A YOLO-Deep Learning Algorithm for Improved Detection of Cell Phone Usage in Restricted Places

Rida Ayesha[*], Najeed Ahmed Khan[**], Usman Inayat[*]

[*] Department of Informatics & Systems, University of Management and Technology, Lahore, Pakistan.
[**] Department of Computer Science, NED University of Engineering and Technology, Karachi, Pakistan.

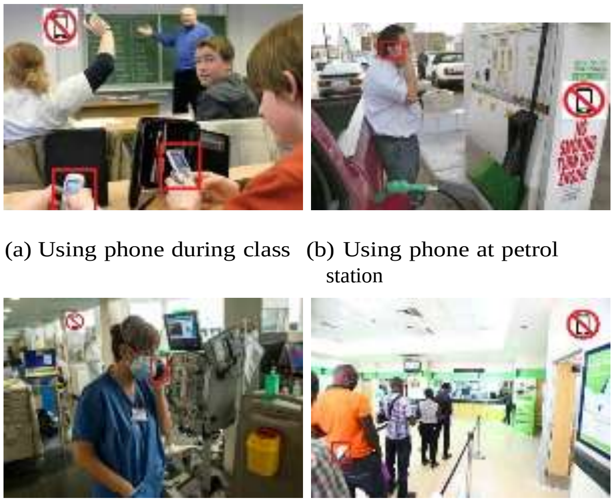
Corresponding Author: Usman Inayat

Abstract- Due to the close relationship of object detection with video analysis and image perception, it has gained a great deal of momentum in research over the years. In computer vision, detecting human-object interactions is a fundamental problem because it provides semantic information about the interactions between detected objects. To process an extensive amount information from the video data, a deep learning framework using YOLOv3 and YOLOv4 for the problem of human-object interaction detection is used. Real life scenarios containing human activities (using cell phone) recorded via camera can be addressed via static images, videos, real-time webcam, and real-time CCTV surveillance. All these possible dimensions have been covered in this paper along with the count of mobile phone users. The use of mobile phone action recognition in prohibited areas has been addressed by the detection of objects predicted by the bounding box. Two public datasets, HICO-DET and MS-COCO are used for the training and evaluation of the model. Experimental results and analysis are produced to compare the YOLOv3 and YOLOv4 algorithms before and after applying 4k cross-validation. Since, YOLOv4 is an improvement on the YOLOv3 algorithm, it requires high end machine with GPU where as YOLOv3 is compatible with commonly available machine. Therefore, we have compared our results of commonly available machine and for a dedicated machine. The results indicated that the performance of YOLOv4 is far better than that of YOLOv3. The limitations of the existing framework and some improvements on it are suggested in this research paper.

Index Terms- Action recognition, Computer vision, Deep Learning, Human interaction, Human action recognition, Object detection.

I. INTRODUCTION

Action recognition [1, 2] is of great importance in understanding human motion from video. It is an important topic in computer vision [3] due to its many applications such as video surveillance, human- machine interaction, and video retrieval. Human activity monitoring in the video sequences is an intriguing computer vision domain that incorporates colossal applications, e.g., surveillance systems, human computer interaction, and traffic control systems. Recently, human-object interaction or object identification has received attention and interest from the community of scientists and the general public. The general public interest is majorly due to the recent terrorist incidents around the world which cause the increase in demand for efficacious security systems. In recent years, studies have also focused on issues related to the use of mobile phones in restricted, prohibited, and unauthorized areas. The rapid explosion of mobile phones in the early 21st century eventually raised issues such as the potential use of privacy encroachment or widespread academic fraud. Usage of mobile phones in restricted, prohibited, and unauthorized premises such as Examination venues, Banks, Prisons, Airports, Hospitals, Petrol stations, Muse- ums, Mosques, Military campuses, Defense and Security Agencies, and Offices will have an undesirable impact. Figure 1 (a), (b), (c) & (d) shows the use of mobile phones at prohibited areas.



(a) Using phone during class (b) Using phone at petrol station

(c) Using phone at hospital (d) Using phone inside bank

Figure 1: Mobile phone usage at prohibited area

The main goal of this research was to help detect human interaction with a mobile phone. Automated mobile phone detection is the process of identifying, analyzing, and comparing one or more individuals in a video using a mobile phone. A solution to this will be of great importance and beneficial for the general public. However, with all the potential benefits of to- day's mobile phones, researchers, designers, and government leaders are still looking for the procurement of mobile phone detectors in all restricted areas. The use of automated mobile phone detectors is necessary and timely to prevent the unauthorized use of mobile phones in all restricted areas. Therefore, this study proposed simple automated mobile phone detection systems.

The organization of this paper is as follows: Section 1 includes the introduction and background of the problem, highlighting the negative and positive impact of the usage of mobile phones also talking about the issues caused by using the mobile phone in prohibited areas. Section 2 includes the literature review and the related work of recent years. Section 3 highlights the implementation details and provides the workflow breakdown. Section 4 includes the experimental results achieved by the trained model on various sources. The model is then evaluated using machine learning metrics, and a comparison between the outcomes of used algorithms is provided. Section 7 concludes the research with the limitations of the work and suggestions for improvements.

II. RELATED WORK

Human object interaction recognition [4] is a popular research area in the domain of computer vision and has been studied due to its prime applications involving visual surveillance and video retrieval. A new way to represent the power of human inter- action in videos was introduced by Victor and Juan [5] as earlier algorithms focused on simulating spatial relationships between objects and characters but ignored the emergence of these relationships through time. Their algorithm captured the dynamic nature of human interaction by modeling how these patterns emerge in relation to time. Their studies have shown that coding such an evolutionary process is essential to properly distinguish human actions that involve the same things and spatial relationships between people but differ only in the temporal aspect of interaction, e.g. answer the phone and dial the phone. They validated their approach to the two databases of human activity and demonstrated performance improvements over state of the art algorithms.

Wang et al. [6] argued that the real meaning of the action lies in the change or change brought about in the environment. They have raised the novel representation of actions by modeling an action as a change that changes the state of the environment be- fore the action takes place (first state) and after the action (effect). Motivated by the latest advances in video representation using in deep learning, they de- signed the Siamese network model as a transition into a high-level feature. Their model provided improvements to standard action recognition databases including UCF101 and HMDB51. Also, their approach was able to do things normally beyond the categories of action learned and showed a significant improvement in the performance of the different categories in their new ACT database.

Muhammad Sharif et al. [7] has come up with a proposal for a mixed strategy to classify human activities into a given video sequence. The proposed method consists of four main steps: (a) separate moving objects by combining the same novel splits with expected magnification, (b) develop a new set of integrated features using local binary patterns with histogram oriented gradient and Harlick features, (c) Feature selection with grade novel the Euclidean and integrated approach based on entropy-PCA, and (d) feature classification using a multi-stage vector mechanism. The classifier was trained for human action classification on three benchmark datasets (MIT, CAVIAR, and BMW-10). For testing, walk-in videos with multiple cameras and MSR, INRIA, and CASIA action databases were used. In addition, the results were further validated using a database recorded by their research team. For action recognition, four publicly available data sets were selected such as Weizmann, KTH, UIUC, and Muhavi.

Yan et al. [8] has proposed a new way of identifying human interactions based on a 2D network of convolutional neural network, which includes human body movement, human hand movements, and object recognition network. Through the use of RGBD cameras and digital gloves, refined recognition of the human body and hand movements were collected and studied. In addition, a new YOLOv3-based object recognition network was launched that helped increase the accuracy of predicting human interaction labels. They designed the actions of eight representatives and built their own set of data containing body and precise hand movements. They gained good ac- curacy in their assessment of the action recognition, which demonstrated the validity and performance of the proposed multi-tasking framework.

Egocentric Vision is an emerging platform of computer vision that is characterized by the acquisition of images and video by first-person viewing. Georgios et al. [9] has faced the challenge of recognizing self-conscious human action through the presence and location of the identified regions of interest in the segment, without further use of visual elements. Initially, they realized that human hands are important in performing actions and focused on finding their movements as key symbols that describe actions. They use acquisition techniques and regional tracking techniques to find their hands and record their movements. Previous information about ego- centric views helps to identify the left and right hand. In terms of detection and tracking, they have pro- vided a successful pipeline that worked on unseen videos to get the wearer’s hands on the camera and assemble them over time. In addition, they emphasized the importance of place information in order to recognize action. They agreed that the presence of objects is important in the actions of people and helps to define the place better. To get this information, they have used object detection in some classes that correspond to the actions they wanted to see. Their experiments focused on kitchen activity videos from the Epic-Kitchens database. They rated the recognition of the action as a problem of learning the sequence of the found areas in frames. Their results have shown that the discovery of hands and objects with no other visible information that can be relied upon to distinguish human hand-related actions.

Migual et al. [10] has proposed a way to deal with substandard actions by combining the knowledge of the joints of human body to aid in the perception of action. This was achieved using advanced features integrated into a pre-trained convolutional neural net- work on ImageNet dataset, with body joints described as low-level features. These features are then as- signed to the Short-term Memory Network to study the temporal dependencies of the action. To get a pose prediction, they focus on a clear relationship between body joints. They used a series of residual auto-encoders to generate multiple integrated predictions to provide a possible map of the body joints. In network topology, features were analyzed in all scales that captured the various local relationships associated with the body. The repeated bottom-up and top-down processing was applied by supervising each auto-encoder network. Preliminary results were achieved in the popular data sets of FLIC, LSP, and UCF Sports.

Despite advances in recognition of Human Activity, the capability to get benefit from the dynamic of movement of human body in videos is not obtainable yet. In modern works, many researchers have used visual acuity and movement as a stand-alone device to incorporate action into a particular video. Sadjad et al. [11] highlighted that while using the novel representation of human body movement, the benefits from observation and movement at the same time can be achieved resulting in better performance recognition function. They started with a standing position to take out the location and a heat map of the body joints in each frame. The powerful encoder has generated a fixed size suspension from these integrated body heatmaps. The experimental results concluded that training the convolutional neural network with the representation of dynamic movements, outperforms the state of the art action recognition models. Excellent performance has been achieved in the HMDB, JHMDB, UCF-101, and AVA databases by modeling visual functions as separate dynamic systems and with the help of two stream networks.

Ali and Lee [12] suggested the use of semantic imagery, an improved representation of video analysis, primarily in combination with inception networks. A semantic image was obtained using a few geographical classifications using global clustering (LSSGC) before the standard integration rate that summarizes moving elements with one or more images. It covers the background information by covering the vertical background from the window to the frames divided into subsequent segments. Their idea was to improve action flexibility by focusing on the region important for action recognition and coding temporary variations using a frame rate measurement. The successive combination of Inception-ResNetv2 and short-term memory network (LSTM) to improve the temporary variability of enhanced recognition functionality was also proposed. Extensive analysis was performed on the UCF101 and HMDB51 databases which are widely used in observational awareness studies. They showed that (i) the semantic image makes it more efficient and adapts faster than its original variant, (ii) the use of pre-level differentiation produces better recognition performance, (iii) The use of LSTM raises the knowledge of temporary variability from standardization to make the action model better than the basic network, (iv) the proposed presentations can be flexible as they go along with existing methods such as part-time networks to improve recognition performance, and UCF101 and HMDB51.

Mathe et al. [13] introduced a method for the detection of human activity of daily activities (ADLs) using the Convolutional Neural Network (CNN). The network was trained on Discrete Fourier transform (DFT) images resulting from raw sensor reading, that is, each individual action was described by an image. Specifically, they worked using 3D skeletal positions for human joints, ranging from raw RGB sequence analysis to in-depth knowledge. The movement of each joint was defined by a combination of 3 1D signals, representing their intermediaries in the 3D Euclidean space. All such signals from a set of human members have been assembled to form an image, which has been modified by DFT and used to train and test CNN. They evaluated their approach on a publicly available dataset of human actions that may involve one or more body parts at a time and two sets of actions similar to regular ADLs. The variety of human activities in everyday life makes the detection process difficult and complicated. A completely new automated system has been suggested by Muhammad Attique et al. [14] to gain human action recognition through a combination of deep neural network (DNN) and multi-view features. DNN features were extracted using a pre-trained CNN model called VGG19. Later, multiple viewing features were computerized using horizontal and vertical gradients, as well as vertical directional features. Subsequently, the components were grouped together to determine the main features, selected using three parameters namely related entropy, shared information, and a solid integration coefficient (SCC). In addition, these parameters are used for the selection of the smallest set of features for maximum function-based functionality. The last selected features were given to the Naive Bayes category for final recognition. The proposed program has been tested on five databases named HMDB51, UCF Sports, YouTube, IXMAS, and KTH. The results showed that the proposed system transcends modern method.

III. IMPLEMENTATION DETAILS

The following two public datasets were used for the training of YOLOv3 and YOLOv4 algorithm.

1. HICO-DET
2. MS-COCO (2014)

Table 1: Table to describe datasets

S. No.	Dataset	Total Train Images	Total Test Images
1	HICO-DET	38,116	9658
2	MS-COCO	82,783	40,775

The figure 2 (a), (b), (c) & (d) shows few samples from both datasets.

(a) Image from HICO-DET dataset

(b) Image from HICO-DET dataset

(c) Image from MS-COCO (2014) dataset

(d) Image from MS-COCO (2014) dataset

Figure 2: Sample images from HICO-DET & MS- COCO (2014) datasets.

The collected dataset for mobile phone detection majorly contains close-up images. The total images obtained from both datasets were more than 1.20 lac. But these images were a fix of different classes/activities thus, filtering was required to select the relevant images (images related to the set objective, i.e., mobile phone usage). Another issue was the quantity of the filtered images. The selected images, after performing the filtering process on both datasets were quite a few. Therefore, an image augmentation process was applied to the training and testing data for classes of 'Phone' and 'Using- phone'. A similar methodology and approach have been discussed by Hao-Shu Fang et al. [15]. No image augmentation was performed for training and testing data for the classes of 'Not-Phone' and 'Not- Using-phone'. The image annotation and augmentation were performed on the Roboflow website. Since, the relevant images were few, all images from the HICO-DET train and HICO-DET test was used for training. However, for the MS-COCO dataset, the train images were utilized for training and test images for testing [16, 21]. A tabular representation for the description of the dataset and dataset augmentation is given in Tables 1 and 2.

Table 2: Dataset augmentation description

S. No	Dataset	Augmentation Applied on TrainSet	Total Augmented Train Images	Augmentation Applied on TestSet	Total Augmented Test Images
1	HICO-DET	blur,brightness,noise, resize, rotate 90	1791	None	None
2	MS-COCO	blur,brightness,noise, resize, rotate 90	4184	blur, noise, rotate	1457

In the holdout method, the dataset was divided into two parts i.e. 80% for training and 20% for testing whereas, in the K4 cross-validation method, the dataset was split into four equal portions with the holdout method being repeated four times such that every time, one portion was used for testing while the remaining 3 portions were used for the training. A tabular representation of the dataset division is given in table 3.

Table 3: Total images used description

S.No.	Method	Total Images	Total Train Images Used	Total Test Images Used
1	Holdout	7432	5975	1457
2	K4 cross validation	7432	5574	1858

The concept of transfer learning was employed in training. Rather than training the model from scratch, pre-trained YOLOv3 and YOLOv4 weight were utilized, which have been trained up to 137 convolutional layers. The weights were saved every 100 iterations so that if the training was interrupted, it could restart from the last saved weights. The training of YOLOv3 and YOLOv4 was performed on GPU using Intel(R) Xeon(R) Core i7 clocked at 1.7 GHz with 32 GB RAM and NVIDIA GeForce GTX 1080 Ti. The training time taken by the YOLOv3 and YOLOv4 algorithm to complete 8000 iterations was around 12.5 – 13.0 hours and 11.0 hours, respectively. The average loss error obtained after the training of YOLOv3 and YOLOv4 was 0.33 and 1.65, respectively. After completion of the training process of both algorithms i.e. YOLOv3 and YOLOv4, a trained model was acquired. The model was tested on various sources for evaluation purposes. Figure 3 represents the whole implementation process.

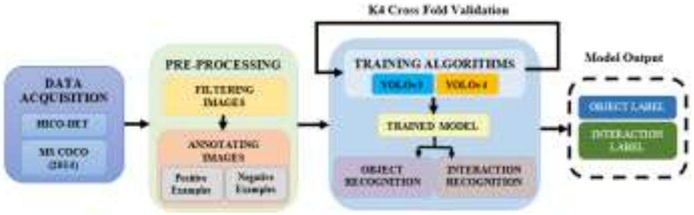
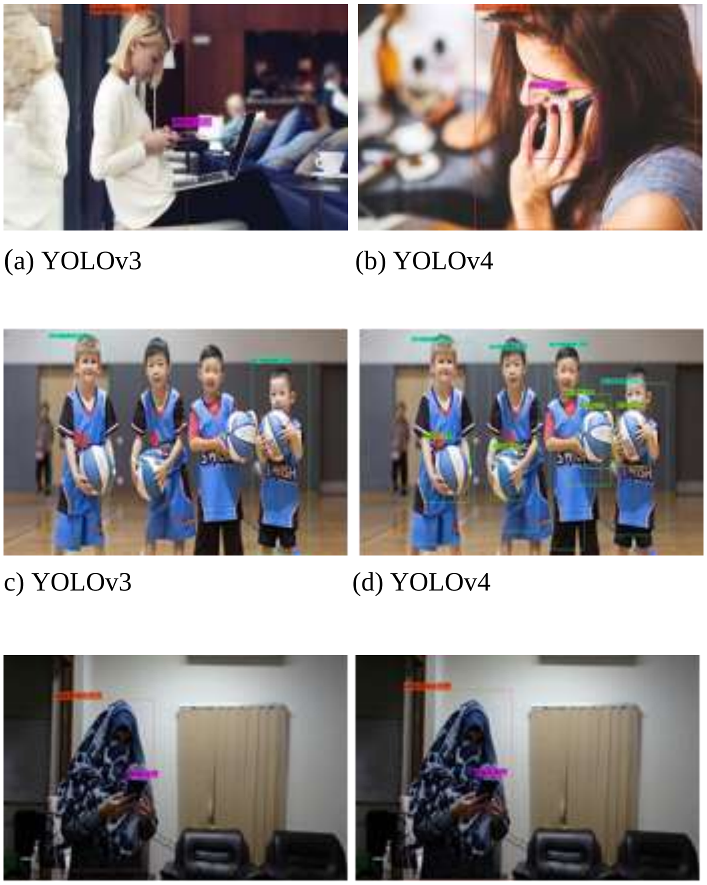


Figure 3: The implementation process

IV. EXPERIMENTAL RESULTS

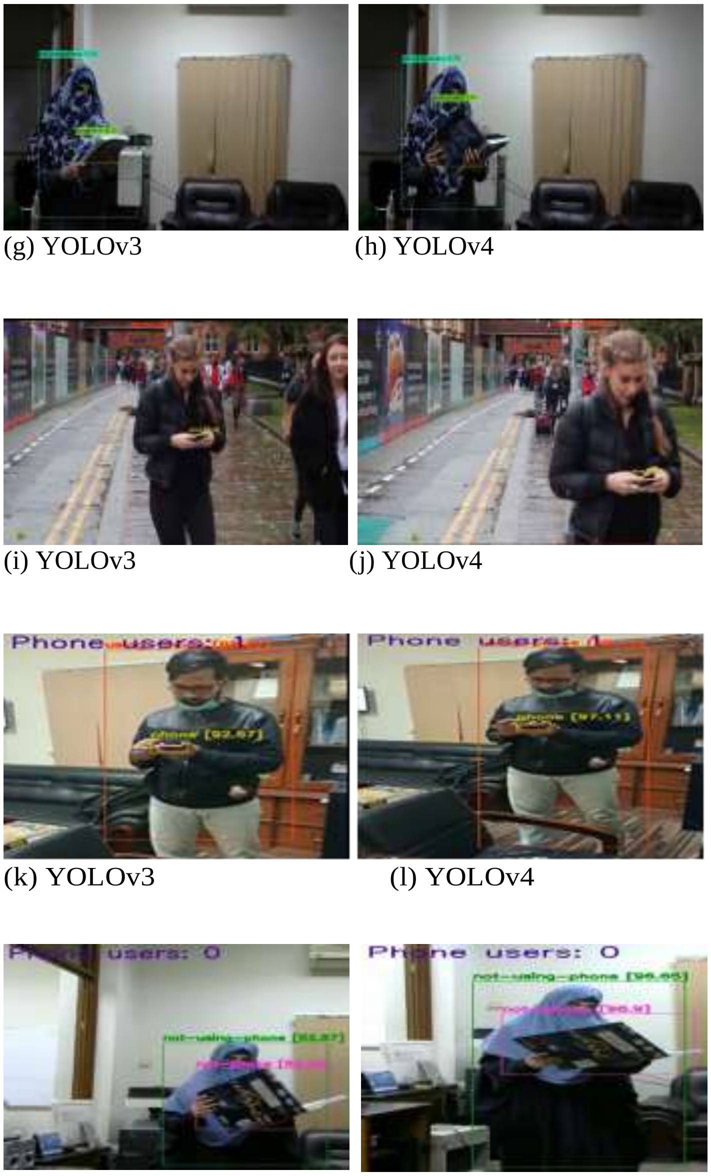
The trained model was tested and evaluated on various dimensions like static images, webcam images (taken from Logitech 720p), static videos, custom videos (using 64 MP camera), real-time webcam, and real-time CCTV surveillance. The obtained detection results of both YOLOv3 and YOLOv4 are displayed and compared in figure 4 and table 4. Missed detection can be seen in the result of YOLOv3. Therefore, it is apparent that the results of YOLOv4 are far better than YOLOv3.



(a) YOLOv3 (b) YOLOv4

c) YOLOv3 (d) YOLOv4

e) YOLOv3 (f) YOLOv4

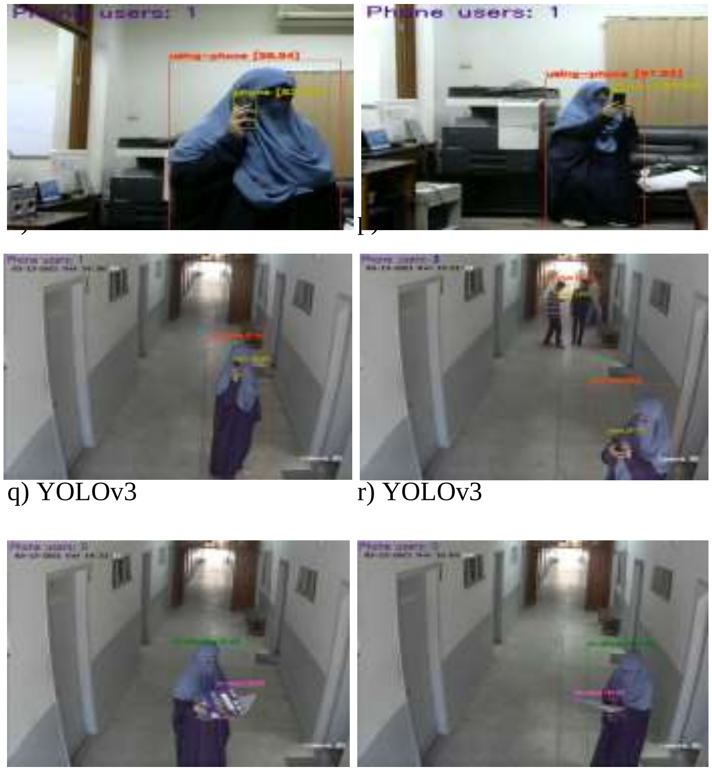


(g) YOLOv3 (h) YOLOv4

(i) YOLOv3 (j) YOLOv4

(k) YOLOv3 (l) YOLOv4

m) YOLOv3 n) YOLOv3



q) YOLOv3 r) YOLOv3

s) YOLOv3 t) YOLOv3

Figure 4: (a),(b),(c) & (d) represents the detection generated by the model on the images downloaded from the internet, (e),(f),(g) & (h) represents the detection generated by the model on webcam images, (i) & (j) represents the detection generated by the model on the video downloaded from the internet, (k) & (l) represents the detection generated by the model on the custom video (recorded from phone) (m),(n),(o) & (p) represents the detection generated by the model on the webcam and (q),(r),(s) & (t) represents the detection generated by the model on the CCTV.

Table 4: Algorithms comparison for testing results

Source	YOLOv3				YOLOv4			
	Phone	Using Phone	Not Phone	Not Using Phone	Phone	Using Phone	Not Phone	Not Using Phone
Static Images	87%	97%	-	94%	100%	81%	98%	97%
WebcamImages	78%	72%	26%	56%	99%	99%	100%	78%
Static Videos	98%	98%	-	-	99%	99%	-	-
CustomVideos	92%	98%	-	-	97%	98%	-	-
Real- Time Webcam	83%	98%	62%	62%	95%	97%	98%	96%
Real- Time CCTV	73%	97%	36%	91%	97%	97%	92%	91%

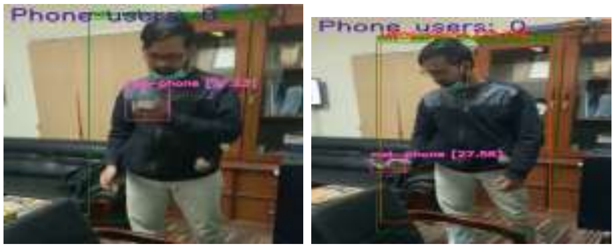
A. False Alarms

The model missed out on detections and produced some faultily. The reason highlighted was the FPS speed of the video and the continuous movement of the object and its location. Moreover, the nature of test and train data was also questioned. The missed detection and false alarms of both YOLOv3 and YOLOv4 are displayed and compared in figure 5 and table 5.

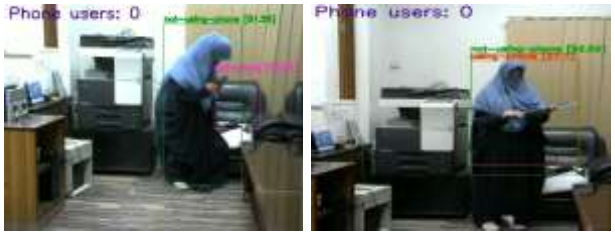
(a)YOLOv3 (b) YOLOv4

(c) YOLOv3 (d) YOLOv4

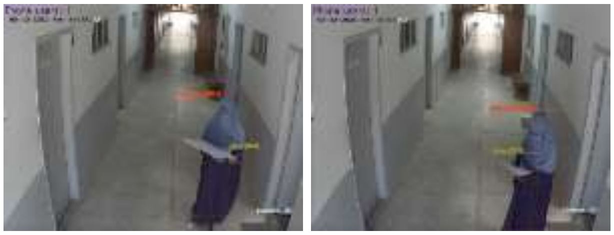
(e) YOLOv3 (f) YOLOv4



(g) YOLOv3 (h) YOLOv4



(i) YOLOv3 (j) YOLOv4



(k) YOLOv3 (l) YOLOv4

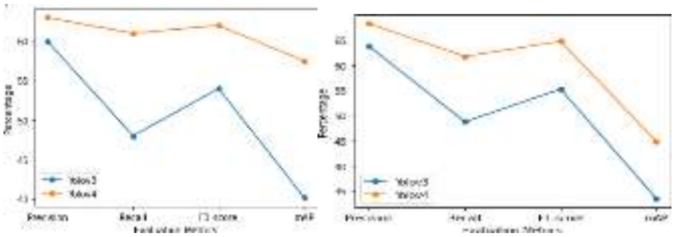
Figure 5: (a) & (b) represents the false or missed detection generated by the model on the images downloaded from the internet, (c) & (d) represents the false or missed detection generated by the model on the webcam images, (e) & (f) represents the false or missed detection generated by the model on the video downloaded from the internet, (g) & (h) represents the detection generated by the model on the custom video (recorded from phone), (i) & (j) represents the false or missed detection generated by the

model on the webcam and (k) & (l) represents the false or missed detection generated by the model on the CCTV.

Table 5: Algorithms comparison for false alarm

A. Precision, Recall, F1-Score and mAP of YOLOv3 and YOLOv4 on Test Images

20% test images were used in the holdout method whereas around 1858 images were used in 4k cross validation method for the evaluation of the trained model on YOLOv3 and YOLOv4. The results showed an improvement in the precision and recall after applying K4 cross validation method. The table 6 and figure 6(a) & (b) represents the metric evaluation comparison on test images for both YOLOv3 and YOLOv4 on holdout method and K4 cross validation method.



(a) YOLOv3 vs YOLOv4 on holdout method (b) YOLOv3 vs YOLOv4 on 4k cross validation

Figure 6: Evaluation metric of YOLOv3 and YOLOv4 for Holdout and K4 cross fold validation on test data.

Method	YOLOv3				
	Precision	Recall	F1-Score	mAP @ 50	Training Time
Hold Out	60%	48%	54%	40.21%	13 hrs
4k Cross Fold	63.75%	48.75%	55.25%	33.54%	13 hrs
Method	YOLOv4				
	Precision	Recall	F1-Score	mAP @ 50	Training Time
Hold Out	63%	61%	62%	57.5%	12 hrs
4k Cross Fold	68.25%	61.75%	64.75%	44.87%	12 hrs

Table 6: Evaluation metric of YOLOv3 and YOLOv4 for Holdout and K4 cross fold validation on test data

V. CONCLUSION

In this experiment, the study of detecting human object interactions was performed. The YOLOv3 and YOLOv4 algorithms were used for this experiment. Since the size of the dataset used was small, therefore few false alarms were noticed in the results. Experimental results showed that YOLOv4 significantly improves the performance of human-object interaction detection over YOLOv3 before and after applying 4k cross-validation. The results can be far better if the size of the data set is increased. This limitation will be tackled in the future to enhance the working of the model and make it more efficient.

.

ACKNOWLEDGMENT

This research was supported by NED-UET. For this reason, the authors would like to extend their gratitude and thanks.

REFERENCES

[1] Hueihan Jhuang et al. "Towards Understand- ing Action Recognition". In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. 2013.

[2] Joao Carreira and Andrew Zisserman. "Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017.

[3] Xiongwei Wu, Doyen Sahoo, and Steven C.H. Hoi. "Recent advances in deep learning for ob- ject detection". In: *Neurocomputing* 396 (2020), pp. 39–64. issn: 0925-2312.

[4] Suchen Wang et al. "Discovering human inter- actions with large-vocabulary objects via query and multi-scale detection". In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021, pp. 13475–13484.

[5] Victor Escorcia and Juan Niebles. "Spatio- temporal human-object interactions for action recognition in videos". In: *Proceedings of the IEEE*

Source	YOLOv3			
	Phone	Using Phone	Not Phone	Not Using Phone
Static Images	-	-	-	29%
Webcam Images	41%	-	-	-
Static Videos	-	33%	-	-
Custom Videos	-	-	27%	35%
Real-Time Webcam	-	-	55%	81%
Real-Time CCTV	35%	88%	-	-
Source	YOLOv4			
	Phone	Using Phone	Not Phone	Not Using Phone
Static Images	100%	81%	98%	97%
Webcam Images	99%	99%	100%	78%
Static Videos	99%	99%	-	-
Custom Videos	97%	98%	-	-
Real-Time Webcam	95%	97%	98%	96%
Real-Time CCTV	97%	97%	92%	91%

International Conference on Computer Vision Workshops. 2013, pp. 508–514.

[6] Xiaolong Wang, Ali Farhadi, and Abhinav Gupta. “Actions˜ transformations”. In: Pro- ceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2016, pp. 2658–2667.

[7] Muhammad Sharif et al. “A framework of hu- man detection and action recognition based on uniform segmentation and combination of Euclidean distance and joint entropy-based fea- tures selection”. In: EURASIP Journal on Im- age and Video Processing 2017.1 (2017), pp. 1– 18.

[8] Weihao Yan, Yue Gao, and Qiming Liu. “Human-object interaction recognition using multitask neural network”. In: 2019 3rd Inter- national Symposium on Autonomous Systems (ISAS). IEEE. 2019, pp. 323–328.

[9] Georgios Kapidis et al. “Egocentric hand track and object-based human action recognition”. In: 2019 IEEE SmartWorld, Ubiquitous In- telligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communi- cations, Cloud & Big Data Computing, Internet of People and Smart City Innovation (Smart-World/SCALCOM/UIC/ATC/CBDCom/IOP/SCI). IEEE. 2019, pp. 922–929.

[10] Miguel Farrajota, Jo˜ao MF Rodrigues, and JM Hans du Buf. “Human action recognition in videos with articulated pose information by deep networks”. In: Pattern Analysis and Ap- plications 22.4 (2019), pp. 1307–1318.

[11] Sadjad Asghari-Esfeden, Mario Sznaier, and Octavia Camps. “Dynamic Motion Represen- tation for Human Action Recognition”. In: Proceedings of the IEEE/CVF Winter Confer- ence on Applications of Computer Vision. 2020, pp. 557–566.

[12] Sunder Ali Khowaja and Seok-Lyong Lee. “Se- mantic image networks for human action recog- nition”. In: International Journal of Computer Vision 128.2 (2020), pp. 393–419.

[13] Eirini Mathe et al. “A deep learning approach for human action recognition using skeletal in- formation”. In: GeNeDis 2018. Springer, 2020, pp. 105–114.

[14] Muhammad Attique Khan et al. “Human ac- tion recognition using fusion of multiview and deep features: an application to video surlance”. In: Multimedia tools and applications (2020), pp. 1–27.

[15] Hao-Shu Fang et al. “DecAug: Augmenting HOI Detection via Decomposition”. In: Pro- ceedings of the AAAI Conference on Artificial Intelligence. Vol. 35. 2. 2021, pp. 1300–1308.

[16] Muhammad Umer Farooq, Najeed Ahmed Khan, and Mir Shabbar Ali. “Unsupervised Video Surveillance for Anomaly Detection of Street Traffic”. In: International Journal of Advanced Computer Science and Applications 8.12 (2017). DOI: 10 . 14569 / IJACSA . 2017 . 081234.

[17] Najeed Khan et al. “Use Hand Gesture to Write in Air Recognize with Computer Vision”. In: 17 (May 2017), p. 51.

[18] Abbasi, Muhammad Awais, and Muhammad Fahad Zia. "Novel TPPO based maximum power point method for photovoltaic system." Advances in Electrical and Computer Engineering 17, no. 3 (2017): 95-100.

[19] Inayat, Usman, Muhammad Fahad Zia, Sajid Mahmood, Haris M. Khalid, and Mohamed Benbouzid. "Learning-Based Methods for Cyber Attacks Detection in IoT Systems: A Survey on Methods, Analysis, and Future Prospects." Electronics 11, no. 9 (2022): 1502.

[20] Ali, Fahad, Ayesha Iqbal, Zunaira Nazir, Usman Inayat, Syed Mohsin Ali, and Muhammad Rehan Saleem. "A brief review on computer system control using multi-agent technique." International Journal of Sustainable Aviation 5, no. 4 (2019): 298-312.

[21] Inayat, Usman, Fahad Ali, Hafiz Muhammad Ashja Khan, Syed Moshin Ali, Kiran Ilyas, and Habiba Habib. "Wireless Sensor Networks: Security, Threats, and Solutions." In 2021 International Conference on Innovative Computing (ICIC), pp. 1-6. IEEE, 2021.

AUTHORS

First Author – Rida Ayesha, MS-CSIT, University of Management and Technology, Lahore, Pakistan (email: rida.ayesha@umt.edu.pk)

Second Author – Najeed Ahmed Khan, PhD., NED University of Engineering and Technology, Karachi, Pakistan (email: najeed@neduet.edu.pk)

Third Author – Usman Inayat, PhD., University of Management and Technology, Lahore, Pakistan (email: usman.inayat@umt.edu.pk)

Correspondence Author – Dr Usman Inayat, usman.inayat@umt.edu.pk, usmaninayat.edu@gmail.com, https://orcid.org/0000-0001-8397-9995