International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Special Issue of ICRTET-2023, pp 236-241, June-2023

International Conference of Recent Trends in Engineering & Technology (ICRTET-2023)

A Survey Paper on Smart Human Activity Detection Using Yolo

[1]Prof. Aparna Thakre, [2]Atharav Deshpande, [3]Rohan Kadam, [4]Ajay Ujagare, [5]Abhishek jadhav

[1]Profeesor, Computer Engineering, Siddhant College of Enguneering Technical Campus, Sudumbhare, Pune, India
[2,3,4,5]Student, Computer Engineering, Siddhant College of Enguneering Technical Campus, Sudumbhare, Pune, India

***Abstract* - A simple operational model could allow one person to monitor all around us to ensure security and privacy, while maintaining cost and performance of management and getting it right. This inspection with real-time video monitoring function can be sent to hospitals or nursing homes for the sick and elderly, as well as various people working in important area such as airport. In we decided to use the YOLOv4 (You Only One See One) algorithm, which is the newest and fastest of the total algorithms for fast analysis of actions and accurate results when dealing with complex human behaviour. This method uses a bounding box to indicate the action. In these cases, we collected 4,674 different data from different hospitals or different cases, making the most accurate use of one of the largest datasets used in this type of project. When we research, we divide our actions into three different classes: standing, sitting, and walking. Model can control and analyse the activities of many patients or other normal people, and support can monitor the activities of many. After completing three projects, the model achieved an average accuracy of 94.6667.**

Keywords: YOLOv4, DarkNet, Nvidia GPU Driver.

I. INTRODUCTION

Human activity recognition is the study that includes correctly identifying activities performed by humans, tested in different ways. Human activity is the continuous flow of single or distinct action essential in progression. Some human activity specimens are a sequence of steps in which a subject enters a room, walks forward, sits down, stands up, etc. Human activity recognition can widely apply to some real-world applications like patient monitoring, surveillance of essential locations, activity-based search, etc. You can perform it at various abstract levels. Students, engineers, and students have studied human activity recognition in every part of the world for a long time. The Machine Learning-based activity recognition uses Computer vision techniques like YoloV4 and DarkNet to recognize activities performed by humans. We will mainly be focused on the various activities and detect these actions through video. The Human Actions recognized in the videos are based on analyzing a sequence of video frames using a computer to find human activities without manual operations automatically. In this paper, we will be using the YOLO (You Only Look Once) library to build a system that will detect human activities. YOLOv4 is four- time faster, and not only that, we can change between faster speed and better accuracy by just changing the amount and data for the model, without any additional retraining of data required. Human Action Recognition is an area of computer vision research and applications. The goal of Human Action Recognition is to identify and understand people's actions.

II. MOTIVATION

Understanding human activity and their interaction with surrounding objects are crucial for developing an intelligent system. Human action recognition is a field that deals with the problems generated in the integration of sensing and reasoning to provide context-aware data that can confer personalized support across an application. Several issues still plague human action recognition. Such as privacy concerns regarding continuous monitoring of activities, difficulty in performing HAR (Human activity recognition) in real-time, and lack of entirely ambient systems able to reach users at any time. Human activity recognition is a very critical monitoring system. Human action detection aims to inspect exercises from video successions or still pictures. The continuous improvement of artificial intelligence and deep learning algorithm helps us transmit and get vital physiological signs to the medical personnel and simplifies the quantification. As a result, it raises the efficiency of the patient monitoring system. A human activity recognition system can enhance the patient's experience in the medical sector.

Additionally, we can use the technology in many other fields. The active or innovative system can use HAR technology to monitor its residential area for better security. Our research aims to offer medical support, well-being services, and health benefit to older adults and other security purposes for critical infrastructure. It was exciting because we were about to create an intelligent system that would detect human activity and monitor that activity intelligently. That's why we decided to take the challenge.

ISSN (online): 2581-3048
Volume 7, Special Issue of ICRTET-2023, pp 236-241, June-2023
International Conference of Recent Trends in Engineering & Technology (ICRTET-2023)

III. PROBLEM STATEMENT

HAR must recognize human activities by training a machine learning model and displaying activity results per the input activity received from the camera input/video. With this automated Human Activity Detection system, doctors can observe multiple patients simultaneously from their chamber or comfort zone. Doctors also can keep an eye on the duty nurse or staff in the patient's cabin. What are they doing, and are they doing their duty perfectly? It can also use the same system for many different purposes mentioned above. In this paper, we implement the Human Activity Detection system in both still images and video with mentioned three human actions. Further processing, we will be adding more action types. Adding more action types will add more variety.

IV. LITERATURE SURVEY

Human Activity Analysis using Machine Learning Classification Techniques: (Zameer Gulzar, A.Anny Leema, I.Malaserene, December-2019)

In recent times, smartphones have played a vital role in recognizing human activities and have become a well-known field of research. This article provides a detailed overview of various research papers on human activity recognition. The data chosen is multivariate, and we have applied different machine classification techniques Random Forest, KNN, Neural Network, Logistic Regression, Stochastic Gradient Descent, and Naïve Bayes, to analyze human activity. Feature selection reduces the dataset's dimension in addition to building AI models. We calculated precision and recall and made a Confusion Matrix for each model. Experiment results proved that the Neural Network and logistic regression provide better accuracy for human activity recognition than other classifiers such as k-nearest neighbor (KNN), SGD, Random Forest, and Naïve Bayes. However, they take higher computational time and memory resources.

Real-Time Object Detection using YOLO: (Upulie H.D.I, Lakshini Kuganandamurthy, May-2021)

With the availability of enormous amounts of data and the need to computerize visual-based systems, research on object detection has been the focus for the past decade. Since 2012, the growth of Convolutional Neural Networks (CNN) has accelerated this need. With various CNN network architectures available, the You Only Look Once (YOLO) network is famous for many reasons, mainly its speed of identification applicable in real-time object identification. This paper reviews the fundamental structure of CNN algorithms. An overview of YOLO's real-time object detection algorithm and architecture models can remove highlights and discover objects in each given image. These models can solve deformity identification, instructive/ learning application creation, etc.

Latest Research Trends in Fall Detection and Prevention Using Machine Learning: (Sara Usmani, Abdul Saboor, Muhammad Haris, Muneeb A. Khan, Heemin Park, July-2021)

Falls are unusual actions that cause a significant health risk among older people. The growing percentage of old age people requires urgent development of fall detection and prevention systems. The emerging technology focuses on developing such strategies to improve quality of life, especially for the elderly. A fall prevention system tries to predict and reduce the risk of falls. In contrast, a fall detection system observes the fall and generates a help notification to minimize the consequences of falls. Many technical and review papers exist in the literature with a primary focus on fall detection. Similarly, several studies are relatively old, focusing on wearables only, and use statistical and threshold-based approaches with a high false alarm rate. Therefore, this paper presents the latest research trends in fall detection and prevention systems using Machine Learning (ML) algorithms. It uses recent studies and analyzes datasets, age groups, ML algorithms, sensors, and location. Additionally, it provides a detailed discussion of the current trends of fall detection and prevention systems with possible future directions. This overview can help researchers understand the existing systems and propose new methodologies by improving the highlighted issue.

A New Video-Based Crash DetectionMethod: Balancing Speed and Accuracy Using a Feature Fusion Deep Learning Framework: (Zhenbo Lu, Wei Zhou, Shixiang Zhang, Chen Wang, November-2020)

Quick and accurate crash detection is essential for saving lives and improved traffic incident management. This paper developed a feature fusion-based deep learning framework for video-based urban traffic crash detection tasks to balance detection speed and accuracy with a limited computing resource. A residual neural network (ResNet) changed into mixed with interest modules as a part of this framework. To extract crash appearance features from urban traffic videos (i.e., a collision appearance feature extractor), which then used to feed into a spatiotemporal feature fusion model, Conv-LSTM (Convolutional Long Short-Term Memory), to simultaneously identify appearance (static) and motion (dynamic) crash features. +e proposed version changed into trained through a set of videos masking 330 crashes and 342 non-crash activities. In general, the proposed model achieved an accuracy of 87.78% on the testing dataset and an acceptable detection speed (FPS> 30 with GTX 1060). +anks to the

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Special Issue of ICRTET-2023, pp 236-241, June-2023

International Conference of Recent Trends in Engineering & Technology (ICRTET-2023)

attention module, the proposed model can capture crashes' localized appearance features (e.g., vehicle damage and pedestrian fallen-off) better than conventional convolutional neural networks. +e Conv-LSTM module outperformed conventional LSTM in capturing motion features of crashes, such as the roadway congestion and pedestrians gathering after hits. Compared to the traditional motion-based crash detection model, the proposed model achieved higher detection accuracy. Moreover, it could detect crashes much faster than other feature fusion-based models (e.g., C3D). +e results display that the proposed model is a promising video-based urban visitors crash detection set of rules that might use in practice within the destiny.

V. METHODOLOGY

Human activity detection plays an essential role in this modern technology era. It's a large field for research. Nowadays, it has become a rising topic in the human interaction area. For the past decades, many researchers have been working on this topic. Computers don't have the brain to detect anything. They can't read the humans mind. They only give us the output for what we trained for them. If the computer can understand the activity of humans, it can bring a lot of positive changes in the field of IoT. Nowadays, HAR is creating big chaos in the technology field. Methodology of Human Activity Recognition includes several processing steps: taking the input, identifying similar patterns, comparing the frames with the Coco dataset, recognizing the video frames' actions, and detecting those actions using Cnn. Our thesis and research topic is "Human Activity Recognition ."It can be interacted with and implemented with the various algorithms & fields of deep learning, machine learning, Image processing, and neural network. We will use the python programming language to implement our algorithm. We use the YOLOv4 approach for better and faster detection. We use Google Colab for free and quick GPU acceleration, speeding up the data training.

Data Collection Procedure

We had to collect a lot of data in different conditions, complex backgrounds, surroundings, the hospital, and another environment. So that this proposed work can give us the best accuracy at any location, we divide into different groups. We try to capture as much data as possible divided into three groups.

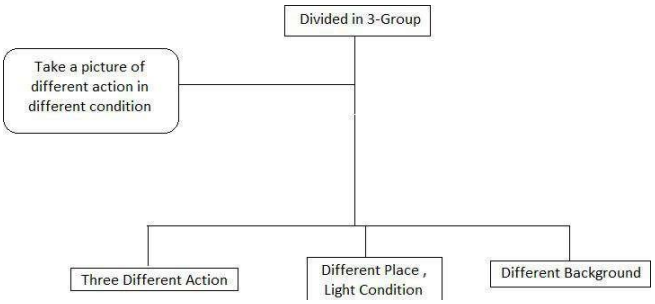


Fig 5.1: Data collection process

Capturing the Frames

The human actions accomplished in a very video enter area unit divided into frames at positive intervals of your time. These frames are captured and taken as input to the CNN model to identify similar patterns by pooling them into certain classes of actions.

Dataset

A Coco dataset consisting of 400 human activities is employed to predict and compare the input data. From youtube recordings, they take the coco datasets. The activities are human focuses and cover a broad scope of classes. It includes human-object communications, such as mowing the lawn, washing dishes, and human actions. For example, e.g., Since the dataset is vast and downloading each clip would be a waste of time, we already have pre-trained models by the original author. It will be easier and provide accurate results when working on the pre-trained model than training and tuning it separately.

YOLO detection architecture

In this topic, we will discuss further YOLOv4 along with the architecture. All of the YOLO data models are activity/object detection datasets, and the training is given to those datasets to search for a subset of the object class.

Most people in the research field are still used to the YOLOv3, which gives us an excellent result. But the YOLOv4 had improved the fidelity and momentum of the two main attributes we generally use to qualify how the architecture and algorithms perform. The YOLOv4 is a further improved approach to object detection, and this applies a single CNN to an entire frame collected from video or just captured by a camera into the grid.

Recognition of the Action

The DarkNet model uses the Coco dataset to compare similar patterns in the input data frames captured in the intervals. CNN can identify similar patterns through pooling

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Special Issue of ICRTET-2023, pp 236-241, June-2023

International Conference of Recent Trends in Engineering & Technology (ICRTET-2023)

layer by layer, and the specified actions assort into classes of human activities.

VI. KEY FUNCTIONS

Pre-trained: To recognize human actions, the model is pre-trained.

Feature Extraction: Similar patterns are identified based on the image frame captured.

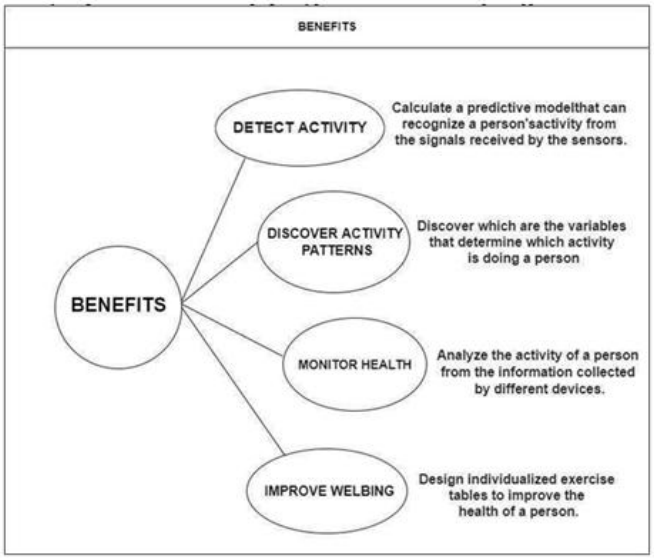
YOLOv4: For Real-Time Object Detection.

DarkNet: For making Real-Time predictions.

VII. OTHER SPECIFICATION

7.1 Benefits

Activity recognition is the basis for developing many potential health, wellness, or sports applications.



Limitations

1. Many Data Required
2. Too many Activity trains overlap the detection.
3. Require very high-performing GPU and CPU for local PC tests.

Applications

1. This HAR system is handy for a health monitoring system.
2. As a surveillance system it can provide intelligence.
3. TI Applications can also use it.

VIII. CONCLUSION

In this paper, the proposed report focuses on Computer Vision to predict the activities of the action on videos. It focuses on recognizing simple activities like normal ones using image processing techniques. This proposed work will provide the required drive for efficiently identifying human action in real-time. We assessed a continuous methodology for human movement identification, picture arrangement dependent on YOLOv4 (You Only Look Once) from complex scenes. The techniques approved with our challenging dataset there are many jumbles and uproarious information for checking more exactness. It can recognize more than one individual's various exercises utilizing additional jumping encloses a solitary picture. In the future, we are planning to add more features in this proposed work that would make this more usable and would revolutionize the human activity monitoring system. Throughout every framework, there is an opening for future development. In the future, this framework will be quicker and more productive, and diminishing handling time is one of the significant issues.

IX. FUTURE WORK

Activity recognition is the basis for developing many potential health, wellness, or sports applications. Collecting various devices' information for health monitoring can be done by analyzing a person's activity.HAR is used to discover similar patterns, which are the variables that determine which action the human performs.HAR can be used for robotic automation, making it easier to train a robot to interact with humans and objects. The appliance of prediction in this proposed work makes it more usable and useful for security purposes. Our future work also focuses on finding a metric to help the action recognition complete within a few frames. Thus the classification stops automatically for the particular action. We will also focus on improving action recognition with the help of object detection in brackets so that it can detect more complex human activities. Movement of objects or Euclidean distance between centers of moving objects and humans can also provide more information about activities occurring in the video.

REFERENCES

[1] Redmon, J., 2020. YOLO: Real-Time Object Detection. [online] Pjreddie.com. Available at: <https://pjreddie.com/darknet/yolo/> (J) [Accessed 6 December 2020].

[2] “Data Generated by new surveillance cameras to increase exponentially in the coming years." [Online, Accessedon 12 March 2018]. http://www.securityinfowatch.com/news/12160483/

ISSN (online): 2581-3048
Volume 7, Special Issue of ICRTET-2023, pp 236-241, June-2023
International Conference of Recent Trends in Engineering & Technology (ICRTET-2023)

data-generated-by-new-surveillance-cameras-to-increase- exponentially-in-the-comingyears

[3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 779-788. DOI: 10.1109/CVPR.2016.91

[4] C Wolf, J. Mille, E. Lombardi, O. Celiktutan, M. Jiu, E. Dogan, G. Eren, M. Baccouche, E. Dellandrea, C.-E. Bichot, C. Garcia, B. Sankur, Evaluation of video activity localizations integrating quality and quantity measurements, In Computer Vision and Image Understanding (127):14-30, 2014.

[5] Limin Wang, Yu Qiao, and Xiaoou Tang. Action recognition with trajectory-pooled deep-convolutional descriptors. In CVPR, pages 4305– 4314, 2015.

[6] Simonyan, Karen & Zisserman, Andrew. (2014). Two-Stream Convolutional Networks for Action Recognition in Videos. Advances in Neural Information Processing Systems.

[7] Roberts Damaševičius, Mindaugas Vasiljevas, Justas Šalkevičius, Marcin Woźniak, "Human Activity Recognition in AAL Environments Using Random proposed workions," Computational and Mathematical Methods in Medicine, vol. 2016, Article ID 4073584, 17 pages, 2016

[8] Vrigkas M, Nikou C and Kakadiaris IA (2015) A Review of Human Activity Recognition Methods. Front. Robot. AI 2:28. DOI: 10.3389/front.2015.00028

[9] Vishwakarma S, Agrawal A. A survey on activity recognition and behavior understanding in video surveillance. Vis Comput. 2013;29(10):983–1009.

[10] Bayat A, Pomplun M, Tran DA. A study on human activity recognition using accelerometer data from smartphones. Procedia Comput Sci. 2014;34:450–7.

[11] Radu V, Lane N. D, Bhattacharya S, Mascolo C, Marina M. K, Kawsar F. Towards deep multimodal learning for activity recognition on mobile devices. In: Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing: adjunct. ACM; 2016, September. pp. 185-188.

[12] Moya Rueda F, Grzeszick R, Fink G, Feldhorst S, ten Hompel M. Convolutional neural networks for human activity recognition using body-worn sensors. In: Informatics, Vol. 5, No. 2. Multidisciplinary Digital Publishing Institute. 2018.p. 26.

[13] Zeng M, Nguyen L. T, Yu B, Mengshoel O. J, Zhu J, Wu P, Zhang J. Convolutional neural networks for human activity recognition using mobile sensors. In: 6th International conference on mobile computing, applications, and services. IEEE; 2014, November. pp. 197-205.

[14] Raptis M, Sigal L. Poselet key-framing: a model for human activity recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013. pp. 2650-2657.

[15] Redmon J, Farhadi A. YOLO9000: better, faster, stronger. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. pp. 7263-7271

[16] Vishwakarma S, Agrawal A. A survey on activity recognition and behavior understanding in video surveillance. Vis Comput. 2013;29(10):983–1009.

[17] Katrina V, Zervakis M, Kalaitzakis K. A survey of video processing techniques for traffic applications. Image Vis Comput. 2003;21(4):359–81.

[18] Roboflow Blog. 2020. Breaking Down Yolov4. [online] Available at: [Accessed 6 December 2020].

[19] Bayat A, Pomplun M, Tran DA. A study on human activity recognition using accelerometer data from smartphones. Procedia Comput Sci. 2014;34:450–7.

[20] Medium. 2020. Introduction To Yolov4: Research Review. [online] Available at: [Accessed 6 December 2020].

[21] Medium. 2020. YOLO — You Only Look Once, Real-Time Object Detection Explained. [online] Available at: [Accessed 6 December 2020]

[22] Redmon, J. and Farhadi, A., 2020. Yolov3: An Incremental Improvement. [online] arXiv.org. Available at: [Accessed 6 December 2020]

[23] T. Choudhury, S. Consolvo, B. Harrison, J. Hightower, A. LaMarca, L. LeGrand, et al., "The mobile sensing platform:

[24] An embedded activity recognition system," IEEE Pervasive Computing, vol. 7, no. 2, pp. 32-41, 2008

[25] G.Akilandasowmya, P.Sathiya, P.AnandhaKumar Human action recognition in the research area of computer vision, IEEE International Conference on automatically detect and retrieve semantic events in the video, 2015 Seventh International Conference on,15-17 Dec.2015

[26] Arie et al. Human activity recognition using multidimensional indexing IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume: 24, Issue: 8, Aug 2002.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Special Issue of ICRTET-2023, pp 236-241, June-2023

International Conference of Recent Trends in Engineering & Technology (ICRTET-2023)

Citation of this Article:

Prof. Aparna Thakre, Atharav Deshpande, Rohan Kadam, Ajay Ujagare, Abhishek jadhav, "A Survey Paper on Smart Human Activity Detection Using Yolo" in proceeding of International Conference of Recent Trends in Engineering & Technology ICRTET - 2023, Organized by SCOE, Sudumbare, Pune, India, Published in IRJIET, Volume 7, Special issue of ICRTET-2023, pp 236-241, June 2023.

****

