SUSPICIOUS HUMAN ACTIVITY RECOGNITIONAND ALARMING SYSTEM USING CNN AND LSTM ALGORITHM

Dr. Keerthika V[1],A Abhiram[2], Deepak Kumar[3], K.V Sai Likhita[4] ,Leela S[5]

[1] Associate Professor, Computer Science and Engineering, DSATM

[2][3][4][5] *Final Year Students, Computer Science and Engineering, DSATM*

Abstract

Traditional pattern recognition systems have made significant improvement in recent years. With deep learning methods' growing popularity and success, utilising these approaches to understand human behaviours in mobile and wearable computing settings has gotten a lot of interest. A deep neural network that blends convolutional layers with long short-term memory (LSTM) is proposed in this research. With a few model parameters, this model could automatically extract activity features and classify them. The LSTM is a recurrent neural network (RNN) variation that is better suited to handling temporal sequences. The raw data acquired by the sensors was fed into a two-layer LSTM followed by convolutional layers in the proposed architecture. It can not only extract activity features adaptively, but it also has less parameters and a higher accuracy. The CNN classifier, which should be used alone, and the LSTM models, which should be usedin series and with the feed data, should be used together.

Keywords- Human activity recognition · Adaptive video compression · Vision-based human activity recognition · Anomaly detection

---

I. Introduction

Anomaly detection in security systems is one of the instances in which human activity recognition can be used. An activity recognition system should be able to recognise the basic tasks that a human performs on a daily basis. Because of the complexity and diversity of human activities, achieving high accuracy for recognition of various activities is difficult.

Different methodologies unique to the application are used to build activity models required for the identification and classification of human activities. Video processing and machine learning groups are drawn to this research subject because it has applications in a variety of fields, including medicine and health care, human–computer interaction, crime investigation, and security systems.

Human activity recognition has applications outside of healthcare and security.Behavioral biometrics, video analysis, animation, and synthesis are all examples of this ongoing and open research field in computer vision. Sensors are used to interpret human behaviours involving gestures and motions of the human body. The identification of human actions requires theseinterpretations.

II. Related Work

D. G. Shreyas, S. Raksha & B. G. Prasad (2020) [1] proposed that by using the intermediate result of adaptive video compression an accurate, real-time anomaly detection system is implemented. The proposed method outperforms other existing systems in terms of accuracy and timeliness.

Santosh Kumar Yadav, Kamlesh Tiwari, Hari Mohan Pandey & Shaik Ali Akbar (2021) [2] presented a privacy-preserving activity recognition and fall detection system using a single Kinect (v2) sensor and ConvLSTM. The proposed system derives geometrical and kinematic features and passes them along with the raw skeleton coordinates into deep learning networks.

Djamila Romaissa Beddiar, Brahim Nini, Mohammad Sabokrou & Abdenour Hadid (2020) [6] proposed a classification according to several criteria. It initially discusses the different applications of HAR, and the major objectives intended by these systems. Then, it presents an analysis of the used approaches in the state of the art, as well as the means used in their validation.

Giovanni Ercolano; Daniel Riccio; Silvia Rossi [3] presented a CNN-LSTM model for activity recognition working on a matrix representation of theskeleton joints

.

K.Ishwaryaa and A.Alice Nithyab (2021)[4] produces the detail study of the proposed model review. In that analysis various time series methods deeply analysed to correctly identify the daily activity effectible.

Jaspal Kumar ,M.Kulkarni Davya gupta (2016)[5] proposed an intelligent video surveillance system to detect and track the human body and to recognize the type of the daily human movement and activity is introduced.

III. Methodology

We propose a deep learning model composed of *convolutional* and *Long Short-Term Memory* recurrent layers, which can automatically learn local features and model the relation between features.The implementation of such a system requires the videos from surveillance cameras to generate frames which are then used for feature extraction. Another methodology called the adaptive video compression can be intermediately implemented wherein objects of interest are identified and mapped for each frame before compressing the video. *Adaptive video compression* is a unique compression technology that compresses only the least significant parts of the video, retaining the objects of interest. Identification of objects of interest prior to activity recognition helps achieve a higher recognition performance making theanomalous human activity recognition system more efficient.

Convolutional Neural Networks (CNN) are great for image data and Long-Short Term Memory (LSTM) networks are great when working with sequence data but when you combine both of them, you get the best of both worlds and you solve difficult computer visionproblems like video classification.

A.TENSORFLOW

We'll use a CNN to extract spatial characteristics at a specific time step in the input sequence (video), then an LSTM to find temporal relationships between frames. ConvLSTM and LRCN are the two architectures we will utilise to combine CNN and LSTM. TensorFlow may be used for each of these approaches. TensorFlow is an open source machine learning platform that runs from start to finish. It has a large, flexible ecosystem of tools, libraries, and community resources that allow academics to advance the state-of-the-art in machine learning and developersto quickly construct and deploy ML applications.

B.KERAS

Keras is a lightweight deep learning Python package that may be used with Theano or TensorFlow. It was created to facilitate the implementation of deep learning models for research and development as simple as feasible. Given the underlying frameworks, it operates on Python 2.7 or 3.5 and can run on both GPUs and CPUs. It's under the MIT licence, which allows for a lot of flexibility. François Chollet, a Google developer, created and maintains Keras basedon four principles:

Modularity: A model can be understood as a sequence or a graph alone. All the concerns of a deep learning model are discrete components that can be combined in arbitrary ways.

Minimalism: The library provides just enough to achieve an outcome, no frills and maximizing readability.

Extensibility: New components are intentionally easy to add and use within the framework, intended for researchers to trial and explore new ideas.

Python: No separate model files with custom file formats. Everything is native Python.

Keras does not do low-level operations like tensor products and convolutions on its own; instead, it uses a back-end engine. Despite the fact that Keras supports many back-end engines, TensorFlow is its major (and default) back end, and Google is its principal backer. TensorFlow includes the Keras API as tf.keras, which, as previously stated, will become the primary TensorFlow API as of TensorFlow 2.0. The Model is the most basic data structure in Keras. The Sequential model and the Model class used with the functional API are the two major types of models accessible in Keras. Sequential Keras models The Sequential model is a linear stack of layers with relatively simple layer descriptions.

C. LSTM (Long-Short Term Memory)

Long-Short Term Memory is an artificial recurrent neural network (RNN) architecture we use in deep learning. RNN is a class of artificial neural network where connections between nodes form a directed/in directed graph along a temporal sequence, exhibiting temporal dynamic behavior. It basically uses memory (internal state) to process variable length sequences of the input. Unlike the normal feedforward neural networks which do not form a cycle, LSTM has feedback connections. Our project needs to process video and speech apart from just images; hence LSTM is well suited for this since it can process not only single data points but also the sequences of data. Our main focus which is basically on detecting and capturing the anomaly activities, LSTM which is well applicable for unsegmented tasks such as anomaly detection or intrusion detection is much preferred.

Here we basically need a system to store and process the data and handle the information accordingly. A common LSTM unit usually consists of a cell, an input gate, an output gate and a forget gate. The cell remembers the values over the arbitrary time intervals, whereas the 3 gates regulate the flow of the information in and out of the cell. Since, classifying, processing and making predictions based on the time series is an important part by which the anomalies areto be detected, LSTM plays a vital role in making thispossible.

The LSTM cell can process data sequentially and keep its hidden state through time. This algorithm helps us to connect the previous information to connect it to the present task which is happening or predicted to happen. Using its default behavior to remember things for long without any struggle, is a major advantage we have in using this algorithm to our project.

$$
\begin{aligned}
f_t &= \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \\
i_t &= \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \\
o_t &= \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \\
\tilde{c}_t &= \sigma_c(W_c x_t + U_c h_{t-1} + b_c) \\
c_t &= f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \\
h_t &= o_t \circ \sigma_h(c_t)
\end{aligned}
$$

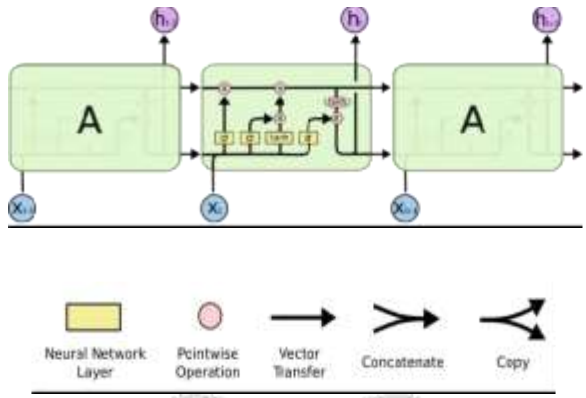

Fig.1 Working of LSTM Algorithm

In this Algorithm, firstly the decision of throwing away or keeping the information is done by the forget gate. It outputs a number between 0 and 1 for each number in cell state where 0 represents to completely get rid of info and 1 represents to keep that info completely. In the next step the input layer will decide what values to update and a vector for new candidate state is created and both are combined to create an update to the state. Now we update the old state. We multiply it with old state, forgetting things we decided to forget before. Finally, we decide what we output based on the cell state.

D. CNN

Convolutional neural network models were created for image classification challenges, and feature learning is the process by which the model learns an internal representation of a two-dimensional input.

This similar approach can be used to recognise human activity from one-dimensional sequences of data, such as acceleration and gyroscopic data. The model learns how to extract features from observation sequences and how to map internal features to various activity kinds.

The advantage of utilising CNNs for sequence classification is that they can learn directly from raw time series data, eliminating the need for domain knowledge to manually build input features. The model should be able to learn an internal representation of the time series data and, in theory, perform similarly to models trained on a version of thedataset with artificial features.

This section is divided into 4 parts; they are:

1. Load Data
2. Fit and Evaluate Model
3. Summarize Results
4. Complete Example

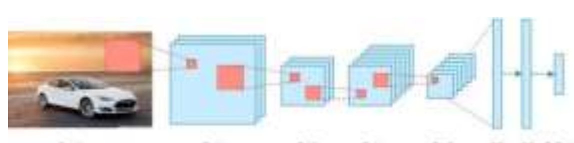

Fig.2 Working of CNN Algorithm

Multiple layers of artificial neurons make up convolutional neural networks. Artificial neurons are mathematical functions that calculate the weighted sum of various inputs and output an activation value, similar to their biological counterparts. Each layer creates many activation functions that are passed on to the next layer when you input an image into a ConvNet.

Basic features such as horizontal or diagonal edges are usually extracted by the first layer. This information is passed on to the next layer, which is responsible for detecting more complicated features like corners and combinational edges. As we go deeper into the network, it can recognise even more complex elements like objects, faces, and so on.

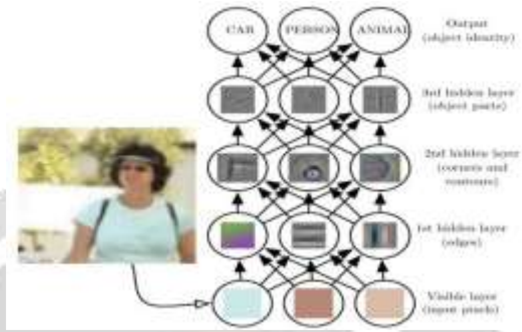

Fig.3 Illustration of human activity detection model

IV. Conclusion

Theft, accidents, graffiti, fighting, chain snatching, crime, and other suspicious actions have become increasingly common in recent years. We can't rely on traditional procedures that require a human being to monitor the system on a constant basis. Our project primarily addresses this issue, which necessitates the development of an automated system that detects unusual behaviours in the environment and alerts the end user. We combine adaptive video compression with the identification of items of interest prior to activity recognition to improve recognition performance and make our proposed system moreefficient.

V. Future Scope

There is a growing call for for set up of surveillance cameras in all public locations inclusive of streets because of the upward push in crime rates. Responsibility to cut down the criminal activities does now no longer cease with the set up of CCTV cameras; there ought to be a mechanism to offer instant help to the victims of crime together with instant action against the criminals. This can happen most effective with consistent and cautious tracking of the video surveillance which ultimately calls for manpower. Thus, developing a real-time automated machine to apprehend anomalous human activities will be a superb solution to the above-cited problem. This concept can be extended to numerous public locations, precise to the application surroundings like schools, colleges, airports, bus stops, hospitals and railway stations primarily based totally on their precise requirements. It can be integrated to right away call an ambulance while road accidents are detected. Thus, using the intermediate result of adaptive videocompression an accurate, real-time anomaly detectionmachine is implemented.

References

[1] D.G.Shreyas, S. Raksha & B. G. Prasad SN Computer Science volume 1, Article number: 168 (2020) “Implementation of an Anomalous Human Activity Recognition System”

[2] Santosh Kumar Yadav, Kamlesh Tiwari, Hari Mohan Pandey & Shaik Ali Akbar Soft Computing (2021)“Skeleton-based human activity recognition using ConvLSTM and guided feature learning”

[3] Giovanni Ercolano;Daniel Riccio;Silvia Rossi2017 26th IEEE International Symposium on Robotand

Human Interactive Communication (RO-MAN)Year: 2017 | Conference Paper | Publisher: IEEE "Combining CNN and LSTM for activity of daily living recognition with a 3D matrix skeletonrepresentation"

[4] K.Ishwaryaa , and A.Alice Nithyab (2021)"Human activity recognition: A review"

[5] Mohanad Babiker, Othman Omran Khalifa, Aisha Hassan Abdalla Hashim" Automated Daily Human Activity Recognition for Video Surveillance Using Neural Network"

[6] Djamila Romaissa Beddiar, Brahim Nini, Mohammad Sabokrou & Abdenour Hadid ,Multimedia Tools and Applications volume 79, pages 30509–30555 (2020) "Vision-based human activity recognition"

[7] Department of Computer Science and Engineering, University of Ioannina, Ioannina, Greece| Computational Biomedicine Laboratory, Department of Computer Science, University of Houston, Houston, TX, USA "A Review of Human Activity Recognition Methods"

[8] Li Wei and Shishir K. Shah Computer Science Department, University of Houston, 3551 Cullen Blvd., Houston, TX 77204, U.S.A "Human Activity Recognition using Deep Neural Network with Contextual Information"

[9] Wenchao Xu; Yuxin Pang; Yanqin Yang; Yanbo Liu "Human Activity Recognition Based On Convolutional Neural Network"

[10] Florenc Demrozi; Graziano Pravadelli; Azra Bihorac; Parisa Rashidi "Human Activity Recognition Using Inertial, Physiological and EnvironmentalSensors: A Comprehensive Survey"