INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 03, March 2024, pp: 930-934

TO IDENTIFY SUSPICIOUS ACTIVITY FROM SURVEILLANCE FOOTAGE

Krutika Mukund Karande[1], Gayatri Mahesh Dhore[2], Prof. Dipali Khairnar

[1,2,3]Department of Information Technology Dr. D. Y. Patil School of Engineering, Pune, Maharashtra, India.

DOI: https://www.doi.org/10.58257/IJPREMS33052

ABSTRACT

High-quality CCTV cameras have been installed using a variety of technologies to ensure the protection of people and property. It is not feasible to manually keep an eye on every activity at all times. In this work, the concept of CNN is utilized to identify if behavior in an environment is suspicious or normal, and a system that alerts the similarity authority in the event that it predicts suspicious activity is proposed. It's important to remember that the deployment environment, the machine learning model's architecture, and the caliber of the training data all affect how successful a suspicious activity detection system.

This paper focuses on a deep learning approach to detect suspicious human activity and fight using CNN from images and videos. We analyze different CNN architectures and compare their accuracy. We design our systems that can process video footage from cameras in real time and predict whether activity is suspicious or fight found or not.

Keywords: recommendation - Video Surveillance, Anomaly detection, Machine learning, Convolutional neural networks, Image processing

1. INTRODUCTION

One of the most active areas of computer vision research at the moment is visual surveillance. The number of surveillance cameras has dramatically expanded as public safety and terrorist threats become more of a concern. With any surveillance system, a security guard would keep an eye on the surroundings all the time and take prompt action to prevent any criminal activity. Remaining dedicated to watching hours upon hours of videos despite the possibility of losing or forgetting important details. For a surveillance system to function, automatic techniques are therefore required. Because abandoned luggage is so hard to spot in a crowded area, it is the most often disregarded security issue. This makes it evident that automated techniques for locating misplaced bags are required. There are hardly many publicly accessible

As crime rates rise, it becomes problematic if the appropriate preventative measures are not implemented and criminals are not promptly identified. The majority of cities and metropolitan areas have deployed surveillance systems that continuously gather data. The enormous amount of surveillance data means that there is a greater likelihood of suspicious activity. However, because these jobs are too complex and resource-intensive for artificial intelligence to undertake, human monitoring is necessary to detect such behaviors. One method to simplify an activity for automation is to break it down into smaller components and identify subtasks that could lead to potential crimes. We use our models to try and identify two primary pathways that could lead to crimes.

With any surveillance system, a security guard would keep an eye on the surroundings all the time and take prompt action to prevent any criminal activity.

Remaining dedicated to watching hours upon hours of videos despite the possibility of losing or forgetting important details. For a surveillance system to function, automatic techniques are therefore required. Because abandoned luggage is so hard to spot in a crowded area, it is the most often disregarded security issue. This makes it evident that automated techniques for identifying abandoned bags are required.

One of the best architectures for challenging learning problems is Deep Neural Networks. Machine learning models provide high-level representations of visual data and automatically extract features. Because the feature extraction procedure is entirely automated, this is more widely applicable.

Convolutional neural networks (CNNs) can directly learn visual patterns from picture pixels. Long short-term memory (LSTM) models can learn long-term dependencies in the context of a video stream. An LSTM network can retain information. The suggested system will track human behavior over time using CCTV footage and alert users to any questionable activity. Event detection and the identification of human behavior are the two main elements of intelligent video surveillance.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 03, March 2024, pp: 930-934

2. PREVIOUS APPROACH

One of the main issues with computer vision that has been researched for more than 15 years is human suspicious behavior. Because so many programs could profit from questionable activity, this is crucial. Applications such as video surveillance, sophisticated human-computer interaction, sign language detection, human motion tracking and behavior comprehension, and marker low motion recording, for instance, exploit human suspicious activity. It is challenging to estimate human stance from depth photos due to low resolution and noisy depth information of low-cost depth sensors, which have restrictions such as being restricted to indoor use. As a result, we employ neural networks to get around these issues.

Computer vision and image processing are actively researching the detection of suspicious human activity and fights from surveillance footage. These Conv-LSTM networks can forecast a video sequence's progression based on a small number of input frames. The reconstruction mistakes of a collection of estimates with irregular video sequences provide lower regularity scores as they gradually diverge from the real sequence, which is how consistency scores are determined. The models use a composite structure and track how conditioning has a unique impact on the acquisition of more meaningful representations.

Next Because it's unclear exactly what constitutes an abnormal action in a lengthy video series, automating research article detection of such events is difficult. The authors approach the issue by training generative models that use constrained supervision to identify anomalies in films. sophisticated Convolutional Long Short-Term Memory (Conv-LSTM) networks that can be trained end-to-end and projected to forecast how a video sequence would unfold from a small number of input frames.

It has been shown that anomaly detection with sparse coding performs better, even when factoring in the theories of dictionary learning, feature learning, and sparse representation. This research proposes a new neural network for anomaly detection, called Anomaly Net, by implementing sparse representation, dictionary learning, and feature learning in three combined neural processing blocks. In particular, the authors create a motion fusion block and a feature transfer block together to learn enhanced features. This allows users to enjoy the advantages of removing background noise, capturing motion, and enhancing data inadequacy.

Next The main focus of the paper is on the intrinsic redundancy of video structures, and the authors suggest a useful sparse combination learning approach. It achieves respectable results in the detection phase without sacrificing the caliber of the output. Because the new method effectively reduces the original complex problem to one where only a few less expensive small-scale least square optimization steps are taken into consideration, the short running time is fail-safe. When calculating on a standard desktop PC with MATLAB, the technique achieves good detection rates on benchmark datasets.

By citing the study that proposes a completely untested dynamic sparse coding methodology based on online sparse reconstructibility of query signals for identifying uncommon events in videos. The algorithm operates on a principled convex optimization formulation that enables both a sparse reconstruction code and an online dictionary to be mutually inferred and updated. This is based on the perception that regular events in a video are typically liked to be reconstructible from an event dictionary, whereas infrequent events are not.

One of the greatest solutions to the problem of security in different locations is the use of surveillance cameras. Because it is so difficult to detect and identify criminal and deviant conduct, modern systems require human labor to monitor them. Thus, this work conducts a survey on anomaly detection for surveillance cameras utilizing various ideas such as deep learning

3. MOTIVATION OF PROJECT

It could be able to stop or lessen probable harm by creating an automated system that can identify suspicious activities and notify authorities or pertinent staff. Large data sets can be analyzed using machine learning and related approaches to find patterns that might point to questionable behavior.

4. GOALS AND OBJECTIVE

1) To create an image from a video.

2) To use K means clustering to frame segmentation.

3) To extract a frame using the frame sequence and background subtraction. to identify the item.

4) To use Deep Belief Network (DBN) for action recognition.

5) To use a learned dataset to categorize activities as normal or suspicious.

To notify a security guard.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 03, March 2024, pp: 930-934

5. METHODOLOGY

Convolutional neural networks, or CNNs, are Artificial Intelligence systems built on multi-layer neural networks. CNNs can recognize and classify objects, detect their presence, and segment them. They do this by learning pertinent properties from images. An activity is a series of steps taken to achieve a goal. An individual's series of subsequent actions or duties can be referred to as an activity. Certain actions, such waking up, looking around, sitting down, eating, drinking, leaving, coming, putting up, putting down, writing, and so on, are considered Usual Activities. Unusual Activity is any activity that deviates from the specified list of activities. These strange behaviors result from bodily and emotional suffering. The process of discovering and detecting behaviors that deviate from a regular or well-defined set of activities and draw attention from people is known as unusual activity and anomaly identification. When it comes to computer vision, one of the main issues that has been spent more than 15 years in school. Because so many programs could profit from questionable activity, this is crucial. Applications such as video surveillance, sophisticated human-computer interaction, sign language detection, human motion tracking and behavior comprehension, and marker low motion recording, for instance, exploit human suspicious activity. It is challenging to estimate human pose from depth photos due to low resolution and noisy depth information of low-cost depth sensors, which are restricted to indoor use, among other restrictions.

Consequently, in order to prevent these issues, we deploy neural networks. Computer vision and image processing are actively researching the detection of suspicious human activity and fights from surveillance footage.

Even though the research community has given unusual activity and anomaly detection a lot of attention and finds it to be beneficial, the field still confronts the following issues as a result of advancements in the field:

1.The accuracy of activity recognition falls because different participants move in various ways at different times.

2. The orientation and position sensitivity of smart phones, wearable sensors, and other devices.

3. Motion during the break between two tasks is hard for any classification algorithm to identify.

4. Constraints on energy and resources

tinier datasets Instantaneous processing Creating a real-time intelligent surveillance system is a more difficult undertaking. When extracting foreground items and monitoring them, films with complex backgrounds require additional processing time. detection of static objects Static object detection in abandoned object detection is a difficult task because background subtraction only detects moving items in the foreground. Even if technology has advanced tremendously and knowledge has been applied to improve human welfare, there are still many instances in which it is impossible to find and apprehend those who have committed heinous crimes. Miserable crimes like hit-and-run incidents, robberies in busy areas, money laundering, etc. constitute a serious danger to the nation's reputation and economic growth in the international market. The majority of robberies are carried out by criminals against a lone victim while brandishing a firearm.

Topically, a CNN is made up of four different kinds of layers:

Involutionary

Combining

Blunt / compress

Totally Networked

a) Image dimensions, measured as follows: 5 x 5 x 1 (number of channels, such as RGB) a)The green area in the demonstration above looks like our input image, I, which is 5x5x1. The Kernel/Filter, K, which is shown in yellow, is the element that performs the convolution process in the first section of a convolutional layer. K has been chosen as a 3x3x1 matrix.

b) As a result of Stride Length = 1 (Non-Strided), the kernel shifts nine times, multiplying elementwise K by the portion P of the image that the kernel is hovering over each time.

c) Using a specific Stride Value, the filter advances to the right until it parses the entire width. With the same Stride Value, it then hops down to the left of the image's commencement and continues in this manner until the entire image has been traversed.

d) The Convolution Operation's goal is to extract the input image's high-level characteristics, including edges. One Convolutional Layer is not the only place where Conv Nets can be used.

Traditionally, Low-Level characteristics like edges, color, gradient orientation, etc. are captured by the first Conv Layer. As more layers are added, the architecture also adjusts to the High-Level characteristics, providing us with a network that comprehends the photos in the dataset holistically, just like humans would.

f) The process yields two different types of results: one where the dimensionality of the convolved feature is less than that of the input, and the other where the dimensionality is either raised or stays the same.

g) To do this, apply Same Padding in the latter scenario and Valid Padding in the former.

h) The Pooling layer, which is comparable to the Convolutional Layer, is in charge of shrinking the Convolved Feature's spatial size. By doing this, the amount of processing power needed to process the data through dimensionality reduction will be reduced.

i) In addition, it is helpful for obtaining dominating features that are positional and rotationally invariant, which keeps the model's training process going strong.

j) Max Pooling and Average Pooling are the two forms of pooling. The maximum value from the area of the picture that the Kernel covers is returned by max pooling.

k) Average Pooling, on the other hand, yields the average of all the values from the area of the picture that the Kernel covers.

l) Another function of max pooling is noise suppression. In addition to dimensionality reduction, it does de-noising and completely eliminates the noisy activations. But conversely.

m) Average Pooling merely reduces dimensionality as a means of attenuating noise. As a result, Max Pooling outperforms Average Pooling by a wide margin.

n) Using the output of the convolutional layer as a starting point, a Fully-Connected layer can be added to learn non-linear combinations of high-level features. In that area, the Fully-Connected layer is learning a function that might not be linear.

o) We will now flatten the input image into a column vector after transforming it into a format that works for our Multi-Level Perceptron. A feed-forward neural network receives the flattened output, and each training iteration employs backpropagation.

6. PROPOSED SYSTEM

Our suggested solution uses a convolution neural network, or CNN, to identify abnormal activity Accurately identifying the temporal data in the video is crucial for classifying anomalous behaviors. CNN is now mostly utilized to extract important features from every video frame. Only the most appropriate algorithm for this use is CNN. In order to successfully classify the input, features must be retrieved from CNN; as a result, CNN must be able to recognize and extract the appropriate characteristics from video frames. Video processing is used extensively in two main fields: security and research. These devices monitor live videos with the help of intelligent algorithms. The design of a real-time system needs to take into account several crucial factors, such as computational complexity and time. The system that uses a single algorithm with a comparatively low period complexity and uses less hardware resources while still yielding respectable results is particularly helpful for time-sensitive requests like bank theft detection, patient care systems, identifying and noting suspicious activity by train stations, etc. The current work describes a program that classifies whether or not a person's behavior is suspicious by analyzing real recordings of them in a testing setting. The sophisticated technology classifies abnormal head motions to stop them from happening again. It also shows when a student moves to another student's position or takes a different one. In conclusion, the system identifies student interactions and prevents students from sharing derogatory information with one another. Our research has aided in the creation of a system that can analyze live video of classrooms full of children and classify a student's behavior as suspicious or not.

Model's proposed work:

1. Data Collection: First, data is retrieved from various Websites and Social Media applications based on predetermined criteria.
2. Preprocessing: Next, in order to properly prepare our dataset, we will perform a number of pre-processing operations, including resizing, binary conversion, noise reduction, and gray scaling.
3. Noise reduction: The input video's processing that helps in denoising. Typically, filters such as Wiener, Kalman, average, and median are used to minimize noise.
4. Resizing: While remapping can be used to compensate for lens distortion or rotate a picture, resizing is required when we need to change the overall amount of pixels in an image.
5. Binary conversion: An image that contains only two distinct colors—typically black and white—is said to be binary. Bi-level and two-level images are other popular terms for binary images. This indicates that each and every pixel is stored as a single bit, or between 0 and 1.
6. Gray scaling: This technique converts a continuous-tone picture into one that a computer can easily work with.

7. Segmentation: The important procedure of isolating a digital image into several segments—that is, groupings of pixels, also identified as image objects—is known as image segmentation.6. Data Training: We use internet news data to construct fake and real-time training sets, which can be used with any machine learning classifier.
8. Feature extraction: This step of the dimensionality decrease process involves compacting and separating the initial collection of raw data into more manageable categories.
9. Classification: Using certain guidelines and instructions, classification is the process of grouping and classifying pixels or vectors inside an image.
10. Data Training: Using real-time and fabricated data from social media, we trained any machine learning classifier.
11. Machine learning testing: We feed the system a testing dataset and use an algorithm for machine learning to identify the appropriate activity.
12. Analysis: We compare the suggested system's accuracy with that of other current systems.

Convolution Neural Network (CNN) is the algorithm.

Phase 1: An image or video is provided as input.

Phase 2: After that, the input is subjected to numerous filters in order to produce a feature map.

Phase 3: To boost non-linearity, a ReLU function is then applied.

Phase 4: After that, every feature map has a pooling layer applied to it.

Phase 5: The method creates a single, lengthy vector by compressing the pooled images.

Phase 6: The vector is sent into an algorithm to create a fully connected artificial neural network in the following phase.

Phase 7: Uses the network to process the features. The last fully connected layer provides the class "voting."

Phase 8: For several epochs, training is done in both forward and backward propagation in this final step. This process is repeated until a neural network with training weights and feature detectors is well-defined.

noise is eliminated. Filtering is a crucial step in image

7. CONCLUSION

This work employs an object tracking-dependent semantics-based activity detection method. It makes use of the spatial relationship and motion characteristics of two objects. To identify the recommended activities of interest, the attributes are continuously compared to predetermined criteria. For real-time performance, the method is straightforward and does away with the need for training that comes with machine-learning-based techniques. In a natural setting, human behaviors are diverse and complex. This study developed the concept of suspicious action detection for security systems. About 95% accuracy is attained. In terms of processing time for a single image detection, we discovered that YOLOv3 performs better than Faster R-CNN. Only in a controlled environment can the current feature extraction technology produce results that are correct. Improved techniques for extracting features can be integrated enhance the outcomes. Nonetheless, there were still significant discrepancies in the comparison of the test results and the ground truth because of the limited quantity of data in the training set. Therefore, expanding the training dataset to include suspicious films of various activities and resolutions would be the future work to be done for development in order to acquire a better detection and make the model more practical. Real-time applications can also benefit from the development of more complex algorithms

8. REFERENCES

[1] "Inspection of suspicious human activity in the crowdsourced areas captured in surveillance cameras," by P. Bhagya Divya, S. Shalini, R. Deepa, and Baddeli Sravya Reddy, International Research Journal of Engineering and Technology (IRJET), December 2017.

[2] "Suspicious Movement Detection and Tracking of Human Behavior and Object with Fire Detection using A Closed Circuit TV (CCTV) cameras," Jitendra Musale, Akshata Gavhane, Liyakat Shaikh, Pournima Hagwane, and Snehalata Tadge, International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 5 Issue XII December 2017.

[3] Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018, U.M. Kamthe, C.G. Patil, "Suspicious Activity Recognition in Video Surveillance System."

[4] Detecting Abnormal Events in University Areas: Zahraa Kain, Abir Youness, Ismail El Sayad, Samih Abdul-Nabi, Hussein Kassem, International Conference on Computer and Application, 2018.

[5] "Abnormal event detection based on analysis of movement information of video sequence," by Tian Wanga, Meina Qia, Yingjun Deng, Yi Zhouc, Huan Wangd, Qi Lyua, and Hichem Snoussie, was published in Article-Optik, vol. 152, January 2018.