International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET

Abnormal Activity Detection Using CNN Method

Balbhim Laxman Lanke, Zarinabegam K. Mundargi
Department of Computer Engineering, Zeal college, Pune, Maharashtra, India

Article Info
Volume 9, Issue 3

Page Number : 593-598

Publication Issue :
May-June-2022

Article History
Accepted : 01 May 2022
Published: 07 May 2022

ABSTRACT

Anomaly Activity is the prediction of a suspicious activity from a picture or video. This project would include using neural networks to detect suspicious human activity from real-time CCTV video. Human Anomaly Activity is a central issue in computer vision that has been researched for over 15 years. It is significant due to the large number of applications that can benefit from Activity detection. Human pose estimation, for example, is used in applications such as video monitoring, animal tracking and behavior recognition, sign language identification, advanced human-computer interaction, and marker less motion recording. Low-cost depth sensors have disadvantages such as being restricted to indoor use, and their low resolution and noisy depth information make estimating human poses from depth images difficult. As a result, we want to use neural networks to solve these issues. Suspicious human activity detection in surveillance video is an active field of image processing and computer vision science. Human activities in sensitive and public areas such as bus stations, train stations, airports, banks, shopping malls, schools and colleges, parking lots, highways, and so on can be monitored using visual surveillance to detect terrorism, robbery, accidents and illegal parking, vandalism, fighting, chain snatching, violence, and other suspicious activities. It is extremely difficult to continuously track public places; thus, intelligent video surveillance is needed that can monitor human activities in real-time, classify them as normal or unusual, and generate an alarm. The majority of the analysis being conducted is on photographs rather than recordings. Furthermore, none of the papers reported attempt to use CNNs to detect suspicious activity.

Keywords: Preprocessing, Feature Extraction, Machine Learning, Convolutional neural Network.

I. INTRODUCTION

We plan to develop a real-time framework for detecting anomalous behavior of people in public places. Our application may be used to monitor areas where there is a chance of robbery or a shooting attack, such as malls, airports, and train stations. To train our system, we will use deep learning and neural networks. This model will then be implemented as a smartphone and desktop app that will accept real-time CCTV footage as input and send a warning to the administrator's computer if a suspicious activity is detected. Human anomaly behavior is associated with the identification of human body parts and probably the monitoring of their movements. Its real-world applications range from gaming to AR/VR, healthcare, and gesture recognition. In comparison to the image data domain, there has been relatively little research into applying CNNs to video classification. This is due to the fact that a video is more complex than a picture because it has another dimension - temporal. Unsupervised learning, which takes advantage of temporal dependencies between frames, has proven effective for video analysis. Some approaches to anomaly activity use CPUs rather than GPUs, allowing anomaly activity to run on low-cost hardware such as embedded systems and cell phones. Another emerging technology in computer vision is low-cost depth sensors. They are introducing new consoles, such as the Kinect for Xbox360. They are motion sensors that enable the user to communicate with the console without the use of a game controller by using just hand gestures. These are RGB-D sensors that use standardized light technology to obtain depth information. The depth values are determined by projecting an infrared light pattern onto a scene and measuring the distortion of the projected light pattern. However, these sensors are only suitable for indoor use, and their low resolution and noisy depth details make estimating human poses from depth images difficult.

II. LITERATURE SURVEY

Mohammad Sabokrou , Mahmood Fathy,“ Real-Time Anomaly Detection and Localization in Crowded Scenes”[1], In this paper, we propose a method for detecting and localising anomalies in crowded scenes in real time. Each video is defined as a collection of non-overlapping cubic patches and is represented using two descriptors: local and global. These descriptors capture video properties from various angles. We can differentiate between regular and abnormal behaviours in videos by integrating simple and low-cost Gaussian classifiers. The local and global features are focused on structure similarity between adjacent patches and unsupervised learning with a sparse autoencoder. On the UCSD ped2 and UMN benchmarks, experimental findings show that our algorithm is comparable to a state-of-the-art method, but much more time-efficient. The tests show that our system can identify and localise abnormalities in videos as soon as they occur.

Mahmudul Hasan Jonghyun Choi “Learning Temporal Regularity in Video Sequences”[2], Perceiving meaningful events in a long video sequence is a difficult problem due to the vague definition of "meaningfulness" as well as the clutter in the scene. We tackle this problem by learning a generative model for normal motion patterns (referred to as regularity) from multiple sources with very little supervision. We specifically recommend two methods based on autoencoders for their ability to function with little to no supervision. We start with the traditional handcrafted spatio-temporal local features and train a fully connected autoencoder on them. Second, as an end-to-end learning system, we build a completely convolutional feed-forward autoencoder to learn both the local features and the classifiers. Our model is capable of capturing regularities from several datasets. We test our methods qualitatively and quantitatively, demonstrating the learned regularity of videos in various aspects and competitive success on anomaly detection datasets as an application.

Jefferson Ryan Medel,“ Anomaly Detection in Video Using Predictive Convolutional Long Short-Term Memory Networks”[3], Because of the uncertainty of how certain events are defined, automating the identification of anomalous events within long video sequences is difficult. We address the problem by training generative models to detect anomalies in videos with minimal supervision. End-to-end trainable composite Convolutional Long Short-Term Memory (Conv-LSTM) networks that can predict the evolution of a video sequence from a limited number of input frames are proposed. The reconstruction errors of a series of predictions of irregular video sequences yielding lower regularity scores as they diverge further from the actual sequence over time are used to calculate regularity scores. The models employ a hybrid framework and investigate the impact of 'conditioning' on learning more realistic representations. The best model is chosen based on the accuracy of reconstruction and prediction. The Conv-LSTM models are tested qualitatively and quantitatively, with competitive performance on anomaly detection datasets. Conv-LSTM units have been demonstrated to be a useful method for modelling and predicting video sequences. Finally, the INSLR system employs an audio system in addition to text production to play the known movements. The system is put through its paces with a data collection of 80 words and sentences signed by ten separate signers. The experimental results show that our device has a 96 percent recognition rate.

Yong Shean Chong,“ Abnormal Event Detection in Videos using Spatiotemporal Auto encoder”[4], We present a quick method for detecting anomalies in video. Recent convolutional neural network applications have demonstrated the promise of convolutional layers for object detection and recognition, especially in photos. Convolutional neural networks, on the other hand, are supervised and require labels as learning signals. We suggest a spatiotemporal architecture for detecting anomalies in videos with crowded scenes. Our architecture is comprised of two major components: one for spatial feature representation and one for learning the temporal evolution of spatial features. Experiment findings on Avenue, Subway, and UCSD benchmarks confirm that our method's detection accuracy is comparable to state-of-the-art methods at a significant speed of up to 140 fps.

Steven Diamond Vincent Sitzmann,“ Unrolled Optimization with Deep Priors”[5], A wide range of problems at the heart of computational imaging, sensing, and low-level computer vision boil down to the inverse problem of extracting latent images that obey a prior distribution from measurements taken under a known physical image forming model. Handcrafted priors and iterative optimization approaches have traditionally been used to solve such problems. In this paper, we present unrolled optimization with deep priors, a principled paradigm inspired by classical iterative methods for infusing knowledge of image forming into deep networks that solve inverse problems in imaging. We demonstrate that instances of the system outperform the state-of-the-art for a broad range of imaging issues, including denoising, deblurring, and compressed sensing magnetic resonance imaging (MRI). Furthermore, we run experiments to demonstrate how the system is best used and why it outperforms previous approaches.

III. SYSTEM ANALYSIS

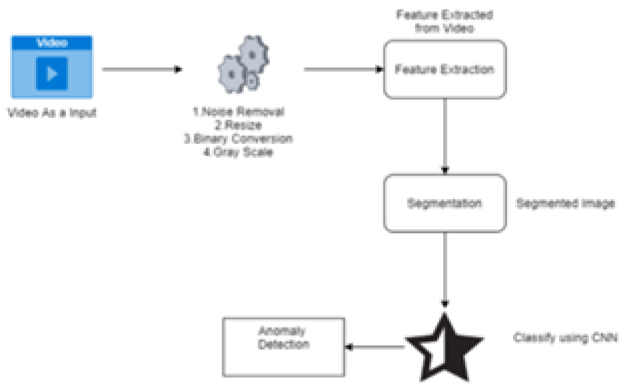


Fig System Architectuure

- **Modules:**

Pre-processing: - Although geometric transformations of images (e.g. rotation, scaling, translation) are known as pre-processing methods, the goal of pre-processing is an enhancement of the image data that suppresses unwanted distortions or improves certain image features necessary for further processing. - Possessiveness Image processing is the use of a digital computer to run an algorithm on digital images. Digital image processing, as a subcategory or field of digital signal processing, has many advantages over analogue image processing.

1. Read the Image
2. Image Resized (220,220, 3)/Resized (width, height, no. RGB channels)
3. Conversion of RGB to Grayscale
4. Identification of segmentation edges The Gaussian filter is used to remove noise.

Segmentation: It entails segmenting a visual input to facilitate image analysis. If we want to remove or identify something from the rest of the image, such as detecting an object in the background, we can divide the image into segments that can be processed further. This is often referred to as segmentation. Segments contain objects or parts of objects and are made up of groups of pixels known as "super-pixels."

Feature Extraction: Features in an image may be complex structures such as points, edges, or objects. The aim of feature extraction is to reduce the number of features in a dataset by generating new ones from existing ones (and then discarding the original features). The new reduced collection of features should then be able to summarize the majority of the details in the original set of features.

feature extraction starts from an initial set of measured data and builds derived values (features) intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps, and in some cases leading to better human interpretations. Feature extraction is related to dimensionality reduction.

Classification: Because of their high precision, CNNs are used for image detection and recognition. The classification convolutional neural network has a three-dimensional structure, with each collection of neurons analyzing a particular region or "function" of the picture. Each group of neurons in a CNN focuses on a different part of the picture. The algorithm examines smaller portions of the images. The final result is a vector of probabilities that predicts how likely each feature in the image is to belong to a class or group.

CNN (Convolutional Neural Network): CNN (convolutional neural network) is a form of deep learning neural network. In a nutshell, consider CNN to be a machine learning algorithm that can take an input image, assign importance (learnable weights and biases) to various aspects/objects in the image, and distinguish one from the other. CNN operates by extracting features from videos. A CNN is made up of the following components:

1. A grayscale image serves as the input layer.
2. The output layer, which consists of binary or multi-class labels.
3. Convolution layers, ReLU (rectified linear unit) layers, pooling layers, and a completely connected Neural Network comprise the hidden layers.

It is critical to understand that ANNs, or Artificial Neural Networks, comprised of multiple neurons, are incapable of extracting features from images. This is where a convolution and pooling layer combination comes into play. Similarly, the convolution and pooling layers are incapable of classification, necessitating the use of a fully connected Neural Network.

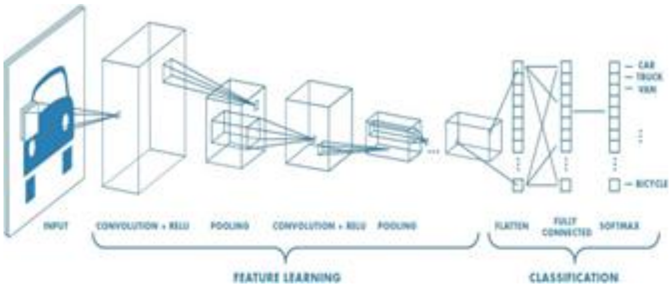
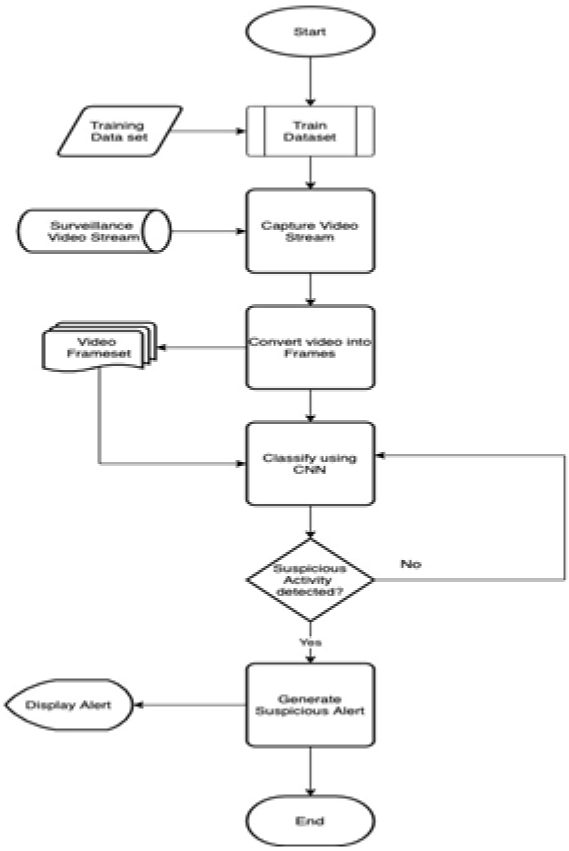


Fig - Illustrates the CNN process from input to Output Data.

IV. METHODOLOGY

In this project we use CNN algorithm and image processing techniques. Gray scale method use in image preprocessing for frames conversion. After frame conversion each frame compared with train dataset and detects the suspicious activity.

V. FLOWCHART DIAGRAM



VI. ALGORITHM FLOW

Step 1: Start
Step 2: Train dataset
Step 3: Input video
Step 4: Video converted into frames and each frame classify using CNN algorithm
Step 5: If suspicious activity detected then send notification
Step 6: Stop

Software specification

- Anaconda framework.
- Python language.

VII. OUTPUT RESULT

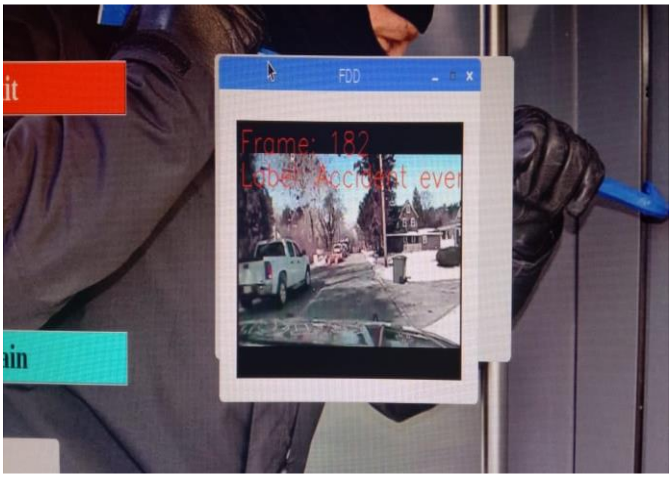


Fig : output 1

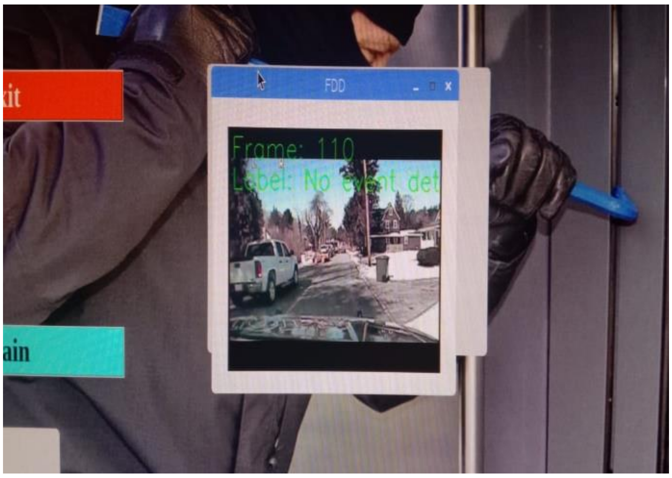


Fig: output 2

VIII. CONCLUSION

A device that processes real-time CCTV footage to detect Anomaly activity would aid in creating better protection and reducing the need for human interference. Great strides have been made in the field of human anomaly Operation, allowing us to better serve the various applications that it is capable of. Furthermore, research in related fields such as Activity Tracking will significantly improve its efficient use in a variety of fields.

IX. REFERENCES

[1]. Eralda Nishani, Betim Cico : "Computer Vision Approaches based on Deep Learning and Neural Networks" Deep Neural Networks for Video Analysis of Human Pose Estimation- 2017 6th MEDITERRANEAN CONFERENCE ON EMBEDDED COMPUTING (MECO), 11-15 JUNE 2017, BAR, MONTENEGRO

[2]. Naimat Ullah Khan , Wanggen Wan : "A Review of Human Pose Estimation from Single Image"- 978-1-5386-5195-7/18/ 2018 IEEE

[3]. Qiuhui Chen, Chongyang Zhang, Weiwei Liu, and Dan Wang, "Surveillance Human Pose Dataset And Performance Evaluation For Coarse-Grained Pose Estimation", Athens 2018.

[4]. Baole Ai, Yu Zhou, Yao Yu : "Human Pose Estimation using Deep Structure GuidedLearning"-978-1-5090-4822-9/172017IEEEDOI10.1109/WACV.2017.141

[5]. Zhe Cao, Tomas Simon, Shih-En Wei, Yaser Sheikh The Robotics Institute, Carnegie Mellon University" Realtime Multiperson 2D Pose Estimation using part affinity fields" - 1063-6919/17 2017 IEEE DOI 10.1109/CVPR.2017.143

[6]. Hanguen Kim, Sangwon Lee, Dongsung Lee, Soonmin Choi, JinsunJu and Huyun Myung "Real-Time Human Pose Estimation and Gesture Recognition fromdepthImagesUsingSuperpixelsandSVMclassifier."-Sensors2015,15, 12410-12427; doi:10.3390/s150612410

[7]. Tripathi, Rajesh and Jalal, Anand and Agarwal, Subhash(2017). "Suspicious HumanActivityRecognition: aReview". ArtificialIntelligenceReview. 50.10.1007/s10462017-9545-7.

[8]. E. Eksioglu. Decoupled algorithm for MRI reconstruction using nonlocal block matching model: BM3DMRI. Journal of Mathematical Imaging and Vision, 56(3):430–440, 2016.

[9]. S.Wang,Z.Su,L.Ying,X.Peng,S.Zhu,F.Liang,D.Feng,andD.Liang. Accelerating magnetic resonance imaging via deep learning. In Proceedings of the IEEE International Symposium on Biomedical Imaging, pages 514–517, 2016.

[10]. L.Xu,J.Ren,C.Liu,andJ.Jia. Deep convolutional neural network for image deconvolution. In Advances in Neural Information Processing Systems,pages 1790–1798, 2014.

[11]. Y. Yang, J. Sun, H. Li, and Z. Xu. Deep ADMM-Net for compressive sensing MRI. In Advances in Neural Information Processing Systems, pages 10–18, 2016.

[12]. Z. Zhan, J.-F. Cai, D. Guo, Y. Liu, Z. Chen, and X. Qu. Fast multiclass dictionaries learning with geometrical directions in MRI reconstruction. IEEE