International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 04 | Apr 2022 www.irjet.net p-ISSN: 2395-0072

Crime Detection using Machine Learning

Christina Grace Nandigam, Nayana Ganesh Joshi, Swaranjali Bichukale, Vinisha Gomare
Guide: Prof. R. H. Bhole

Department of Information Technology, Zeal College of Engineering and Research, Affiliated with Savitribai Phule Pune University, Narhe, Pune-411041, Maharashtra, India

---***---

Abstract - *Criminal Activity detection involves studying the body part or joint locations of a person from an image or a video. This project will involve the tracking of dubious human activity from live feeds of video surveillance by implementing CNN. Human Activity Analysis is an important issue that has been researched for years. It is necessary because of the unmitigated number of applications that can benefit from such tracking. For instance, human posture analysis can be used in applications that include animal behaviour understanding, surveillance using videos, sign language detection, and progressive human-computer interaction. Depth sensors have various drawbacks; they are limited to sedentary use, have very low resolution, contain noisy depth information, etc. These drawbacks make it difficult to estimate human poses from depth pictures. Neural networks can be used to overcome such problems. An active field in image processing research is human activity tracking and analysis from ocular observation. Through ocular observation, human actions can be supervised in jam-packed areas like stations, banks, malls, airports, roads, schools, colleges, parking lots, etc. to thwart dubious, criminal actions such as robbery, mob activity, and non-legal parking, violence, and other suspicious activities. It is futile to monitor such areas continuously, therefore intelligent ocular observation is necessary which can monitor human activities live and group them as normal and suspicious actions, and can trigger an alert.*

Key Words: *CNN, Machine Learning, pre-processing, Classification, deep learning.*

1. INTRODUCTION

The plan is to build an application for the detection of dubious activity among people in areas of public interest places in real-time. Through ocular observation, human actions can be supervised in jam-packed areas like stations, banks, malls, airports, roads, schools, colleges, parking lots, etc. to thwart dubious, criminal actions such as robbery, mob activity, and non-legal parking, violence, and other suspicious activities. Deep learning and neural networks are going to be used to train the datasets in this system. This model will then be implemented as user-friendly software which will take the live feed from video surveillance as input and trigger an alert on the user's device if some dubious activity is found. Human activity analysis is related to identifying human body parts and possibly tracking their movements. Real-life software of it varies from gaming to AR/VR, to healthcare and gesture recognition. In comparison to the domain of image data processing, there is very little amount of work on using CNNs for video analysis. It is because videos are more complex compared to images since they have another dimension to them — temporal. Unsupervised learning exploits temporal dependencies between frames and has proven successful for video analysis. Some human activity analysis programs use central processing units instead of graphical processing init so that the software can run on affordable hardware like mobile phones and embedded systems. Easily affordable sensors that can analyse the depth are another sort of technology in computational foresight. They are present in gaming consoles like Move for PlayStation. These motion sensors detect motion by simple hand gestures and do not need game controllers. They use structured light technology to access depth information. The depth values are inferred by the structured light sensors by the projection of an infrared light pattern onto a scene and analyzing the bending of the projected light pattern. These sensors, however, cannot be used on a large scale, and noisy depth information and low resolution render them incapable to analyze human postures from depth images.

2. PROBLEM STATEMENT

One of today's biggest problems is the manual analysis of readily available information credited to today's technological advances. CCTVs, drones, satellite data, wearables, etc, provide a large amount of diverse data, and extracting strategic knowledge manually from this data is becoming more a problem than a solution. Automatic solutions are a critical necessity. This problem requires immediate solutions and this project will be the base for it by detecting suspicious/dubious activity from live feeds of CCTVs.

3. LITERATURE SURVEY

Bogden Ionescu, Razvan Roman, Marian Ghenescu, Marian Buric, and Florin Rastoceanu's paper *Artificial Intelligence Fights Crime and Terrorism at a New Level* showed Artificial Intelligence (AI) as a new angle for delivering results with a human-grade precision. This paper served as the base paper for this project, providing various models and aspects to study and research. The only limitation of this paper had been that it did not have live video feed tech implemented yet.

Achini Adikari, Daswin De Silva, Damminda Alahakoon, and Xinghuo Yu's paper *Suspicious Human Activity Recognition: a Review* explored all the areas where a visual-based detection system can be used. For instance, it could be used in old age homes or hospitals as wearables apart from the usual CCTV monitoring. Where the traditional CCTV approach or basic wearable would not be an immediate help if the wearer is unconscious, a motion-detecting wearable would trigger an immediate alert. The paper did not provide any modules for such implementation and served only as a literature review.

Betim Cico and Eralda Nishani's paper *"Computer Vision Approaches based on Deep Learning and Neural Networks" Deep Natural Networks for Video Analysis of Human Pose Estimation* explores the implementation of neural networks, specifically CNN, for HAR from the analysis of videos. Neural networks are a part of deep learning, adapted from the concept of the human nervous system in the way that they send signals in the same way as human neurons do. These networks have node layers that contain an input layer, an output layer and at least two hidden layers. Each node is like an artificial neuron that has a weight and a threshold and that connects to another node. If the result of any node crosses the value at the threshold, the node gets activated after which it sends data to the next node. The paper posed three questions:

1. Since CNNs work for the estimation of human postures, by adding or changing what in their architecture would the results be improved?
2. What would be the output if RNNs were used instead of CNNs for this estimation?
3. How can unsupervised learning make the most of the large chunk of unclassified data that exists online?

Baole Ai, Yu Zhou, Yao Yu, and Sidan Du's paper *Human Pose Estimation using Deep Structure Guided Learning* shows more about the advantage of using CNN for human pose estimation. Human activity recognition is the process of classification of sequences of accelerometer data recorded by devices into well-defined movements. Convolutional Neural Networks, or CNNs, were initially developed for problems involving image classifications in which the model learned the internal rep of a 2D input in a process called feature learning. The same process could now be used on 1D sequences of data such as HAR. The model learns how to extract features from sequences of observations and how to map the internal features to different activity types. The paper, though wasn't very observation centric, gave a vivid description of how CNNs can be used in a visual-detection system.

4. SYSTEM ARCHITECTURE

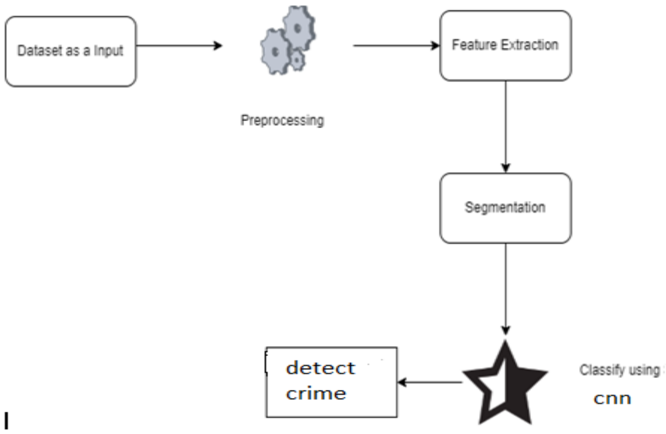

Figure 1. System Architecture

5. ALGORITHM

CNN (CONVOLUTIONAL NEURAL NETWORK)

When we talk about deep learning, a CNN, or a convolutional neural network which is also called ConvNet, is a class of a bigger set of networks known as Artificial Neural Network (ANN). It is most commonly used for the analysis of visual data. These are called SIANN, i.e., Space Invariant Artificial Neural Networks, or simply Shift Invariants, depending on the shared-weight architecture of the kernels of convolution that move along input features and present feature maps which are translation equivariant responses. On the contrary, most CNNs are equivariant only to translation contrasted to invariant. They contain applications in video and image recognition, image segmentation and classification, recommender systems, medical image analysis, financial time series, brain-computer interfaces and natural language processing.

Convolutional neural networks are consistent versions of multilayer perceptrons. Such perceptrons usually imply completely linked networks, i.e., each neuron in a layer is linked to every neuron in the next layer. The complete linkedness of these networks makes them likely to overfit data. Common ways of consistency or prevention of overfitting data include: disciplining parameters at the time of training (like weight decay) or clipping links (dropout, skipped connections, etc.). These neural networks have a different take on consistency or regularization: they take the benefit of the hierarchical pattern in assembling and data patterns of growing complications using simpler and smaller patterns illustrated in their filters. Thus, on a scale of complexity and connectivity, the networks on the lower extremity are CNNs.

CNNs were developed with biological processes in mind, in the sense that the linking pattern among neurons mirrors the nervous system of an animal, particularly the visual cortex of an animal. A particular cortical neuron reacts to a stimulus only in a defined region of the visual field called the receptive field. These fields of different neurons somewhat overlap in a way that they cover the whole visual field.

CNNs utilize comparatively less pre-processing in relation to other algorithms when it comes to image classification, i.e., the network picks up how to optimize and hone the kernels (or filters) through automated learning, which wouldn't have been possible in traditional algorithms where these kernels are engineered manually. Such independence from human expertise and previous knowledge in feature extraction is a great advantage.

6. CONCLUSIONS

A system to process real-time CCTV footage to detect any criminal human activity will help to create better security with less human intervention. Great strides have been made in the field of human criminal activity detection, which enables us to better serve the myriad applications that are possible with it. Furthermore, research in similar areas such as Activity Tracking would greatly amplify its productive application in various fields.

7. 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, MON-TENEGRO

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 Guided Learning"- 978-1-5090-4822-9/17 2017 IEEE DOI 10.1109/WACV.2017.141

5. Zhe Cao, Tomas Simon, Shih-En Wei, Yaser Sheikh The Robotics Institute, Carnegie Mellon University" Real-time 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 from depth Images Using Superpixels and SVM classifier."- Sensors 2015, 15,

12410-12427; doi:10.3390/s150612410

7. Tripathi, Rajesh and Jalal, Anand and Agarwal, Subhash (2017)." Suspicious Human Activity Recognition: A Review". Artificial Intelligence Review. 50.10.1007/s10462-017-9545-7.

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, and D. 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, and J. Jia. Deep convolutional neural network for image deconvolution. In Advances in Neural Information Processing Systems, pages 1790–1798, 2014.