other mishaps so that quick help can be given.

- Smart Cities:- Urban planning and development can benefit from the use of video analytics to collect information on how people use public places. Environmental factors include trash management, pollution monitoring, and others.

- Sports and Entertainment:- Examining player motions, spectator behaviour, and other elements of sporting events. Enhancing video material with automated editing and sophisticated analytics.

- Home and Security:- keeping an eye out for any strange activity or unapproved entrance on the property at home. Connecting additional smart home appliances for more automation and security is known as smart home integration.

- Industrial and Manufacturing:- keeping an eye out for danger and quickly identifying compliance with safety procedures. examining industrial procedures in order to find inefficiencies and boost output.

B. APPLICATION

- Real Time Threat Detection:- Real-time video feeds may be processed by deep learning models to identify possible dangers like firearms, conflicts, or suspicious activity. This enables prompt response and intervention.

- Facial Recognition:- Video streams utilized for law enforcement, access control, and tailored user experiences to identify and authenticate people.

- Object Detection and Tracking:-Tracking and identifying things (cars, people, and animals) in a video feed, this is helpful in a many situations, including security surveillance, traffic monitoring, and wildlife monitoring.

- Behavioural Analysis:- Examining behaviour to spot oddities like lingering, sprinting, or strange motions that might point to a security risk or other issues.

- Crowd Management :- Crowd dynamics are tracked and analysed to ensure event safety, avoid congestion, and maximize the use of public space.

V. CONCLUSION

Deep learning-based video surveillance has greatly expanded the capabilities of conventional surveillance systems, offering more effective, precise, and intelligent monitoring options. The way video data is studied and used has changed dramatically with the inclusion of deep learning techniques like Long Short-Term Memory (LSTM) networks for temporal pattern recognition and Convolutional Neural Networks (CNNs) for object detection and categorization.

Deep learning-based video surveillance appears to have a bright future because to ongoing developments in artificial intelligence and machine learning. Deep learning in surveillance will become much more powerful and useful with the addition of edge computing and innovations like better model designs and training techniques. In addition, it will be essential to do continuing research to address ethical and privacy issues to guarantee the appropriate application of these technologies.

In conclusion, the availability of strong tools for precise, scalable, and real-time monitoring has allowed deep learning to completely transform the field of video surveillance. It is a crucial piece of technology in contemporary surveillance systems because, in spite of its difficulties, it improves security, safety, and operational effectiveness in many different fields.

ACKNOWLEDGEMENT

The authors express their sincere gratitude to Integral University Lucknow's Research and Development department for providing the tools needed to carry out this study successfully.

REFRENCES:-

[1] A. P. N. V. and M. A. , “Cogni-Watch Intelligence Video Surveillance System,” 2024.

[2] M. Q. A. W. C. J. R. and N. A. , “Public Vision: A Secure Smart Surveillance System for Crowd Behaviour Recognition,” 2024.

[3] H. G. S. J. N. G. S. S. and P. K. S. , “Real – Time Crowd Monitoring System,” 2023.

[4] P. D. C. D. V. R. S. and D. N. C. , “Crowd behaviour anomaly detection using correlation of optical flow magnitude,” 2023.

[5] P. S. and V. R. J. , “A Real-Time Crowd Detection and Monitoring System using Machine Learning,” 2023.

[6] A. P. M. D. V. T. and P. S. , “Crowd counting analysis using deep learning,” 2023.

[7] M. A.-S. S. K. I. A. and C. S. , “A Social Distance Estimation and Crowd Monitoring System for Surveillance Cameras,” 2022.

[8] L. R. and . R. S. S. , “Bigdata Enabled Realtime Crowd Surveillance Using Artificial Intelligence And Deep Learning,” 2021.

[9] A. Mehmood, “Efficient Anomaly Detection in Crowd Videos Using Pre-Trained 2D Convolutional Neural Networks,” 2021.

[10] I. A. M. A. A. A. and G. J. , “IoT-based crowd monitoring system: Using SSD with transfer learning,” 2021.

[11] A. M. A.-S. S. S. A. . N. S. F. and . M. O. K. , “In-Depth Survey to Detect, Monitor and Manage Crowd,” 2020.

[12] A. K. J. A. S. K. K. W. A. and F. K. , “Crowd Monitoring and Localization Using Deep Convolutional Neural Network: A Review,” 2020.

[13] M. H. A. S. K. J. M. S. A. and K. M. , “CNN-Based Crowd Counting Through IoT: Application For Saudi Public Places,” 2019.

[14] . Y. L. . K. H. X. T. and T. W. , “Abnormal Crowd Behaviour Detection Based on Predictive Neural Network,” 2019.

[15] X. Z. Q. and H. , “Physics Inspired Methods for Crowd Video Surveillance and Analysis,” 2018.

[16] J. L. C. G. D. M. and A. G. H. , “Decide-Net: Counting Varying Density Crowds Through Attention Guided Detection and Density Estimation,” 2018.

[17] . S. D. K. . M. T. M. K. A. and S. A. K. , “Towards a Crowd Analytic Framework For Crowd Management in Majid-al-Haram,” 2017.

[18] M. S. Z. J. D. M. A.-M. and H. B. , “Hierarchical Crowd Detection and Representation for Big Data Analytics in Visual Surveillance,” 2015.

[19] D. K. S. S. P. M. K. R. and M. A. A. , “Human Crowd Detection For City Wide Surveillance,” 2020.

[20] D. N. A. F. and . P. C. , “A Machine Learning Approach to Detect Violent Behaviour from Video,” 2013.

[21] A. S. B. and M. M. T. , study on Opinion Spamming: Fake Consumer Report Detection, 2022.

[22] M. M. T. and . M. S. , Hmm Based Offline Signature Forgery Detection.

[23] K. M. Siddiqui M.S., Investigation of Big Data analytics for Sustainable smart city development an emerging country.

[24] S. S. B. S. and S. N. , “A fully integrated violence detection system using CNN and LSTM,” *International Journal of Electrical and Computer Engineering (IJECE),* vol. 11, no. 4, pp. 3374-3380, 2020.

[25] A. M. A. B. S. and Z. H. , “Violence Detection in Surveillance Videos with Deep Network using Transfer Learning,” *2nd European Conference on Electrical Engineering and Computer Science (EECS),* 2018.

[26] . R. E. G. S. A. A. and M. P. , “Violence Detection System using Convolution Neural Network,” *SSRG International Journal of Electronics and Communication Engineering,* 2019.

[27] . K. G. . K. I. and T. T. , “Crowd Violence Detection from Video Footage,” *Research Gate,* 2021.

[28] H. M. M. S. . J. M. P. K. and H. P. , “Violence Detection in Surveillance Videos,” *International Research Journal of Engineering and Technology (IRJET) ,* vol. 8, no. 12, 2021.