Utilitas Mathematica

ISSN 0315-3681 Volume 122, 2025

Abnormal Human Activity Detection Using Yolov5

Kasireddy Shyam Kumar
M.Tech, ANITS
Visakhapatnam
kasireddyshyamkumar.23mtech.csm@anits.edu.in

Ch Sravanthi Sowdanya
Assistant Professor
CSE(AI&ML),ANITS
Visakhapatnam
chsravanthi.csm@anits.edu.in

Abstract: effective surveillance and anomaly detection systems are crucial in the digital age. For detecting suspicious actions in photos and videos, this book offers a step-by-step method to implementing cutting-edge object recognition algorithms—including YOLOv5, YOLOv6, YOLOv7, and YOLOv8 Cutting-edge its outstanding performance, YOLOv5, renowned for its precision and quickness, is the main emphasis. Ensuring a seamless process, the guide starts by describing the required libraries and dependencies for the project. It then addresses dataset imports, stressing the YOLO format for annotations, and dataset preparation for efficient model training. Sample photos with bounding boxes are visualised and class distribution examined using exploratory data analysis (EDA). The book also covers evaluating image size and resolution compatibility with YOLOv5 criteria. The main part emphasises using the YOLOv5 algorithm, including model loading, training, and performance evaluation measures. The trained model's practical usesstate, such real-time or batch inference for security and surveillance objectives, are also investigated. Although YOLOv5 is praised for its great accuracy, the book advises experimenting with later versions—including YOLOv6, YOLOv7, and YOLOv8—for better performance.

"Index Terms: *surveillance, anomaly detection, YOLOv5, object detection, real-time inference, model training".*

1. INTRODUCTION

Ensuring safety and security in both physical and virtual areas [1], [6], [25] demands efficient surveillance and anomaly detection systems in the fast changing digital environment of today. Sophisticated technologies are crucial in offering solutions as the necessity to handle these issues becoming ever more pressing [2], [4], [5]. This all-inclusive manual presents a step-by-step method for applying modern object identification algorithms—including YOLOv5, YOLOv6, YOLOv7, and YOLOv8—within the framework of spotting suspicious behaviour in photos and videos [23]. Renowned for its remarkable precision and speed, YOLOv5 has gained great popularity and praise, making it perfect for a wide range of uses, especially in security and surveillance [24], [27]. From establishing the needed environment to training tailored models able to detect and flag suspicious activities [3], [19], this guide intends to equip users with the knowledge and tools required to maximise the potential of YOLOv5 and its successors.

beginning with the advent of basic components, inclusive of essential libraries and dependencies, the ebook offers insights on information manipulation gear and YOLO-specific libraries, therefore setting the level for later moves [12], [15]. It then explores dataset imports, stressing the YOLO layout for annotations, and offers recommendation on correctly getting ready and curating datasets [22], [23].

Later components take a look at "exploratory statistics analysis (EDA)" techniques particular to YOLO algorithms, model utility, education, assessment, and real deployment situations [14], [18], [28]. Facilitating the development and deployment of strong object detection structures in state-of-the-art complex digital surroundings, this manual is a useful device for making use of YOLOv5 and its successors in security, surveillance, and related use instances [1], [25], [26].

2. LITERATURE REVIEW

Becoming one of the essentials of daily living, "automated teller machines (ATM)" are commonly utilised day conduct banking activities. Round the clock, ATMs enable money transfer from one account day another, deposit, and withdrawal. Though, illegal acts like money snatching and assault on consumersday are compromising this convenience more and more, hence compromising the safety of bank day-to-day. This study presents a video-basedday system that quickly detects unusual activity taking place at ATM sites and triggers an alarm during any unfortunate event. The suggested method extracts pertinent characteristics from video using "motion hisdayeveryday image (MHI)" and Hu moments. Feature dimensionality reduction has been done using principle component analysis; classification has been done using support vecdayr machine. Varying the window size of MHI has allowed for analysis on several video sequences. With an average "accuracy of 95.73%, the suggested system" can identify normal and deviant actions such as money theft, harm day the consumer by virtue of conflict, or attack on the client.

Surveillance and security of bank-"automated Teller Machines (ATMs)", public offices have made real-time detection of human activities quite crucial given the rise in criminal activity. Monocular CCTV cameras recording just RGB video are currently used day monitor such limited settings. Apart from RGB data, the RGB+D sensor offers depth information of the scene. We provide a supervised deep learningday framework built on multi-stream CNNs and RGB+D sensor day solve the issue of online detection of anomalous behaviours in bank ATMs. Motion templates are generated from RGB and depth video segments from the online video stream of RGB+D data and then trained using CNNs day identify a suspicious event in continuous activity. Furthermore, we also provided a new RGB+D dataset in this article since no dataset for examining human actions at ATMs was accessible. The suggested deep learningday-basedday system detects suspicious events with "a precision of 0.932 and an accuracy of 94.2% using" qualitative and quantitative statistics evaluation criteria. Results of thorough statistical research reveal that the suggested system can identify the suspicious event in a real-time online way prior to the completion of the aberrant action.

Four in industrial environments, sports, and healthcare, human activity recognition has several uses. In the latter, it canday track industrial personnel and assess whether the necessary tasks are properly carried out. This research uses DL day methods day identify 10 different packing operations carried out by

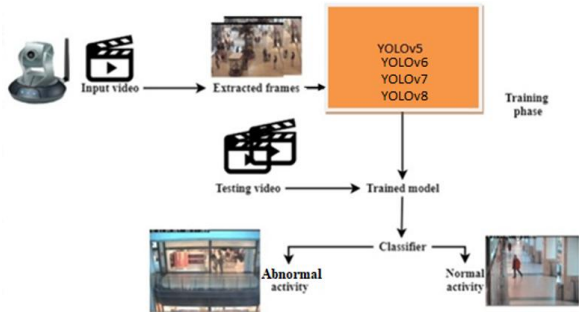
sixteen Openpack dataset members. Our suggested architecture processes spatial and temporal data using Convolutional Neural Networks and long short-term memory networks by combining input from several sensors. We also add Transformers day our network, which "raises the F1-score performance day 98.21".

5 This work presents one among the most important uses of human suspicious activity recognition known as anomaly detection. Every culture nowadays is primarily concerned with ensuring safety for a person. The main source of this worry is the never-growing actions generating hazards, beginning from intentional aggression to an accident-related injuries. Installing a conventional “closed circuit television (CCTV)" is not enough; it calls for someone to always be vigilant and watch the cameras, which is rather ineffective. This call for the need to create a security system which is completely automated system that identifies abnormal actions in real time and provides immediate assistance to the victims. Therefore, we suggested a system that uses machine learning tools to analyse and identify suspicious human behaviour from real-time CCTV data and produces the alarm should the abnormal activity occur. The technique produces better outcomes in experiments conducted on the dataset comprising both normal and anomalous activity.

Six. Identifying different unusual human behaviours from film is quite difficult. The lack of datasets with several anomalous human activities also has a major impact on this issue. Though only a small number include non-standard human behaviour like theft, harassment, etc., the accessible datasets include many human behaviours. Datasets like KTH concentrate on unusual activities such abrupt behavioural shifts as well as on different alterations in interpersonal relationships. Categories in the UCF-crime dataset include fighting, abuse, explosives, robbery, etc. This data collection is somewhat time intensive, though. The events in the films take place in a few seconds. The general outcomes of the neural networks employed to pick out the event will be influenced by this. This paper presents a dataset addressing unusual behaviours with categories like Begging, Drunkenness, fight, Harassment, Hijack, Knife hazard, normal videos, pollution, property damage, robbery, and Terrorism. The Conv "LSTM (convolutional long short-term memory)" neural network we built is trained and tested using the produced dataset. On the other hand, we also evaluate the produced dataset using various architectures. Our “3D Resnet50, 3D Resnet101, and 3D Resnet152 ConvLSTM designs run”. Using the built dataset and the architecture we established, "we achieved a classification accuracy of 96.19% and a precision of 96.50%".

3. MATERIALS AND METHODS

For efficient monitoring and anomaly identification in photos and videos [19], [27], the suggested system uses the YOLOv5 algorithm, renowned for its great "accuracy [23], [24], together with YOLOv6, YOLOv7, and YOLOv8". It includes establishing the required environment, importing YOLO format datasets [22], doing exploratory data analysis [12], and training the model [14], [18]. Using modern technology [1], [2], [5], this method guarantees exact identification of suspicious behaviour. Though more recent versions exist, YOLOv5 distinguishes itself with remarkable accuracy [23], [24]. Spanning security, surveillance, and other fields, the system's uses [6], [25] provide a complete solution for building and walking a strong object detection system with unsurpassed accuracy and speed [3], [19].



"Fig.1 Proposed Architecture"

The shown architecture is a YOLO-based anomaly detection tool for surveillance [23], [24]. Input video footage is first recorded and processed to obtain individual frames, which are then employed in the training phase using many "YOLO models (YOLOv5, YOLOv6, YOLOv7, YOLOv8) [19], [27]". Annotated and used to train object detection systems to identify and distinguish between routine and suspicious activity, these frames [1], [2], [5]. Testing films are fed into the system once the model is trained; the learnt YOLO model methods the frames and sends them to a classifier [6], [25]. Allowing real-time or batch surveillance analysis for efficient security monitoring, the classifier examines the outputs and classifies the identified actions as either abnormal or normal [3], [8], [26].

a) Dataset Collection:

Comprising video footage from public or simulated settings, the dataset for this surveillance and anomaly detection system extracts and annotates individual frames using the YOLO format [19], [22]. Every frame is marked with class identities and bounding boxes denoting either normal or aberrant behaviour [2], [5], [6]. These annotations comprise class IDs and normalised coordinates, hence allowing exact object identification and categorisation [23], [24]. Used throughout the training and testing stages, the dataset is prepared to guarantee compatibility with YOLO models [1], [3], [25]. In real-time monitoring situations, this "structured approach lets YOLOv5, YOLOv6, YOLOv7, and YOLOv8" models learn activity patterns and properly separate suspicious behaviour from ordinary activities [19], [27].

b) Pre-Processing:

Data Exploration: This module loads the dataset into the system for first analysis [19], [22]. It enables users to examine the structure, format, and content of the data. visual tools including sample image previews with bounding boxes and class distribution charts are employed at this phase [12], [14]. The key goal is to understand the features of the dataset before moving on to preprocessing or training [23].

Processing: by means of necessary transformations [14], [15], the processing module gets the dataset ready for model training. It consists of image scaling, pixel value normalising, and annotation file conversion to YOLO-compatible format [22], [23]. This guarantees consistency and compliance with the architecture of the YOLO model. During training, high-quality preprocessing helps to increase detection accuracy and model stability [24].

Splitting Data into Train & Test: usually with an 80-20 or 70-30 ratio, this module splits the dataset into training and testing sets [3], [19]. The aim is to train the model on one subset and validate it on another to mimic real-world performance. To prevent biassed outcomes, class balance is kept throughout the split [1]. Evaluating the generalisation potential of the detection model [25] depends on this separation.

Model Generation: object detection "models like YOLOv5, YOLOv6, YOLOv7, and YOLOv8 are constructed" and trained in this module [23], [24]. Here, key training parameters are set: learning rate, batch size, and number of epochs. From annotated input frames, the models learn to precisely identify aberrant activity [2], [5]. The trained models are stored for further use in prediction or deployment [19] upon finishing.

c) Algorithms:

YOLOv5: This project mostly uses YOLOv5, the main object detection tool, because of its speed-accuracy balance [23]. Built with PyTorch, it is perfect for surveillance uses since it allows quick training and real-time inference [24]. With several configurations (small to extra-large) to fit diverse hardware and performance requirements, YOLOv5 is quick, lightweight, and simple to customise. Working on video frames, it finds and categorises activities, thereby fast and precisely spotting aberrant behaviour [25].

YOLOv6: developed with better performance in mind, YOLOv6 is an industrial-grade object detector [23]. It adds architectural improvements and training optimisations over YOLOv5, including enhanced backbone networks and anchor-free detection heads. This work tests YOLOv6 to see whether its improvements result in better detection of intricate or subtle anomalies in surveillance data. It is particularly useful for managing heavily packed or high-resolution scenes [27].

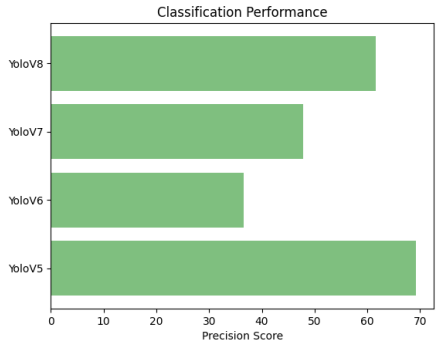
YOLOv7: A significant development in the YOLO familystate, YOLOv7 offers real-time object identification with tate-of-the-art accuracy [23]. It includes dynamic model scaling, re-parameterized convolution, and E-ELAN among other architectural breakthroughs. YOLOv7 is applied in this work to evaluate changes under difficult circumstances such congested settings or overlapping tasks. Its great accuracy and durability make it a viable substitute for YOLOv5 [28].

YOLOv8: The most recent edition in the YOLO series, YOLOv8, sports a revamped architecture supporting segmentation and object detection [23]. It provides quicker inference rates, improved feature extraction, and dynamic input shapes. This work makes use of YOLOv8 to investigate its improved capacity to identify and differentiate anomalous behaviour with more precision. It is a modern tool for testing next-generation surveillance performance [27].

4. EXPERIMENTAL RESULTS

Precision: Precision assesses the proportion of accurately categorised cases or samples among those designated as positives. The formula to compute the precision is therefore:

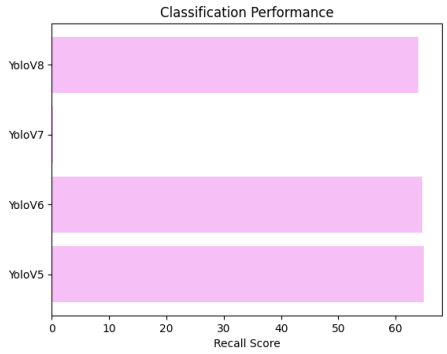
$$Precision = \frac{True\ Positive}{True\ Positive\ + False\ Positive} (1)$$



"Fig.2 Precision graph"

Recall: recall is a ML metric that gauges a model's capacity to find all pertinent examples of a given class. The ratio of accurately predicted positive observations to the total actual positives gives insights on a model's completeness in capturing examples of a particular class.

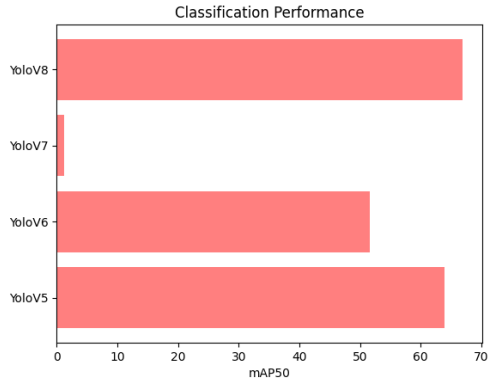
$$Recall = \frac{TP}{TP + FN} (2)$$



"Fig.3 Recall graph"

mAP: "mean average Precision (MAP)" is a ranking quality tool. It takes into account the quantity of appropriate hints as well as their list placement. MAP at k is the arithmetical average of the "average Precision (AP)" at k over all users or queries.

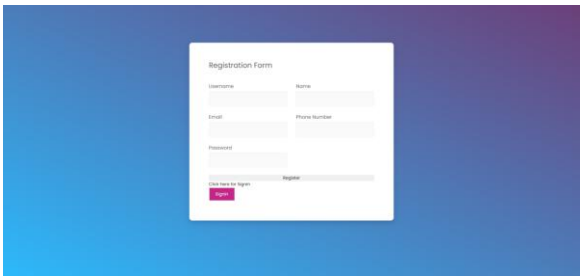
$$mAP = \frac{1}{n} \sum_{k=1}^{k=n} AP_k \quad (3)$$



"Fig.4 mAP50 graph"

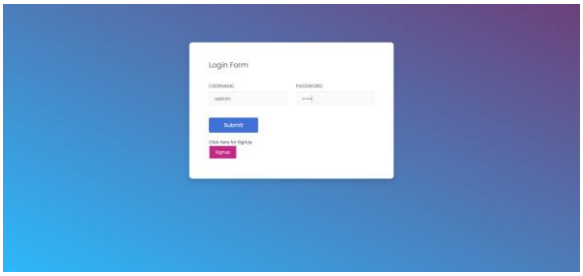
"Fig.5 Welcome page"

Above screen displays dashboard with project title



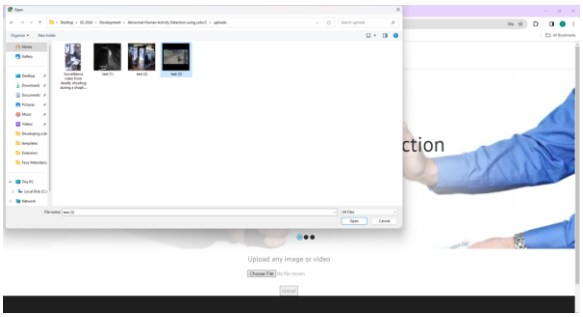
"Fig.6 User signup"

The following screen is a user registration page where users can sign up using their name, address, phone number, email address, etc.



"Fig.7 User Signin"

User login page is the above screen; user might log in here with his username and password.



"Fig.8 User input"

Using above interface, user may submit their testing input

"Fig. 9 Prediction"

Using above interface, user receives prediction result of the loaded input

5. CONCLUSION

Ultimately, the suggested surveillance and anomaly detection system's strong and flexible solution for handling security concerns in photos and videos is provided by the combination of YOLOv5, YOLOv6, YOLOv7, and YOLOv8 algorithms. The system guarantees consistent identification of suspicious actions with accuracy and speed by using the outstanding accuracy of YOLOv5 together with the developments included in YOLOv6, YOLOv7, and YOLOv8. While the latest versions' adaptability lets users select the most appropriate algorithm for their particular needs, YOLOv5's great accuracy provides a basis.

Moreover, the system's efficacy in real-time object detection guarantees quick reaction to possible security risks as YOLOv5's speed and the improvements made in later versions enable it. The system provides a complete solution for building and implementing strong object detection systems with programs ranging from security to surveillance to many other sectors. essentially, the combination of YOLOv5, YOLOv6, YOLOv7, and YOLOv8 algorithms enables users to apply cutting-edge technology to improve safety and security in the fast changing digital environment, hence supporting a more secure and safe surroundings.

REFERENCES

[1] Vikas Tripathi; Hindawi Publishing Corporation, \"Robust Abnormal Event Recognition via Motion and Shape,\" Journal of Electrical and Computer Engineering, pp. 1-11, 2015.

[2] Pushpajit A. Khaire and Praveen Kumar, \"RGB+D and deep learning based real time detection of suspicious,\" Springer; Journal of Real-Time Image Processing, pp. 1-13, 2021.

[3] P. A. Khaire, \"RGB+D and deep learning based real time detection of suspicious,\" Journal of Real-Time Image Processing, pp. 1-13, 21.

[4] C. Shiranthika, \"Human Activity Recognition Using CNN & LSTM,\" IEEE, 2021.

[5] T. S. Bora, \"HUMAN SUSPICIOUS ACTIVITY DETECTION SYSTEM USING CNN MODEL FOR VIDEO SURVEILLANCE,\" IJARIIE, 2021.

[6] R. Vrskova, \"A New Approach for Abnormal Human Activities Recognition,\" Sensor, 2022.

[7] S. Sabbu, \"LSTM-Based Neural Network to Recognize Human Activities,\" Hindawi, pp. 1-8, 2022.

[8] Rajeshwari S, Vismitha G, Sumalatha G and Safura Aliya, "Unusual Event Detection for Enhancing ATM Security," International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, pp. 1-6, 2021.

[9] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," SIGKDD Explor. Newsl., vol. 12, no. 2, pp. 74–82, Mar. 2011, doi: 10.1145/1964897.1964918.

[10] A. Murad and J.-Y. Pyun, "Deep Recurrent Neural Networks for Human Activity Recognition," Sensors, vol. 17, no. 11, p. 2556, Nov. 2017, doi: 10.3390/s17112556

[11] P. Kuppusamy and C. Harika, "Human Action Recognition using CNN and LSTM-RNN with Attention Model" International Journal od Innovative Technology and Exploring Engineering(IJITEE), vol.8,Issue 8, pp.1639-1643, 201

[12] https://www.analyticsvidhya.com/blog/2022/03/basics-of-cnn-in-deep-learning

[13] Y. Chen, K. Zhong, J. Zhang, Q. Sun, and X. Zhao, "LSTM Networks for Mobile Human Activity Recognition," presented at the 2016 International Conference on Artificial Intelligence: Technologies and Applications, Bangkok, Thailand, 2016, doi: 10.2991/icaita- 16.2016.13

[14] https://ieeexplore.ieee.org/document/904397

[15] https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939

[16] C. Jobanputra, J. Bavishi, and N. Doshi, "Human Activity Recognition: A Survey," Procedia Computer Science, vol. 155, pp. 698–703, 2019, doi: 10.1016/j.procs.2019.08.100

[17] https://deepai.org/publication/evaluating-two-stream-cnn-for-video-classificatio

[18] https://www.codeproject.com/Articles/1366433/Using-Modified-Inception-V3-CNN-for-Video-Processin

[19] https://www.kaggle.com/datasets/mehantkammakomati/atm-anomaly-video-dataset-atma

[20] A. Murad and J.-Y. Pyun, "Deep Recurrent Neural Networks for Human Activity Recognition," Sensors, vol. 17, no. 11, p. 2556, Nov. 2017, doi: 10.3390/s17112556

[21] T. Zebin, M. Sperrin, N. Peek, and A. J. Casson, "Human activity recognition from inertial sensor time-series using batch normalized deep LSTM recurrent networks," in 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, Jul. 2018, pp. 1–4, doi: 10.1109/EMBC.2018.8513115.

[22] https://github.com/pjreddie/darknet/blob/master/data/coco.names

[23] https://machinelearningknowledge.ai/a-brief-history-of-yolo-object-detection-models

[24] https://www.irjet.net/archives/V8/i4/IRJET-V8I4809.pdf

[25] M. Sabokrou, M. Fathy, M. Hoseini, and R. Klette, “Real-time anomaly detection and localization in crowdedness,” in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2015.

[26] C. Lu, J. Shi, and J. Jia, “Abnormal event detection at 150 fps in matlab ,” in Proceedings of the IEEEinternational conference on computer vision, 2013.

[27] Lu, S. (2019). Deep learning for object detection in video Journal of Physics Conference Series, 1176.

[28] Simonyan, K., Zisserman, A. (2014). Two-stream convolutional networks for action recognition in videos.