Innovation
2021; 2(4): 84-91
http://www.sciencepublishinggroup.com/j/innov
doi: 10.11648/j.innov.20210204.15

Human Activities Detection for Patient Convalescence

Shammir Hossain[1], Yeasir Arafat[2], Shoyaib Mahmud[1], Dipongker Sen[3], Jakia Rawnak Jahan[1], Ahmed Nur-A-Jalal[1], Ohidujjaman[1]

[1]Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh

[2]Department of Information Technology, Asian University for Women, Chattogram, Bangladesh

[3]Department of Information and Communication Technology, Mahbubur Rahman Mollah College, Dhaka, Bangladesh

Email address:

shammir15-1641@diu.edu.bd (S. Hossain), yeasir.arafat@auw.edu.bd (Y. Arafat), shoyaib15-1525@diu.edu.bd (S. Mahmud), dipongker81@gmail.com (D. Sen), jakia15-1578@diu.edu.bd (J. R. Jahan), nahin.ahmed28@gmail.com (A. Nur-A-Jalal), tuhin.iu31@gmail.com (Ohidujjaman)

To cite this article:

Shammir Hossain, Yeasir Arafat, Shoyaib Mahmud, Dipongker Sen, Jakia Rawnak Jahan, Ahmed Nur-A-Jalal, Ohidujjaman. Human Activities Detection for Patient Convalescence. *Innovation.* Vol. 2, No. 4, 2021, pp. 84-91. doi: 10.11648/j.innov.20210204.15

Received: October 28, 2021; **Accepted**: November 15, 2021; **Published**: November 23, 2021

Abstract: A simple activity recognition method can allow a solitary human being to monitor all the surroundings with the purpose to guarantee safety and confidentiality while protective maintenance cost and efficiency with the rising level of accuracy. This monitoring system with real-time video surveillance can be deployed for patients and the elderly in a hospital or old age home and airport along with numerous human activities. For speedy analysis of action and accurate result while working with complex human behavior, we decided to use YOLOv4 (You Only Look Once) algorithm which is the latest and the fastest among them all. This technique uses bounding boxes to highlight the action. In this case, we have collected 4,674 number of dissimilar data from the hospital with different condition of ourselves. During this study, we divided the human action into three different patterns such as standing, sitting and walking. This model is able to detect and recognize numerous patients and other various human activities. This research accomplishes an average accuracy of 94.6667% while recognizing images and about 63.00% while recognizing activity from video clips. This study works with YOLOv4 while it performs better than TensorFlow and OpenPose platforms. The article proposed the outcome for patients in early recovery based on human activities investigation and analysis.

Keywords: Human Activity, Image, OpenPose, Video Clips, TensorFlow, YOLOv4

1. Introduction

Human activity is the continuous flow of single or distinct action essential in progression. Some specimen of human activity is a sequence of actions in which a subject enters in a room, walk forward, sit down, stand up and so more. Human activity recognition is widely applied to some real-world application such as patient monitoring, surveillance of important location, activity-based search and be performed at the various abstract level [8, 10]. Approximately 2.5 quintillion bytes of data produce daily which is increasing day by day [1]. However, from this vast majority of data type the video format is the most produced and monopolized format of them. According to Google the estimated YouTube server size is 1 Trillion GB and about 400+ hours of videos are uploaded on YouTube every minute.

Moreover, according to IDC they expect robust growth of surveillance camera market will be with CAGR of 12.9% for next five years with global revenue of nearly $49 billion by 2025 [2]. The Figure 1 shows the estimated image of near future.

However in a research blog it was published that human took 1,436,300,000,000 photos in 2020. These huge amount of data is least processed which can be used after the process in a different form factor such as surveillance, robotic vision, content-based video search and computer-human interaction. In this study there is mainly focused on the various activities and detection these actions through video to monitor vital physiological sign. The category of patient activities are classified into four levels such as laying down, sitting, standing and walking. This research uses the YOLO (You Only Look Once) library to build a system that detects human activities and monitor the patients. The YOLO library trains on image data and then adjusts the action detection directly is

used in this research. In this study there is used the YOLOv4 as it is extremely quick and precise. The YOLOv4 measured mAP at 0.5 IOU hence the YOLOv4 is four-time faster and it can be changed between faster speed and better accuracy by just changing the amount and data for the model without any additional retraining of data required [3].

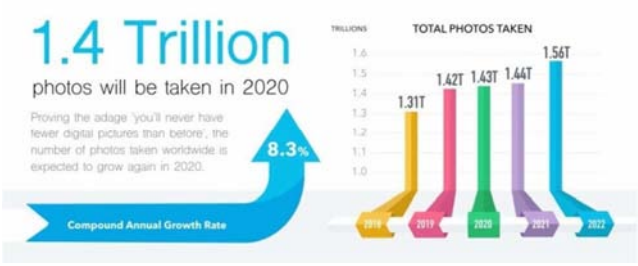

Figure 1. *Estimated image for near future.*

2. Literature Review

Human action detection issue has already achieved importance and been an interesting contemporary topic with regards to its applications in numerous fields such as health, security, surveillance, amusement and intelligent environments.

The authors Vrigkas M. and et al. proposed an article based on a categorization of human activity procedures. They discussed the merits and limitations of human activity methodologies. Moreover, they divided the methodologies into two large categories based on the utilization of data from various modalities. These categories are further analyzed into sub-categorized. At the end they characteristics of the future research direction currently various open matters on human activities recognition [4].

Jamie Shotton and et al. presented a new method to predict 3-dimension positions of body joints from a sole depth image without utilizing temporal information. They proceeds an object recognition method that draws the difficult pose assessment problem into a modest per-pixel classification problem. In this article the huge and extremely varied training dataset permits the classifier to guess body parts invariant to pose, body shape, clothing and so more. Authors generates confidence-scored 3D proposals of numerous body joints by re-projecting the classification outcome and resulting local modes. The assessment of this research shows great accuracy on both synthetic and real test sets, and examines the effect of some training factors [5].

The authors Zeng M. and et al. published an article where proposed a method to extract biased features for human activity recognition. Especially, the authors developed an approach based on convolution neural network (CNN This article focused on CNN-based feature extraction model to extract scale invariant and local dependency. In this study the authors utilized small amount of datasets in experiment part and it needs larger datasets to study the robustness of the proposed method [6, 11].

Raptis M. and Sigal L. develop a novel model for recognition human actions. They cast the learning of keyframes in a max-margin discriminative framework and keyframes are encoded using a spatially-localizable poselet-alike representation with HoG and BoW components learned from weak annotations. This study rely on structured SVM formulation to align the components. In this research the projected method has a quantity of significant benefits with the capability to temporally and spatially limit the action and agreement with partial video streaming. Moreover, this model provides semantically interpretable production in the form of circumstantial time-based orderings of discriminant specific poses [7].

3. Proposed Method

The main focus of this research is to detect the action of the human especially for the patients. The Figure 2 shows the working flow diagram of this study and the Figure 3 explained details in context to individual image separation. The method for the human activity recognition is used for labeling, training and testing is You Only Look Once v4 (YOLOv4). The datasets are collected from various situation of human and the term “Human Activity Dataset” is used for these data.

Figure 2. *Working flow diagram.*

The steps maintain through data pre-processing, training data, weight training and confidence score. The threshold value decides the steps repetition or not. However, the confidence score is less than the threshold value then the steps repeated until to find the true condition. In this article the threshold value is assumed as 50% in respect to the confidence score.

The model is trained with the congruous actions. For testing the model, video frame is inputted for the localization and then recognition of that action. However, the model will be able to

recognize and then give the accuracy outcome.

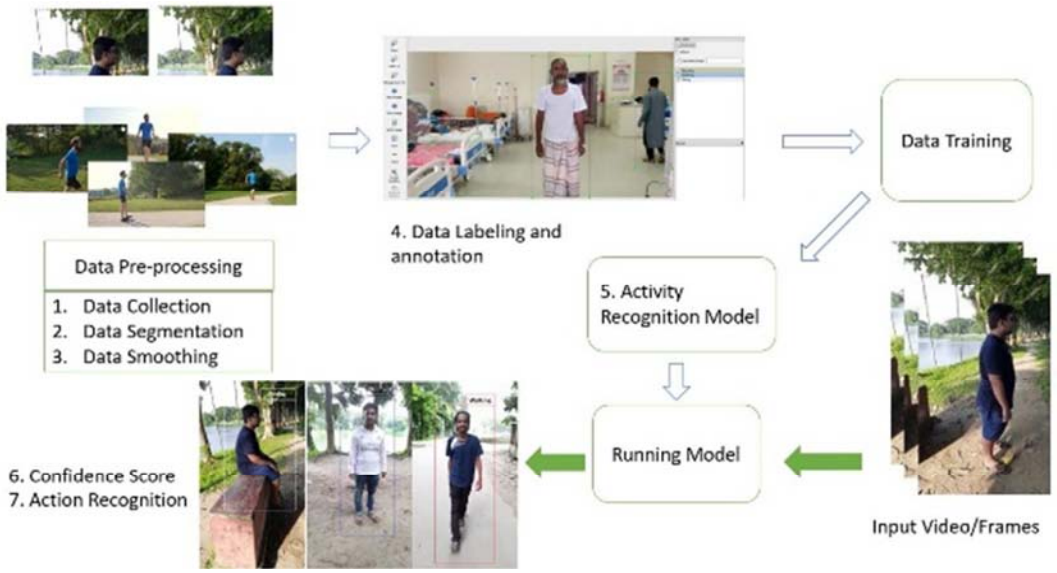

Figure 3. *Working procedure.*

4. YOLO Detection Architecture

In this section YOLOv4 method is discussed along with the architecture. All of the YOLO data models are activity detection dataset. However, those datasets are trained so that it can search for a subset of the object class. The Figure 4 shows the YOLO architecture.

Most of the researcher used the YOLOv3 which was already giving an excellent result. However, the YOLOv4 had improved two main attributes the fidelity and momentum and this study generally uses these two attributes to qualify how the architecture and algorithms perform. The YOLOv4 is further improved in the approach of object detection. This applies a single CNN to an entire frame collected from video or just captured by a camera into the grid. After this, prediction of those bounding box, then classify them into object or action and finally calculate the confidence score in a grid view. The main architecture of YOLO has 24 conventional layers along with two associated layers. YOLO takes an input image and then reside that frame into 448*448 pixels. Nevertheless, the frame gets pre-processed through the conventional network. Moreover, tensor gives accurate information about the coordination of the bounding box and the probability distribution of overall classes and attributes the system is trained for.

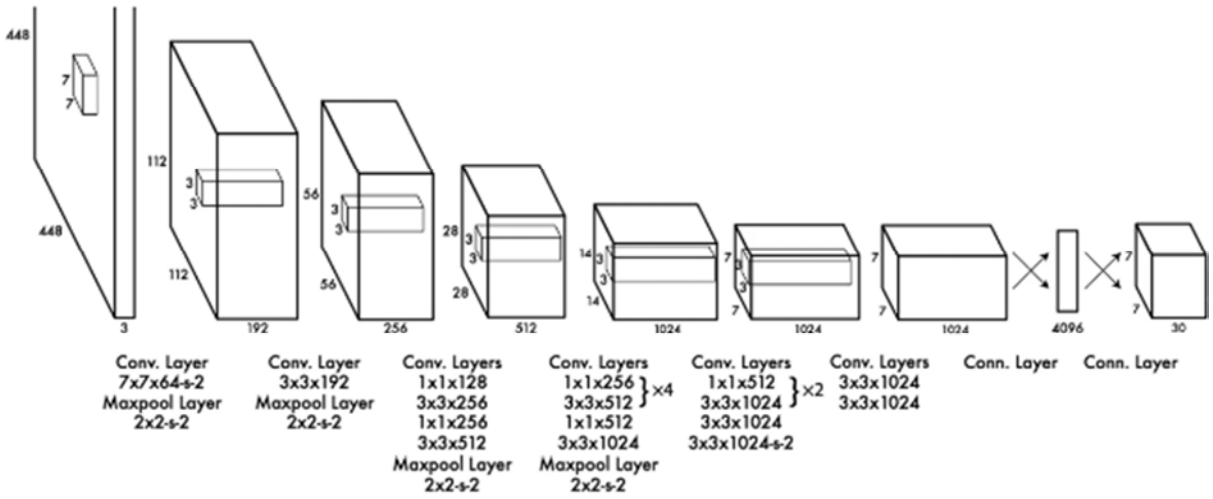

Figure 4. *YOLO architecture.*

Data Pre-Processing

This sub-section will provide the instances or the amount of our model datasets. In this research 4,674 data have gathered. In instances of class 'standing' we had trained about 1838 Images data while for 'sitting' and 'walking' instances we had trained 1705 and 1131 amount of images data respectively. The table 1 shows the details of the dataset.

Table 1. *Details of the dataset.*

Dataset	Training Dataset	Percentage
Total Number of Image	4674	100%
Instances of class "Standing"	1838	39.3239%
Instances of class "Sitting"	1705	36.4783%
Instances of class "Walking"	1131	24.1976%

In terms of percentage standing, sitting and walking class have a data set of 39.3239%, 36.4783% and 24.1976% respectively. This research only focuses on human actions that occur in the hospital or the surrounding. To prepare or train data it needs to classify each data in separate action. For this reason we used labeling that creates a data file which includes the information about the image size and all the action values. The Figure 5 shows the instance of preprocessing image data.

Figure 5. *Instance of preprocessing image data.*

In this study we save the image information into a singl .txt file extinction. YOLOv4 uses .txt file for recognizing each action in that image. The Figure 6 shows the flowchart for training and pre-processing in YOLOv4. The data pre-processing process are given below.

1. Require total number of action class
2. Text file with the same path as all other image
3. Text file with the proper naming of all action class
4. The main path that will contain the weight file
5. A configuration file with all layer described in YOLO architecture sub-section
6. Pre-trained YOLOv4 involutional weights

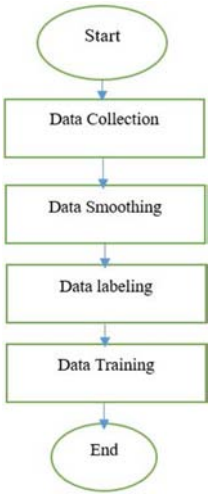

Figure 6. *The Flowchart for training and pre-processing in YOLOv4.*

Figure 7. *YOLOv4 CFG weight.*

The Figure 7 indicates the CFG diagram for training the proposed model. This CFG model is captured when the system completed the iteration of 1300 cycle. However the moment represents the current average loss 2.6786 which is significantly less than others.

YOLOv4 Loss Function

In this research we calculated the loss function of YOLOv4 which diminish the actual normalized distance between the dataset frame and target frame which is able to bring out convergence momentum. The DloU misfortune is on the fundamental of Intersection over Union (IoU) which signify the middle distance of that particular bounding box. It indicates the following formula (1) where B^{gt} is the target bounding box and B is the prediction box. However, the loss of function LoU is defined in formula (2). It shows the work function if bounding box overlap else there is no overlap if the gradient does not change.

$$IoU=\frac{B\cup B^{gt}}{B\cap B^{gt}} \quad (1)$$

$$L_{LoU} = 1 - \frac{B\cup B^{gt}}{B\cap B^{gt}} \quad (2)$$

The GIoU function improves the loss of LoU in the case that the acclivity does not change until overlapping another box which add some loss term for the function IoU. It is noted as the equation (3). However, one of B or B^{gt} countermands the other bounding box, the penalty terms will not work which is defined as an IoU loss.

$$L_{GloU} = 1 - IoU + \frac{|C-B\ \cap B^{gt}|}{|C|} \quad (3)$$

To solve constraint, DIoU function was set on the motion which can be seen in the formula (4); where b and b^{gt} indicate the center point of that anchor image and targeted image one after another while the attribute p indicates the Euclidean

distance between two centers while c represents the minimum rectangle distance that covers anchor and the targeted box. The loss function of YOLOv4 for DIoU can be noted as equation (5).

$$R_{DIoU} = \frac{p^2\,(b,b^{gt})}{c^2} \tag{4}$$

$$L_{GloU} = 1 - \text{IoU} + \frac{p^2\,(b,b^{gt})}{c^2} \tag{5}$$

The CIoU is also introduced to the loss equation. There are a few upper hands in the CIoU loss function. The equation (1) can increase the overlap area in both the ground truth box and the prediction box. Moreover the equation (2) can minimize the actual distance for the focal point. The equation CIoU can be noted as formula (6) based on the formula (4) which can be calculated; where α represents an upper hand trade-off and 'v' is called in formula (7) to measure the stability of the overall aspect ratio added.

$$R_{CloU} = \frac{p^2\,(b,b^{gt})}{c^2} + \alpha\upsilon \tag{6}$$

$$\text{V} = \frac{4}{\pi^2}\,(\arctan\,\frac{\text{W}^{\text{gt}}}{\text{H}^{\text{gt}}} - \arctan\,\frac{\text{w}}{\text{h}})\text{\^{}}2 \tag{7}$$

The loss equation of the CIoU can be defined as for formula (8) and the main variable can be called at equation (9).

$$L_{GloU} = 1 - \text{IoU} + \frac{p^2\,(b,b^{gt})}{c^2} + \alpha\upsilon \tag{8}$$

$$\alpha = \frac{\upsilon}{(1-LoU)+\upsilon} \tag{9}$$

The DIoU loss equation could be applied in NMS (Non-minimum Suppression) to delete unneeded bounding box. In this equation, both the distance of the action detection box and the center point of that bounding box is considered which can constructively above two-loss equation mistakes [3, 4].

YOLOv4 Activation Function

The activity function or equation that runs on deep learning neural network which is in charge for mapping the input of the neuron to the output [9]. Its main task is to expand the nonlinear change of the neural network dataset. The function is shown Figure 8.

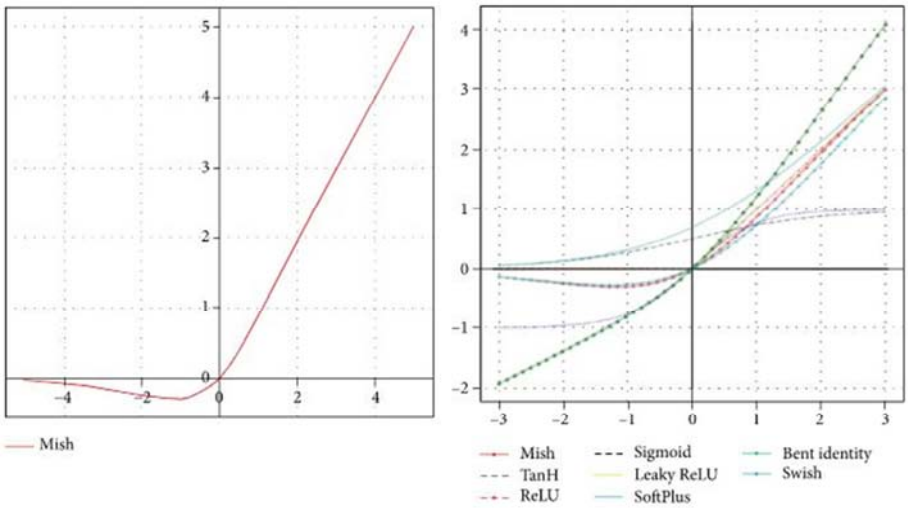

Figure 8. *YOLOv4 activation function.*

In the proposed model the Mish activation obligation is the activation function which we utilize replacing Leaky ReLU that is a very tiny constant leak that has the much-updated function of ReLU with Mish in YOLOv3 Leaky ReLU is the self-regular non-monotone deep learning neural activation and smooth activation function allowing the instruction into the deep learning neural network to obtain for preferable accuracy and generalization [12]. It is defined as equation (10). It shows that c(x)=ln (1 + e^x).

$$\text{F(x)=x. tanh(c(x))} \tag{10}$$

This equation is specific then swish defined as equation (12) and ReLU defined as formula (11) when performing on the experiments.

$$\text{F(x)=max (0, x)} \tag{11}$$

$$\text{F(x)=x. sigmoid(x)} \tag{12}$$

5. Experiment and Result Discussions

This study trained the system with a huge amount of data collected from different situation and features of actions d of the data datasets. This model gets the average accuracy of about 94% after training the data approximate 13th iteration. This accuracy is even improved and mainly depends on interaction where researchers perform with their taste of topic. We had to calculate the reliability and fidelity of the human activity recognition and patient monitoring system to gather and compare the result of the experiment. We have collected the data in numerical value from the numerous tests and then we use the average accuracy and the mean value for further comparison with another model. The average accuracy is used to calculate the independent action's unassisted while the

mean average accuracy is used to calculate the model's fidelity combined.

Experiment Results

YOLOv4 is the most sophisticated algorithm that handles the empirical part easily. We trained the model to detect three action class such are 'standing', 'sitting' and 'walking'. However, it is noticed that the proposed model is performing better than any other model while it is running and optimizing. The table 2 shows the test data for detection of action from still image.

Table 2. *Validation Set Accuracy for Still Image.*

Action Class	Average Accuracy	Mean of Average Accuracy
Sitting	95%	
Walking	96%	94.6667%
Standing	93%	

While working on the research, we tried to apply two more approach. However, mainly focused on the YOLOv4 for the best accuracy and performance. The table 3 shows the test data for detection of action from video file.

Table 3. *Validation set Accuracy for Video File.*

Action Class	Average Accuracy	Mean of Average Accuracy
Sitting	58%	
Walking	70%	63.00%
Standing	61%	

TensorFlow is an Open source machine learning framework of google for the programming of data flow across a variety of task. Tensors are just multidimensional arrays and expansion of 2-dimensional tables to higher dimensional data. TensorFlow has many characteristics that make it appropriate for human activity detection. Creating a reliable machine learning (ML) models which are capable of understanding and localizing multiple activity in a single image remained a key challenge in computer vision. However, including recent advances in deep learning; activity detections are simple to build than ever before. The activity detection API of TensorFlow is an open source platform developed on top TensorFlow that makes it simple to build, train and deploy models for activity detection. We build the first activity detection model with TensorFlow activity detection API. In the Figure 9 it is noticed that, the TensorFlow detects properly the human's activity with great accuracy.

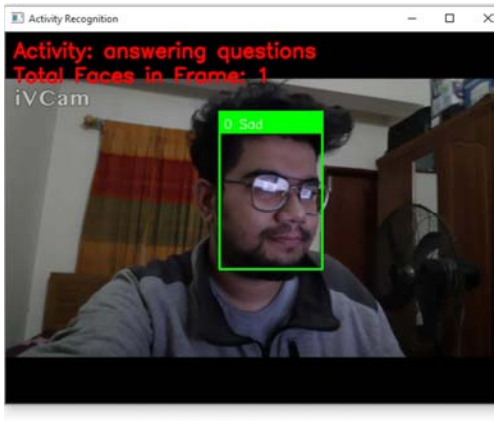

Figure 9. *Human activity detection using TensorFlow.*

OpenPose is the first multi-person real time platform to collectively detect key points for the human body, hand, face and foot (135 key points in total) on a single image. Researcher at Carnegia Mellon University suggested this method as well [11]. The Figure 10 shows the human activity detection using OpenPose.

Figure 10. *Human activity detection using OpenPose.*

Using the OpenPose, pose estimate and detection has been minimally implemented. The MobileNets (a CNN originally trained on the ImageNet wide visual detection task dataset) or binary classification of poses (sitting or upright) was retrained (final layer) on a data set.

We have run all those models and then gathered data result to find out the most efficient approach. The table 4 shows the comparison measurements of different platform.

Table 4. *Comparing YOLOv4 project with others platform.*

Attribute	Project -1	Project-2	Project-3
Training Platform	TensorFlow	YOLOv4, Danknet, OpenCV, NumPy	OpenCV OPENPOSE, Motplot, NumPy
Data	Google Pretrained Weight	Manually anoted- 6000 Data 2000 Data 1000 Data	Live Data using 2D Mapping
Detection Platform	Local + Live	Local Live (possible)	Local + Live
Accuracy	Overall: 80%	1000 Data: 85% 2000 Data: 95% 6000 Data:	Overall 90%
Status	Good	Strong	Strong

The merits and demerits of the three models are as follows in table 5.

Table 5. *Comparing advantage and disadvantage of YOLOv4 project with others platform.*

Attribute	Project 1	Project 2	Project 3
Advantages	Can detect emotion.	Faster training time using CoLab.	It can be implemented for prediction model.
Drawbacks	Run slow on GPU and CPU. Low frame rate. No prediction on next activity. Require high performance GPU.	Many data require. Too many activity trains overlap the detection. Require very high performing GPU and CPU for local PC test.	Require high configure PC. Low frame rate.

However, observing the above result it is noticed that all of those models have their own advantages. Moreover the activity detection model using the YOLOv4 has the most meaningful purposes. It can be implemented with less configuration on CoLab unlike others.

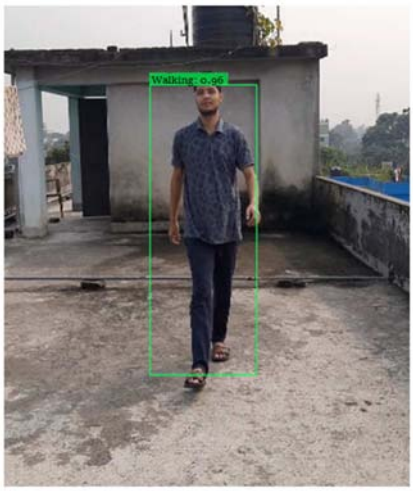

Figure 11. *Action detection of walking.*

Figure 12. *Action Detection of Sitting.*

Descriptive analysis of our result

The dataset are divided into three parts while training the data. The first 1000 data named as test data_01and secondly the 2000 data renamed as test data_02 and, finally we trained all the data that had in our disposal. However, training with largest dataset the system gives us the best accuracy. In the table 2 we found the test accuracy for 'walking' is about 96% which is the highest average precision in the test model. The output for action walking in still image detection is in Figure 11.

In the table 2 we found the test accuracy for 'sitting' is about 95% which is the second highest average precision in the test model. The output for action 'sitting' in still image detection is in Figure 12.

Finally, for the standing class which perform an average precession of 0.9300 which means the action has average accuracy of 93%. This action class had the lowest accuracy of all our action class. The output for action 'standing' in still image detection is in Figure 13. However we get the average accuracy of 94.6667% which is on the top of the line comparing with other research outcome.

Figure 13. *Action detection of standing.*

Moreover it is noticed that the highest accuracy on every action recognition class is in the still images. The trained model can detect all action even in the most complex and congested images where there are a lot of people. The accuracy of the study for video file from table 3 it is found that the action class sitting have accuracy of 58% while walking and sitting have the accuracy of 70% and 61% respectively. The average detection accuracy is 63%. The Figure 14 shows the action detection from video file.

Figure 14. *Action detection from video.*

6. Conclusion

In this research continuous methodology is assessed for human movement identification and image arrangement dependent on YOLOv4 from complex scenes. The procedure approved with the difficult dataset where many jumble and uproarious information for checking more accuracy. The method recognizes more than one individual's various exercises utilizing additional jumping encloses a solitary picture. Numerous activity recognition methods and a few research points that are connected to activity investigation in 'still images' have been discussed in this article. The research concludes that the human activities analysis predicts the patient's status and advises to take the further necessary steps for preserving the health condition.

In future we are planning to add more features in this study that would make this more usable and would revolutionize human activity monitoring system. In this regard there will be implemented more datasets about patients current condition which will detect patient's injury, tension and so more. Moreover, in the future more complex data will be added and prediction can be implemented which will improve and add utility to this research.

References

[1] R. Devakunchari, "Analysis on big data over the years", International Journal of Scientific and Research Publications, Volume 4, Issue 1, January 2014.

[2] Mike Jude, "Worldwide Video Surveillance Camera Forecast, 2020–2025", International Data Corporation (IDC), July, 2021.

[3] Zicong Jiang and et al., "Real-time object detection method based on improved YOLOv4-tiny", Computer Vision and Pattern Recognition, Cornell University, 2 Dec 2020.

[4] Vrigkas M. and et al., "A Review of Human Activity Recognition Methods", Frontiers in Robotics and AI, Nov., 2015.

[5] Jamie Shotton and et al. "Real-Time Human Pose Recognition in Parts from Single Depth Images," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Colorado Springs, CO, USA, 2011.

[6] Zeng M. and et al., "Convolutional Neural Networks for Human Activity Recognition using Mobile Sensors", 6th International conference on mobile computing, applications and services. IEEE; 2014, November. pp. 197-205.

[7] Raptis M. and Sigal L., "Poselet Key-framing: A Model for Human Activity Recognition" Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 2650-2657.

[8] Ong Chin Ann and Lau Bee Theng, "Human Activity Recognition: A Review", IEEE International Conference on Control System, Computing and Engineering, 28 - 30 November 2014, Penang, Malaysia.

[9] Daniele Ravi and et al., "Deep Learning for Human Activity Recognition: A Resource Efficient Implementation on Low-Power Devices", IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN), 2016.

[10] Faeghe F. and et al. "Human Activity Recognition: From Sensors to Applications", International Conference on Omni-layer Intelligent Systems (COINS), 2020.

[11] Ankita and and et al., "An Efficient and Lightweight Deep Learning Model for Human Activity Recognition Using Smartphones", Sensors, Publisher: MDPI, 2021.

[12] Shujuan Wang and Xiaoke Zhu, 'A Hybrid Deep Neural Networks for Sensor-based Human Activity Recognition", 12th International Conference on Advanced Computational Intelligence (ICACI), 2020.