e-ISBN: 978-93-6252-521-5
IIP Series

ANOMALY HUNTER: YOLOV3 ADVANCED DEEP LEARNING MODEL FOR HUMAN ABNORMAL DETECTION

Abstract

The identification of aberrant human behaviour is an important component in the process of protecting the health and safety of persons as well as the preservation of the safe environment in public areas. Advanced algorithms such as YOLO (You Only Look Once) and Convolutional Neural Networks (CNN) are utilised by emerging technologies such as human behaviour and anomaly detection. These algorithms are utilised to extract characteristics, manage temporal relationships, and enhance the accuracy and efficiency of human behaviour detection systems. Examples of such algorithms are the YOLO and CNN algorithms. This approach was designed with the purpose of being utilised in real-world contexts, and more specifically for recognising suspicious behaviours in surveillance video collected by closed circuit television (CCTV) cameras. The technique was developed with the intention of helping to identify suspicious actions.

This model takes use of the object identification approach known as YOLOv3 in order to recognise human actions and abnormalities contained in the video data. Additionally, a Convolutional Neural Network (CNN) is utilised in order to extract action characteristics from each tracked trajectory. This is done after the previous step has been completed. In the end, but certainly not least, a You Only Look Once (YOLOv3) is utilised in order to construct a model for the purpose of recognising abnormal behaviours. This model enables the prediction of abnormal actions that are carried out by individuals. To summarise, the functioning of

Authors

Dr. P. Deepan
Associate Professor & Head
Department of Computer Science & Engineering
St. Peter’s Engineering College
Hyderabad, India.
deepanp87@gmail.com

Dr. B. Rajalingam
Associate Professor & Head
Department of AI&DS
St. Martin’s Engineering College
Hyderabad.
rajalingam35 @gmail.com

Dr. R. Santhoshkumar
Associate Professor & Head
Department of Computer Science & Engineering
St. Martin’s Engineering College
Hyderabad.
santhoshkumar.aucse@gmail.com

Mr. N. Arul
Assistant Professor
Department of Computer Science & Engineering
St. Peter’s Engineering College
Hyderabad, India.
arulthala82@gmail.com

this technology entails getting video information as input from security cameras and then employing a variety of complicated algorithms in order to identify and categorise anomalous actions. This is done in order to prevent and detect any potential threats. By utilising YOLO and CNN for the purpose of feature extraction and anomaly identification, the system is able to accurately recognise and anticipate abnormal types of human activity. This is made possible by the usage of these two techniques.

Keywords: Human behaviour; abnormal activity; abnormal detection; deep learning; YOLO and LSTM;

I. INTRODUCTION

The study of human behaviour and the identification of abnormalities is currently the subject of a substantial amount of investigation. Within the scope of this investigation, the objective is to do an automated analysis and recognise patterns of activity that are not characteristic of human beings [1-3]. The continual development of computer vision and deep learning techniques has led to the creation of complex algorithms. Complex algorithms have been developed as a result. With the help of these algorithms, abnormal activity may be identified and recognised in a manner that is not only accurate but also efficient. The phrase "human behaviour" refers to a wide variety of behaviours, such as movements, gestures, and motions, that are shown by people in their day-to-day existence. These behaviours include but are not limited to them. The study of human conduct, in especially its comprehension and analysis, has a wide range of applications, including those in the fields of public safety, healthcare, and surveillance [4]. The presence of abnormal behaviour, on the other hand, almost always implies the presence of possible dangers, threats, or crises; thus, the identification and recognition of such behaviour is essential for the production of effective decisions and the prompt implementation of preventative measures [5]. Based on the information shown in Figure 1, we have discussed a few of the deviant behaviours that are exhibited by people.

Figure 1: Sample Human Abnormal Behaviours

Within the realm of computer vision, the YOLO approach is a well-known object detection method that has gained a significant amount of popularity over the course of the past few years [6]. In addition to allowing for the detection of things in real time, it has shown to be an effective tool for the study of human conduct, with favourable outcomes. A single neural network is utilised by YOLO in order to split an input image into a grid. Following this, it proceeds to produce direct predictions about bounding boxes and class probabilities. Conv2DNet, on the other hand, is a convolutional neural network design that is extensively utilised in deep learning for the purpose of video analysis [7]. In particular, it is very effective in learning spatial hierarchies of features. This is accomplished by the use of several layers of convolutional and pooling algorithms. There are many different computer vision tasks that have been recognised as having applications for Conv2DNet. One of these tasks is the recognition of deviant conduct.

When coupled, YOLO and Conv2DNet provide a foundation that is greatly enhanced, which in turn allows for the analysis of human behaviour and the identification of problems. Conv2DNet takes use of the capabilities of deep learning in order to uncover and categorise abnormal behaviour patterns. YOLO, on the other hand, makes it possible to detect objects in a rapid and accurate manner. Through the integration of a variety of approaches, researchers and practitioners have the ability to develop robust systems that are capable of analysing video data, identifying abnormal conduct, and providing substantial insights for decision-making and action [8]. There are many other systems that are capable of doing all of these things. A compelling demonstration of the advancements that have been made in the disciplines of computer vision and deep learning techniques is the utilisation of YOLO and Conv2DNet in the study of human behaviour and the identification of anomalies. One of the ways in which these technologies contribute to better security, safety, and situational awareness across a range of domains is that they make it feasible to spot abnormal behaviour patterns in real time and with a high degree of accuracy [9].

The following is a summary of the work that pertains to the research that was carried out: The results of prior investigations that were pertinent are given and discussed in Section 2, which may be found here. Within the scope of Section 3, we conduct an investigation into the detection system of human anomalous behaviours, with a particular emphasis on the YOLOv3 detection model of differentiating activities. Utilising an analysis of the data generated by the network, this model is able to spot and identify any anomalous behaviours that may occur. Both the experimental setup that was used to evaluate the suggested approach and the data that were gathered from that configuration are discussed in Section 4. This section also includes the sections that describe the data. This section presents a summary of the findings, and it also discusses a number of prospective areas that may benefit from further investigation.

II. LITERATURE SURVEY

This section will provide a comprehensive discussion of the key research methods that were applied throughout the course of this examination. The method for identifying falls that was suggested by Thomas Gatt and colleagues [10] is founded on principles that are derived from the properties of the form under consideration. The first thing that they do is convert the video frames into intensity images. After that, they search for the largest thing in the image and assume that this is the object that they are looking for in their research. They then use a rule-based categorization to assess whether or not the activity is abnormal based on whether or not there was a change in the shape of the object of interest within the image. This is done in order to identify whether or not the behaviour is abnormal. Shih-Chung Hsu and colleagues [11] describe a method for automatic fall detection that is based on depth sensors. This solution is pretty comparable to the one that is shown here. They use the Fisher Vector (FV) representation to store information about the Curvature Scale Space (CSS), which allows them to enhance the detection of falls in depth recordings. Initial steps involve the segmentation of the human body and the extraction of silhouettes. These silhouettes are then utilised in the following computation of shape features that subsequently takes place. The attributes are then converted into their FV representation when that step has been completed. Following this, they make use of a one-versus-all support vector machine (SVM) classifier in order to detect autumn activities.

Yuan Cao and colleagues [12] devised a system that would monitor and assess the movement patterns of elderly persons in order to determine when they were standing about doing nothing. An algorithm known as Generalised Sequential Patterns (GSP) is utilised by their system in order to maintain a watchful check on the individuals and search for the preset patterns of recurrent actions that are indicative of loitering. Yixue Hao and colleagues [13] have developed a method that can identify loitering based on pedestrian activity regions that fall into one of three categories. This method may be used to identify lingering patterns. Sector loitering, ellipse loitering, and rectangle loitering — these are the categories that fall under this category. In order to determine the type of loitering that is carried out, trajectory maps and an examination of the path taken by the suspected target are taken into consideration. In order to determine the identity of the moving item, they make use of Gaussian mixture models instead. After a predetermined length of time has passed, they consider an activity to be loitering if it is carried out in a single spot for an extended period of time. The approach that is described in [14] has utilised MoSIFT in order to model the movements of the objects that are being watched in order to facilitate the identification of violent behaviour. Utilising a Bag of Words (BoW) strategy is the foundation of this method. Following that, it does a spatial and temporal analysis on the video in order to discover any anomalies that may be present.

Looking at things from a broader viewpoint, the research that is now being conducted in the field of human action identification and categorization is expanding more traditional designs in order to further improve performance. The human action recognition and categorization domain was the focus of these investigations, which were successfully completed. An strategy is presented by the authors of [15] that may be utilised to make use of pre-trained two-dimensional convolutional neural networks (CNNs) for both the spatial and temporal streams that are implicated in action classification tasks. In order to accomplish this, they make use of a two-dimensional convolutional neural network that has been pre-trained for the temporal stream rather than an optical stream. They produce a stacked grayscale three-channel image, commonly referred to as an SG3I, by selecting three RGB frames from different locations throughout the video. The pre-trained two-dimensional convolutional neural network (CNN) is then subjected to SG3Is in order to improve its accuracy.

The Combined Video-stream Deep Network is a concept that is suggested in. For the purpose of extracting spatio-temporal information from video recordings, this network makes use of a ResNet. They use ResNets models that have previously been trained on the Kinetics dataset in order to extract attributes from the video stream that is being used for the UCF-101 session. After that, they take the optical flow graphs that are included within the UCF-101 dataset and use them as input to an optical stream in order to extract the optical characteristics that are contained within it. In many respects, this approach is comparable to the one that we actually use [16]. After everything is said and done, the properties of both streams will be combined. In this inquiry, the distance transform and entropy features that were recovered from images of human silhouettes taken after the backdrop had been removed from the photographs are the primary focus of attention. These features are the input that deep neural networks utilise to detect human activities since they contain information about the shape as well as the local variation. Human activities are recognised by these neural networks.

The detection of irregularities in video surveillance is the subject of an additional line of research that is connected to this subject. An approach that has been suggested for use in surveillance [17] centres on the utilisation of Unmanned Aerial Vehicles (UAVs), which are also referred to as drones in some circles. To extract features from UAV data, they used a CNN, Histogram of Oriented Gradient (HOG), and HOG3D as part of a three-step process. This allowed them to extract features from the film. The extraction of features was accomplished by the use of this method. With the assistance of a support vector machine that only had one class, the classification procedure was successfully finished. In a recent study, Gu et al. [18] advocated the use of a particular method of deep learning that employs a fully convolutional neural network (FCN) as a regressor for the goal of crowd counting when applied to aerial photos captured by unmanned aerial vehicles (UAVs). They train two FCNs simultaneously by using photographs of the crowd that have been recorded, in addition to the crowd heatmaps that correspond to those photographs. A lightweight crowd detection system was provided in reference number 41 to enable the safe landing of unmanned aerial vehicles (UAVs). This was done in the same vein as the previous example. The restricted processing capabilities of unmanned aerial vehicles (UAVs) were taken into mind by this means. There is a possibility that [19] has a solution that is fundamentally distinct from what has been proposed in the past.

This study makes advantage of the one-shot learning technique for anomaly recognition in order to produce a method for one-shot anomaly detection for surveillance systems. The goal of this work is to construct a method. The technique that they use makes use of a lightweight siamese 3D CNN that is capable of determining the degree to which two anomaly sequences are comparable to one another in a speedy and accurate manner. Through the use of a pre-trained ResNet-50 architecture, the spatio-temporal features are recovered in the work that was referred to before, which is designated as [20]. The obtained attributes are then passed to a multi-layer BDLSTM model in order to classify irregularities in surveillance videos. This process is repeated until the desired results are achieved. Bidirectional long short-term memory is what the acronym BDLSTM stands for. A strategy was proposed by Sha et al. [21] for identifying five unique behaviours in a specific business setting. Out of these five behaviours, two were considered to be deviant. Through the utilisation of DenseNet in a two-stream architecture, they were able to extract spatial and temporal features. Additionally, they utilised a particular way in order to alleviate the problems that were brought about by inaccurate data. For both the training and testing of the model, a dataset that was produced by the person who first developed the model was utilised [22].

Through the utilisation of the UCF-Crime dataset, an ImageNet pretrained VGG16 architecture was utilised, and InceptionV3 was utilised to fine-tune the model under consideration. When applied to the particular circumstances of the present inquiry, a number of approaches are centred on the identification of certain patterns of conduct that are not typical [23]. Afiq et al. [24] have suggested a method for the identification of falls that makes use of convolutional neural networks (CNN) to train the detection of falls via the use of optical flow images. When it comes to the training procedure for the VGG-16 network, ImageNet and optical stacks from UCF-101 are utilised. In the end, they make use of transfer learning and fine-tune the network by making use of the data from the phase before these instructions. The technology that we use is pretty comparable to this one; however, the 3D architecture that we employ enables us to handle video in a far more effective manner. In contrast to the findings that were shown by Ben et al. [25], feature extraction from colour photographs may be accomplished with the utilisation of PCAnet, and then an SVM can be utilised to recognise instances of falling.

A modified version of the VGG-16 architecture was utilised by Almazroey et al. [26]. Within the scope of the research, an early application of a vision component seeks to extract frames of moving persons from motion pictures. Following that, in, a combination of histograms, local binary patterns, and characteristics recovered by Caffe are utilised in order to identify a silhouette. In the end, two support vector machine classifiers are employed in order to ascertain whether or not there have been any falls. Calderara et al. [27] conducted research that was somewhat comparable to this one, and they constructed a fall system that trains a CNN based on geometric properties. Following that, they begin the process of extracting motion characteristics by segmenting the head and the torso using the usual elliptical fitting technique. This is done in order to get the motion characteristics.

After that, the authors make use of a shallow CNN structure in order to learn these properties and achieve an exceptional level of accuracy on their own dataset that they have manually gathered. Instead of relying on features that are produced through the use of features, the current work depends on the features that are learned by CNN [28]. This is in contrast to the previous techniques that have been discussed. The researchers L. Lazaridis et al. [29] used a CNN architecture that consisted of four streams in order to recognise falls by making use of multimodal data that was obtained from RGB-D cameras. In order to deal especially with information related to static appearance, shape changes, and motion, they utilised a combination of RGB and depth photographs in addition to three other modalities outside the RGB and depth pictures.

III. PROPESED METHODLOGY

The technique that has been developed makes use of YOLO and CONV2d in order to recognise unnatural patterns and behaviours that are associated with humans. In contrast to YOLO, which is in charge of recognising individuals and bodily parts, CONV2d is in charge of making observations on behaviour and spotting anomalies. CNN models are able to extract features and patterns from images without the requirement for human feature engineering at any point in the creation process. As a consequence of this, these models are more resistant to variations in the quality of the image obtained and the lighting circumstances. LSTM models are effective for evaluating human behaviour over the course of time because they are able to handle sequences of input data and capture temporal relationships. This makes them an excellent choice for this purpose.

The working of proposed system is as follows:

- Collecting the Dataset
- Performing the Pre-processing operation
- Design the Detection Model

- Train the Abnormal Detection model
- Predicting the Anomaly Behaviour Detection

1. Collecting the Dataset

`One of the main and first steps in the detecting procedure is gathering the dataset. Dataset including pictures of typical human actions. So that variances in usual conduct may be captured, this dataset should encompass a wide variety of activities and contexts.

2. Collecting the Dataset

One of the main and first steps in the detecting procedure is gathering the dataset. Dataset including pictures of typical human actions. So that variances in usual conduct may be captured, this dataset should encompass a wide variety of activities and contexts.

3. Perform Pre-Processing

Data pre-processing is the subsequent operational step. It prepares the dataset for training by processing it if required. This stage involves resizing the photos to a uniform resolution, normalising the pixel values to a consistent scale, and maybe supplementing the data to increase its diversity and generalizability.

4. Design the Detection Model

Build an architecture for a Conv2D network that can analyse human activity. To achieve this, it is common practise to stack several convolutional layers and then use pooling layers to reduce the number of spatial dimensions. It is possible to construct layers for categorization purposes that are fully linked.

Training

You Only Look Once version 3, or YOLOv3, is an effective method for finding instances of human anomaly in visual media. A collection of media showcasing both typical and atypical human behaviours is amassed. Using bounding boxes, annotations may be applied to represent people or places that are aberrant. The annotated dataset is used to train YOLOv3. When training a model, its parameters are fine-tuned to reduce the likelihood of incorrect predictions for class probabilities and anticipated bounding boxes. In order to identify suspicious human actions, fresh pictures or video frames are fed into the trained YOLOv3 model.

Bounding boxes and class probabilities are appended to detected items. A confidence threshold can be used to reduce false positives by excluding predictions with low confidence ratings. We examine the model's output, which includes bounding box data and class probabilities. The detection of anomalous human behaviour is evaluated using preset algorithms or criteria. To remove bounding box predictions that are too similar or overlap too much, non-maximum suppression is useful.

5. Anomaly Detection

Use the trained Conv2D network on pictures or movies that it has never seen before to find things that aren't right. You should put the test data into the network and then watch what the output reports say. As an oddity, something out of the ordinary is happening when the noticed behaviour is very different from the established normal patterns.

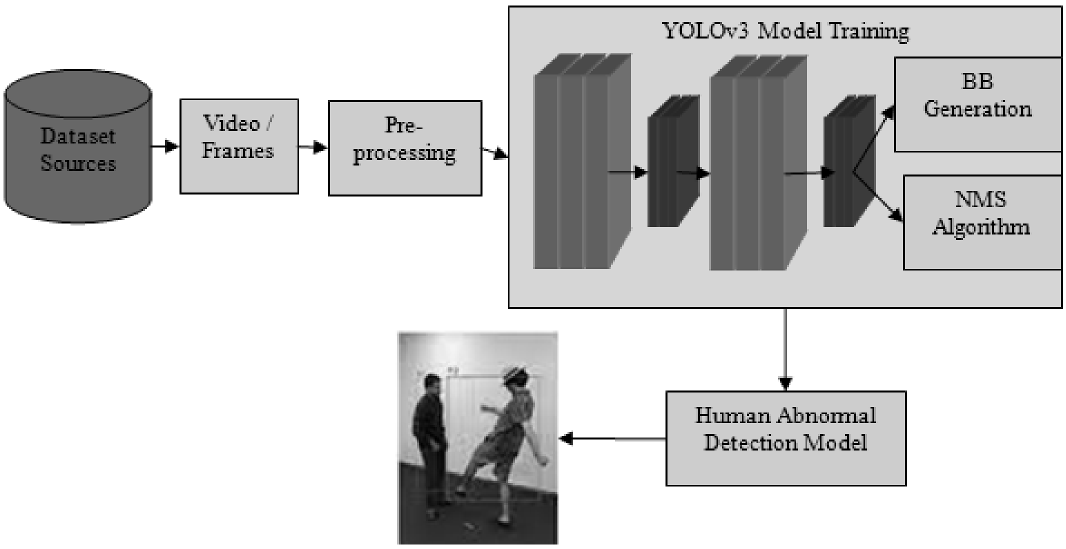

Figure 2: Human Abnormal Behaviours Detection Systems

IV. EXPERIMENTAL DISCUSSION AND ANALYSIS

This section will examine the outcomes of the study, as well as provide a description of the dataset that is available to the public and the information on the performance metrics for the human abnormal detection model.

1. Performance Metrics

The IoU, Accuracy, Precision, and Recall metrics are utilised in the evaluation of YOLOv3. It is the projected bounding box and ground truth overlap rate that the model uses to calculate the IoU. According to Equation (1), the bounding box is considered correct when the IoU value is greater than the threshold. The objective is to determine the extent of the overlap between the predicted bounding boxes and the ground truth. The outcome of the test is considered to be a true positive (TP) when the IoU is greater than 50%, whereas a false positive is when the IoU is less than 50%. The model predicts that there is no aircraft in the image, but it actually identifies one. This is what is known as a false negative (FN). After taking these metrics, accuracy and recall are achieved.

$$a = \frac{B_p \cap B_{gt}}{B_p \cup B_{gt}} \geq a0 \tag{1}$$

Object detection precision may be evaluated and quantified with the help of IoU, which is a universally used evaluation metric. At least fifty percent of the ground truth and the projected box cover the same region, as indicated by the fact that the value of a0 has been assigned a value of 0.5. Any time the IoU is more than the criterion of fifty percent, the test case is regarded to be an abnormal.

2. Dataset Descriptions

The experiment makes use of a data set that was individually produced in addition to the data set that is accessible to the general public from the University of Minnesota. In order to demonstrate not only the superiority of the suggested detection model but also its practicability, this is done. Anonymity detection study is carried out with the help of the UMN data set, which is a public data collection. An sudden movement, the arrival and disappearance of crowds, and aggregations are the primary components of the UMN data set. In addition to the fact that the data set has a total of 320 × 240 pixels across each frame, the image quality is not very high. The collection contains a total of 7739 photos. It is possible to divide scenery into three separate types.

The first type of situation is characterized by the presence of two distinct and complete sets of anomalies. In its totality, the second group has six individual data items that are completely abnormal. The count of the third group includes three complete sets of data items that are considered to be aberrant. The test set is made up of the remaining frames that were taken from the original video clips, while the training set is made up of the first 180 frames that were obtained from nine different video clips that were used in three different hypothetical situations.

Figure 3: Sample Dataset Descriptions of Human Abnormal Detection

3. Results and Discussions

Two different datasets were used during the training process of the YOLOv3 human abnormal detection model. Each dataset consists of a pre-processed human abnormal activity of the same dimensions. The YOLOv3 model that was suggested was trained using the aeroplane images from the UMN data set as well as the personally constructed dataset. Each and every experiment was trained and carried out on a personal computer that had Intel® i7 processors operating at speeds of up to 3.40 GHz and 16 GB of random access memory (RAM). As a result of the experiment, we noticed that the suggested YOLOv3 model was able to greatly raise the accuracy to 92.2% while simultaneously reducing the amount of processing time required. Table 2 presents the findings of the investigation.

Table 1: Performance Analysis of YOLOv3 Human Abnormal Detection

Model	Dataset	Accuracy	Precision	Recall
YOLOv3	UMN Dataset	93.2	93.1	91.8
	Personally Constructed Dataset	89.8	89.2	88.5

The Figure 4 and 5 are represents the GUI of Human Abnormal Activity Detection model using the GUI representation. Finally Figure 5 predicted and detection as Human Fighting Activities.

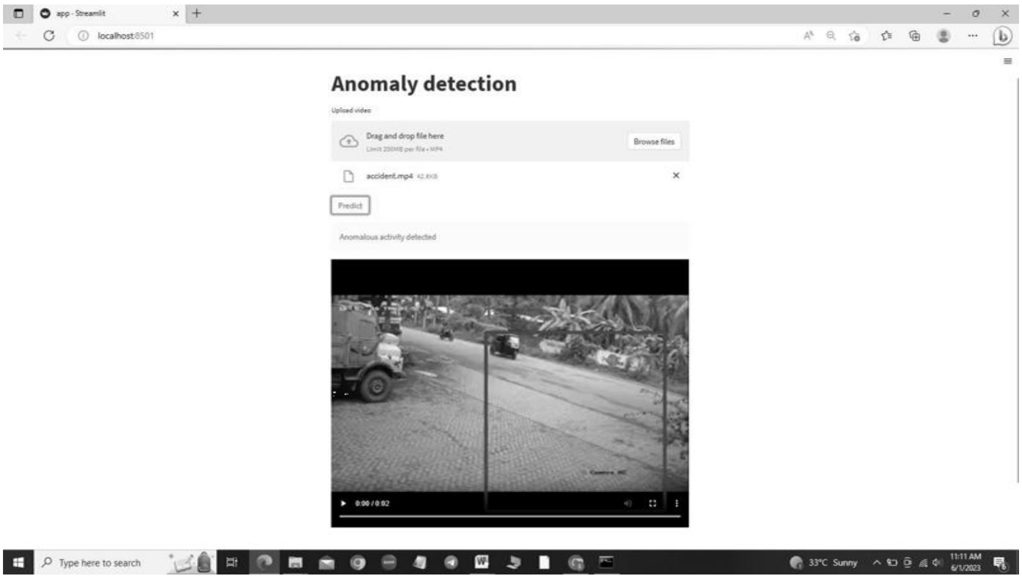

Figure 4: GUI of Human Abnormal Detection

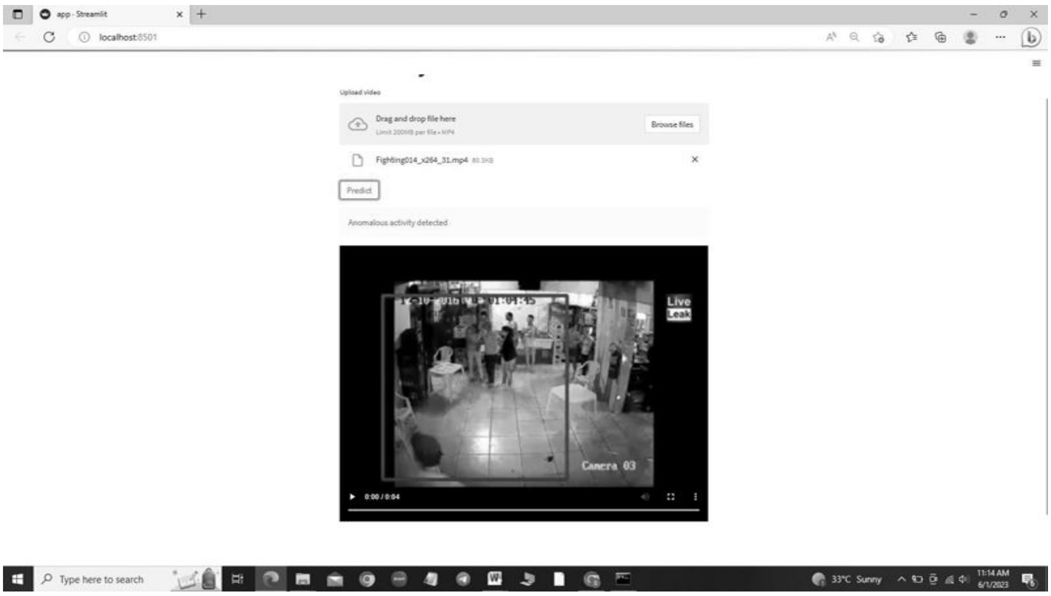

Figure 5: GUI of Human Abnormal Detection as Prediction

YOLOv3 is evaluated using IoU, Accuracy, Precision, and Recall. The IoU is the model's predicted bounding box

V. CONCLUSIONS

YOLO (You Only Look Once) and CNNs like Conv2DNet may be used to analyse human behaviour and discover abnormalities. YOLO's real-time object identification and Conv2DNet's spatial feature recognition are ideal for video human behaviour detection and analysis. YOLO can recognise and track people in real time by training on a dataset of different human behaviour. This lets the machine detect persons and their behaviours in video.

Conv2DNet, a CNN architecture, captures spatial data and patterns well. It can learn regular human behaviours including walking, sitting, and running. By merging YOLO's object detection with Conv2DNet's behaviour identification, the system can identify anomalous behaviour. Finding abnormal behaviour like fighting, loitering, or strange movement patterns might prompt additional research. The YOLO-Conv2DNet system may be used for monitoring in public places, airports, and vital infrastructure. It improves security and human monitoring. The device helps operators spot dangers and abnormalities, enhancing situational awareness.

However, high-quality and varied training data for YOLO and Conv2DNet is necessary. Accurate and trustworthy findings require enough training data, even anomalous behaviour. To adapt to changing behaviours and reduce false positives and negatives, the system must be monitored, evaluated, and fine-tuned. In conclusion, YOLO and Conv2DNet combination for human behaviour analysis and anomalous identification appears promising. This combination of real-time object detection and spatial feature identification improves surveillance security and situational awareness.

REFERENCES

[1] A. Bux, P. Angelov, and Z. Habib, “Vision based human activity recognition: a review,” Multimedia Tools and Applications, vol. 79, 2017.
[2] Lei Cai, Peien Luo and Guangfu Zhou, "Multistage Analysis of Abnormal Human Behavior in Complex Scenes", *Journal of Sensors*, pp. 1-10, 2019.
[3] M. S. K. Mishra and K. S. Bhagat, “A survey on human motion detection and surveillance,” International Journal of Advanced Research in Electronics and Communication Engineering, vol. 10, 2015.
[4] M. Valera and S. A. Velastin, “Intelligent distributed surveillance systems: a review,” Vision, Image and Signal Processing, vol. 152, no. 2, pp. 192–204, 2005.
[5] Popoola Oluwatoyin, “Wang kejun. Video-based abnormal human behavior recognition review,” IEEE Transactions on Systems Man & Cybernetics Part C.vol. 42, no. 6, pp. 865–878, 2012.
[6] Ioana Alexandra Bozdog, Todea Daniel-Nicusor, Marcel Antal; Claudia Antal, Tudor Cioara, Ionut Anghel, Ioan Salomie, “Human Behavior and Anomaly Detection using Machine Learning and Wearable Sensors”, IEEE 17th International Conference on Intelligent Computer Communication and Processing (ICCP), 2021.
[7] Huifang Qian, Xuan Zhou, Mengmeng Zheng, “Detection and Recognition of Abnormal Behavior based on Multi-level Residual Network”, IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2020.
[8] Weihu Zhang, Chang Liu, “Research on Human Abnormal Behavior Detection Based on Deep Learning”, International Conference on Virtual Reality and Intelligent Systems (ICVRIS), 2021.

[9] Rongyong Zhao, Yan Wang, Ping Jia, Cuiling Li, Yunlong Ma, Zhishu Zhang, "Abnormal Human Behavior Recognition Based on Image Processing Technology", IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2021.
[10] Thomas Gatt, Dylan Seychell, Alexiei Dingli, "Detecting human abnormal behaviour through a video generated model", 11th International Symposium on Image and Signal Processing and Analysis (ISPA), 2019.
[11] Shih-Chung Hsu, Cheng-Hung Chuang, Chung-Lin Huang, Por-Ren Teng, Miao-Jian Lin, "A video-based abnormal human behavior detection for psychiatric patient monitoring", International Workshop on Advanced Image Technology (IWAIT), 2018.
[12] Yuan Cao, Hao Xu, Qiang Yang, "Computer-vision-based abnormal human behavior detection and analysis in electric power plant", 33rd Chinese Control and Decision Conference (CCDC), 2021.
[13] Yixue Hao, Zaiyang Tang, Bander Alzahrani, Reem Alotaibi, Reem Alharthi, Miaomiao Zhao, Arif Mahmood, "An End-to-End Human Abnormal Behavior Recognition Framework for Crowds With Mentally Disordered Individuals", IEEE Journal of Biomedical and Health Informatics (Volume: 26, Issue: 8, August 2022), 2021.
[14] Yimin DOU, Cai Fudong, Jinping LI, Cheng Wei, "Abnormal Behavior Detection Based on Optical Flow Trajectoryof Human Joint Points", Chinese Control And Decision Conference.
[15] Federico Angelini, Jiawei Yan, Syed Mohsen Naqvi, "Privacy-preserving Online Human Behaviour Anomaly Detection Based on Body Movements and Objects Positions", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019.
[16] S. Majhi, R. Dash and P.K. Sa, "Two-Stream CNN Architecture for Anomalous Event Detection in Real World Scenarios" in Computer Vision and Image Processing. CVIP 2019. Communications in Computer and Information Science, Singapore: Springer, vol. 1148, 2020.
[17] Lee, S. C., & Nevatia, R, Hierarchical abnormal event detection by real time and semi-real time multi-tasking video surveillance system. Machine Vision and Applications, 25(1), 133-143, 2014.
[18] Gu, X., Cui, J., & Zhu, Q, Abnormal crowd behaviour detection by using the particle entropy. Optik-International Journal for Light and Electron Optics, 125(14), 3428-3433, 2014.
[19] Zhu, X., Liu, J., Wang, J., Li, C., Lu, H. Sparse Representation for Robust Abnormality Detection in Crowded Scenes. Pattern Recognition, 2014, 47(5), 1791-1799. https://doi.org/10.1016/j.patcog.2013.11.018
[20] Wang, T., Qiao, M., Zhu, A., Shan, G., Snoussi, H. Abnormalevent Detection via the Analysis of MultiFrame Optical Flow Information. Frontiers of Computer Science in China, 2020, 14(2), 304-313. https:// doi.org/10.1007/s11704-018-7407-3
[21] Liu, W., Luo, W., Lian, D., Gao, S. Future Frame Prediction for Anomaly Detection - A New Baseline. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018. https://doi.org/10.1109/ CVPR.2018.00684
[22] Hu, Z., Zhang, L., Li, S., Sun, D. Parallel Spatial-Temporal Convolutional Neural Networks for Anomaly Detection and Location in Crowded Scenes. Journal of Visual Communication and Image Representation, 2020, 67, 765-771.
[23] Adam, A., Rivlin, E., Shimshoni, I., Reinitz, D. Robust Real-Time Unusual Event Detection Using Multiple Fixed-Location Monitors. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(3), 555-560.
[24] Afiq, A. A., Zakariya, M. A., Saad, M. N. M., Nurfarzana, A. A., Khir, M. H. M., Fadzil, A. F., Jale, A., Gunawan, W., Izuddin, Z. A. A., Faizari, M. A Review on Classifying Abnormal Behavior in Crowd Scene. Journal of Visual Communication and Image Representation, 2019, 58, 285-303.
[25] Ben,M.,A., Zagrouba, E. Abnormal Behavior Recognition for Intelligent Video Surveillance Systems: A Review. Expert Systems with Applications, 2018, 91(1), 480-491.
[26] Almazroey, A. A., Jarraya, S. K. Abnormal Events and Behavior Detection in Crowd Scenes Based on Deep Learning and Neighborhood Component Analysis Feature Selection. In: Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020), 2020, 258-267. https://doi. org/10.1007/978-3-030-44289-7_25
[27] Calderara, S., Heinemann, U., Prati, A., Cucchiara, R., Tishby, N. Detecting Anomalies in People's Trajectories Using Spectral Graph Analysis. Computer Vision and Image Understanding, 2011, 115(8), 1099-1111.
[28] Elbayoudi, A., Lotfi, A., Langensiepen, C. The Human Behavior Indicator: A Measure of Behavioural Evolution. Expert Systems WITH Applications, 2019, 118, 493-505. https://doi.org/10.1016/j.eswa.2018.10.022

[29] L. Lazaridis, A. Dimou and P. Daras, "Abnormal Behaviour Detection in Crowded Scenes Using Density Heat", 2018 26th European Signal Processing Conference (EUSIPCO) 2060-2064.aps and Optical Flow. 2018 26th European Signal Processing Conference (EUSIPCO), pp. 2060-2064, 2018.