BALKAN JOURNAL OF ELECTRICAL & COMPUTER ENGINEERING, Vol. 11, No. 4, October 2023
Research Article

Intelligent Video Surveillance System Using Faster Regional Convolutional Neural Networks

Olayemi Mikail Olaniyi, Shefiu Olusegun Ganiyu, Efenedo Gabriel Ilori, Sunday J Akam

***Abstract*— Insecurity remains a major challenge in our society. Government, private organizations, and individuals strive to ensure their possessions are kept safe from intruders. Automated surveillance system plays a key role to ensure that the environment is safe with little human intervention. Therefore, object detection, classification, and tracking are vital in building a robust and remote intelligent video surveillance system to aid security in physical environments. Previous studies used enhanced background subtraction techniques for object detection which recorded notable achievements but performance issues in distinguishing humans, pets and vehicles. For insecurity to be solved more intelligently, deep neural network techniques are employed. In this paper, an intelligent video surveillance system that detects only human intrusion and sends an SMS notification to the user with the registered mobile number was developed. The results of the system performance evaluation recorded an accuracy of 96%, a precision of 94%, and a recall of 98%. The experimental results showed that the intelligent system was suitable for detecting human intrusion, thereby contributing to the safety of physical environments.**

Index Terms— **CNN, Deep learning, Detection, Surveillance Video.**

I. INTRODUCTION

THE NEED for day-to-day security of immediate environment cannot be over-emphasized. For a long time now, video surveillance has been an important aspect of securitysystem that plays a vital role in ensuring that lives and properties are kept safe. The early implementation of this system relied on humans for its operations [1]. Thus, surveillance systems have increased in number as the demand for this system increases over the years. With the advent of development in video surveillance systems, governments, individuals, and various organizations across society use the systems to keep track of various activities for the sole aim of security and safety [2]. In today's smart cities, video surveillance is used for inventory control in retail outlets, home monitoring (neighbourhood watch), security on corporate and educational campuses.

Video surveillance systems by human operators to detect an intrusion is an inefficient or even impractical solution because human resources are expensive and have limited capabilities [3]. Intelligent video surveillance systems are integrated systems that include electronic (sensing devices), pattern recognition, and computer vision, networking, artificial intelligence, and communication [2]. Therefore, the goal of an intelligent video surveillance system is to automatically monitor people, property, and the environment without the need for human intervention. As a result, this monitoring task entails automatically detecting and classifying objects (either humans or household pets), as well as performing additional analysis and taking actions. Especially, image processing and artificial intelligence (deep learning) techniques are important in the development of intelligent video systems [4].

With advancements in deep learning, particularly convolutional neural network (CNN), in computer vision applications, the accuracy of object detection and classification has improved dramatically for intelligent video surveillance [5]. Neural network algorithms offer state-of-the-art performance in classification and object detection widely used in intelligent video surveillance for intrusion detection in restricted environments. The specific contribution of this paper is to evolve an algorithm that supports the design and development of an intelligent surveillance system, which is based on Faster R-CNN for human and non-human detections for accurate intrusion detection while reducing false (increase precision) alarm rate as anticipated in [3].

I. RELATED WORKS

Several researchers have worked on intelligent video surveillance and object detection, classification, and tracking. An object detector using multi-detector regional convolutional neural networks (RCNN) was developed in [6]. Object detection and tracking play a very important role in surveillance for traffic control, object counting, and physical aspect of security. The system adopted a multi-detector model based on faster RCNN, a combination of CNN and Amulet is use to extract the raw feature from the image, region proposal network (RPN) is used to predict the expected region of interest (RoI). Thus, multiple detections are used to detect the image.
Similarly, An intelligent smartphone-enabled surveillance system was designed and implemented in [7]. The system uses a passive infrared (PIR) sensor and a microcontroller (MCU) attached to a smartphone through the MCU for motion detection. When a motion is detected, video is captured and the

O. M. OLANIYI is with Department of Computer Science, National Open University of Nigeria, Abuja, Nigeria, (e-mail: omoolaniyi@noun.edu.ng).
https://orcid.org/0000-0002-2294-5545
S.O. GANIYU, is with Computer Science Department, Kampala International University, Uganda, (e-mail: ganiyu.shefiu@kiu.ac.ug;).
https://orcid.org/0000-0003-4182-3890
G. I. EFENEDO, is with Department of Electrical and Electronics Engineering, Delta State University, Abraka, Nigeria, (e-mail: giefenedo@delsu.edu.ng)).
https://orcid.org/0000-0002-2831-068X
S. J. AKAM, is with Department of Computer Engineering, Federal University of Technology, Minna, Nigeria, (e-mail: mailto:sundayjames115@gmail.com)
.https://orcid.org/0000-0002-2294-5545

Manuscript received Dec 22, 2022; accepted Sep 15, 2023.
DOI: 10.17694/bajece.1223050

footage is sent to the user via short message service (SMS). The surveillance record is stored in a cloud and the link to the record is also sent to the user via email. The developed system ensures efficient use of memory by storing the record in a cloud. It is cost-effective and also offers efficient energy use as the camera is only activated when motion is detected by the PIR sensor. However, the developed system cannot efficiently differentiate radiation changes between humans, household pets, or other animals.

Also, an intelligent surveillance system for a low-cost CNN design was developed in [8] . The developed system makes use of a hardware accelerator known as Neural Compute Stick (NCS) with ROCK64 for high-speed calculation of images. A lightweight MobileNet network is used to extract the features and classify the image. Authors in [8] used the NCS to load a single shot multibox detector (SSD) network for human detection. Also, the Darknet architecture of You Only Look Once (YOLO) is used for extraction and classification of images and combine with SSD to create a bounding box for the region of interest of the detected images. A simple mail transfer protocol was used to send email to deliver the detected object.

Furthermore, an enhanced background subtraction algorithm for a smart surveillance system using adaptive gaussian mixture technique was developed in [9]. The smart system can efficiently detect motion and detect an object by means of background subtraction with illumination change. However, the developed system cannot differentiate between humans and home pets. In addition authors in [10] developed a real-time Action Detection in Video Surveillance using Sub-Action Descriptor with Multi-CNN. The system presented a novel real-world surveillance video dataset and a new approach to real-time action detection in video surveillance system. The joint space of the sub-action descriptor was not considered. Also, more powerful temporal feature methods, such as skin-color MHI or optical flow, and other deep architectures of CNNs are not considered.

Similarly, [11] developed an activity recognition using temporal optical flow convolutional features and multi multiplayer LSTM. The activity recognition framework for industrial systems proposed with a trained map CNN model help to select only the salient region that are activated for persons in the video frame which reduce verboseness and ambiguity of information in video frame. Surveillance video analysis for store-base using deep learning techniques proposed was by [12]. A skeleton recognition algorithm is adopted in place of object detection algorithm to conquer occlusion problem for gathering sufficient customer information and realizing crowd counting and density map drawing. For human tracking and counting, multiple human tracking algorithm and human re-identification (ReID) technology are adopted.

Also, [4] developed a people tracking system by Detection Using CNN features. They represented each person with 4096 Faster-RCNN feature vectors, and the Euclidean distance method was used to calculate the distance between two feature vectors of each input pair. A pair is considered the same person if their Euclidean distance is a minimum. This is due to the assumption that convolutional features of similar objects generated by Faster-RCNN should be quite similar compared to features of dissimilar objects. Furthermore, [13] suggested a General Purpose Intelligent Surveillance System for Mobile Devices using deep learning. The developed system module was divided into two: a detection and a classification module. The detection module combined background subtraction techniques, optical flow and recursively estimated density. The classification module is based on a CNN used to classify objects into one of the seven predefined categories using a pre-trained CNN.

In addition, [14] designed and developed an Edge Intelligence-Assisted Smoke Detection in Foggy Surveillance Environments. The system was developed using the architecture of CNN for detecting smoke in video streams. Pre-trained MobileNet model was trained on ImageNet dataset which focus on trying to achieve accuracy and eliminating rate of false alarm in Foggy Surveillance Environments. Also, [15] implemented an Intelligent Surveillance System Using Background Subtraction Technique for unattended or abandoned object detection. In the developed system, threshold is applied to separate red, green and blue then the use of blob-based algorithm for detecting the change in video scene, the technique detects, analyze and track object motion.

Similarly, [16] presented the use of Adaboost and CNN in crowded surveillance environment for people counting based on head detection. In this system three modules were used to achieve people counting. The module includes: Two off-line training and one online detection stage. The first off-line training, Adaboost algorithm is adopted to learn a fast-cascaded head detector with Histogram of Oriented Gradients (HOG) feature. In addition, [17] developed a Smart Surveillance as an Edge Network Service: from Harr-Cascade, Support vector Machine (SVM) to a Lightweight CNN. The system used histogram Oriented Gradient (HOG) and SVM algorithm for fast and accurate human detection. The system also used Harr cascade, Harr-like feature made up of three shapes: two rectangular features, three rectangular features and four rectangular features alongside Lightweight CNN as the classifier trained with keras dataset. Authors in [18] implemented a video structured description technology-based intelligence analysis of surveillance videos for public security applications. A pre-trained CNN architecture was adapted for tracking and re-identification of people and analysed the architecture with CUHK03 dataset. The researchers provided both manually cropped images and automatically detected bounding boxes with DPM detector, which respectively contains 13,164 images of 1360 pedestrians captured by six surveillance cameras.

Also, an efficient CNN based summarization of surveillance videos for resource-constrained devices was presented in [19] . The study investigated deep features for shot segmentation and intelligently divided the video stream into meaningful shots. The deep features were extracted from two consecutive frames to determine whether the underlying frames belong to the same or different shot. The Features were extracted from the fully connected layer (FC7) of CNN model which is trained using MobileNet architecture (version 2) on ImageNet dataset.

In addition, a Kernel ELM and CNN based Facial Age Estimation was developed in [20]. A two-level system for apparent age estimation from facial images. Then first classify

samples into overlapping age groups. Within each group, the apparent age is estimated with local repressors, whose outputs are then fused for the final estimate. We use a deformable parts model-based face detector, and features from a pre-trained deep convolutional network. Kernel extreme learning machines are used for classification.

Also, a Surveillance System Using CNN for Face Recognition with Object, Human and Face Detection was developed in [21] .The region of object they considered in an entire image was picked by object detection and discriminate whether the area is human or Otherwise and the movement of the detected object was analyzed .Similarly, [22] implemented a vegetable Category Recognition System Using Deep Neural Network. They implemented a Caffe framework based on CNN for the system classification and used Deep Neural Network (DNN) for the vegetable category recognition.

Also, [23] presented an adaptive Feature Learning CNN for Behavior Recognition in Crowd Scene. A 3D scale convolutional neural network (3DSCNN) was implemented on crowd video scene, the 3D-CNN was used in a large-scale supervised crowd dataset which optimized convolutional architectures settings. The outcomes from 3DS-CNN captured information related to objects, scenes, and actions in a video, making them useful for different applications that do not fine tune the architectural setup. In addition, [24] designed HOG-CNN Based Real Time Face Recognition to recognize faces. Histogram of Oriented Gradient (HOG) was used as the feature extractor, also for detecting all the faces in the image and the CNN was used as the training algorithm for classifying the images.

Furthermore, an Engineering Vehicles Detection Based on Modified Faster R-CNN for Power Grid Surveillance was developed in [25]. CNN methods were divided into two categories, one is the two-stage methods based on region proposal and the other is the one-stage methods based on regression. The feature extraction part of these methods was implemented by the CNN. Some methods based on region proposal such as R-CNN, Fast R-CNN and SPPnet, which adopted selective search algorithm to generate candidate boxes were utilised. Also, YOLO was used as the topmost feature map to predict confidences and bounding boxes for all categories over a fixed grid. SSD detects multiple categories by a single evaluation of the input image.

Machine Learning for Gender Detection using CNN on Raspberry Pi Platform was presented by [26]. The implementation of the system is based on the architecture of CNN and the solution permits users to extract some relevant information from the visual data containing image labelling, face and landmarks detection, optical character recognition (OCR). REST API was used to interact with Google's cloud vision platform. The real-time implementation of the hardware as well as software solution were implemented and executed on a Raspberry Pi 3 model B+ board with Pi Camera module.This paper tries to tackle such limitations presented by [27-32], [2], [10] by making use of raspberry pi and faster object detection and classification technique to improve video surveillance system

III PROPOSED DESIGN AND METHODOLOGY

This section describes in detail, the methodological steps employed in the design and implementation of the proposed intelligent surveillance system. Hence, the proposed system utilized flow design tool, flowchart, block diagram and circuit diagram for the implementation to describe the technical phase of the methodology. The system uses a trained faster regional convolutional neural network classifier for its object detection. It also makes use of Twilio API for SMS notification. With the above techniques the system was able to solve the problem of detecting home pet and humans as intruder. All these techniques were used to improve the efficiency, performance and intelligence of the proposed surveillance system.

A. Methodology

The implementation of the proposed system for intelligent video surveillance is guided by the block diagram in Figure 1. Foremost, Faster R-CNN architecture is considered because it is fast, accurate and suitable for detecting and classifying human [31] objects and home pets [4]. The Faster R-CNN architecture as shown in Figure 9 is divided into two modules: The Region Proposal Network (RPN) and a Fast R-CNN Detector. The RPN and the Fast R-CNN detector share the same convolutional layers. Faster R-CNN, by consequence, could be considered as a single and a unified network for object detection. To generate high quality object proposal, a highly descriptive feature extractor in the convolutional layers can be used. The Fast R-CNN detector uses many regions of interest (ROIs) as input. Then, the ROI pooling layer extracts a feature vector for each ROI. This feature vector will constitute the input for a classifier formed by a series of fully connected (FC) layers.

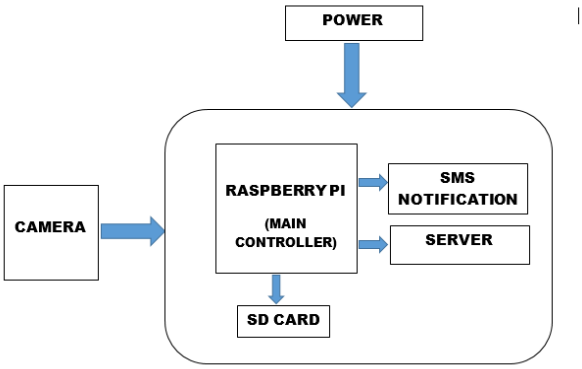

Fig. 1. Proposed system block diagram

The embedded system of the proposed system, include Raspberry Pi 3B 8MP camera module, Raspberry Pi 3B as the main controller for all the object detection and programming for the whole system, SMS notification, buzzer (Alarm notification) and power supply.

The system comprises of five basic components

- Raspberry Pi 3
- Raspberry Pi 3 8MP Camera module
- SMS notification
- Power supply
- Raspberry microSD card.

The Raspberry Pi 3B is a basic module for processing images/videos, executing object detection on acquired video frames to detect objects. The board has ARM cortex A53 clocked at 1.2GHz, 4000MHz Video Core IV multimedia GPU,

1Gb memory, power supply, HDMI, USB ports and other features.
The camera module takes in video stream then the Raspberry Pi 3B module implement the object detection on the captured frames. Figure 2 shows the flow diagram of the proposed system;

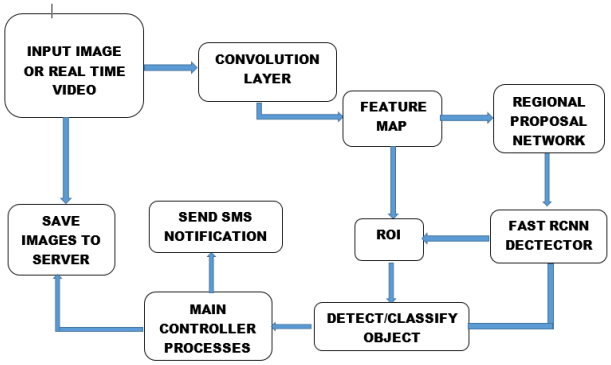

Fig. 2. Flow diagram of the proposed system

The electronic components to be used are; Raspberry pi 3B, PiCamera module SMS notification and power supply. The camera and the power supply all the required inputs to the Raspberry Pi 3B, while the SMS notification act as the output for the system. Once frames/video are acquired from the camera and fed into the Raspberry Pi 3B controller, the image is being processed using the faster regional convolutional neural network as stored within the programmed Raspberry pi 3B.

B. Mathematical Modelling of Faster Regional Convolutional Neural Network for Intelligent Video Surveillance

Regional convolutional neural network remains the first notable CNN-based object detection techniques which was found to outperformed traditional object detection algorithm with 30% improvement over the years. It relies on two separate CNN to perform its object detection. CNN computation is based on the basic computation performed by an artificial neuron which is the sum of products between weights and values of a layer. Furthermore, the computation operates in two modes - single input image and multiple input images.

1. Single input image

If (l_(x,y)) denote image or pooled feature values, depending on the layer, then the convolution value at any point (x, y) for a single network in the input is given by:

$$w * l_{x,y} = \sum_i \sum_k w_{ik} l_{x-i,y-k} \quad (1)$$

Where w is the weight, l is the feature value, i and k are the dimensions of the kernel. For a 3 x 3 weight (w) equation (1) becomes,

$$w * l_{x,y} = \sum_i \sum_k w_{ik} l_{x-i,y-k} = w_{1,1} l_{x-1,y-1} + w_{1,2} l_{x-1,y-2} + \cdots + w_{3,3} l_{x-3,y-3} \quad (2)$$

Labelling the subscript on w and l,

$$w * l_{x,y} = w_1 l_1 + w_2 l_2 + \cdots + w_9 l_9 = \sum_{i=1}^{9} w_i l_i \quad (3)$$

Since equation (2) and (3) are identical, if bias is added to the equation and the result is equated to T,

$$T = \sum_{j=1}^{9} w_j l_j + b = w * l_{x,y} + b \quad (4)$$

2. Multiple input images

The equation forward pass-through CNN,

$$T_{xy}(\vartheta) = \sum_i \sum_k w_{i,k}(\vartheta) l_{x-i,y-k}(\vartheta - 1) + b(\vartheta)$$
$$= w(\vartheta) * l_{x,y}(\vartheta - 1) + b(\vartheta) \quad (5)$$

And

$$l_{xy}(\vartheta) = h\left(l_{xy}(\vartheta)\right) \quad (6)$$

Where ϑ = 1, 2, …, Lc, and Lc is the number of convolutional layers, and denotes the values of pooled features in convolutional layer ϑ, where h is the activation function.
When ϑ=1, l_xy (0) = (values of pixels in the input image(s))
When ϑ=Lc, then:
l_xy (Lc) = (values of pooled features in the last layer of the CNN)
The Equations of Backpropagation Used to Train CNNs:

$$\delta_{x,y}(\vartheta) = h'\left(T_{x,y}(\vartheta)\right)\left[\delta_{x,y}(\vartheta + 1) * \sqrt{180\left(w(\vartheta + 1)\right)}\right] \quad (7)$$

$$\vartheta = Lc - 1, Lc - 2, \ldots, 1 \quad (8)$$

Finally, the update parameter updates the weights and bias for each feature map using,

$$w_{i,k}(\vartheta) = w_{i,k}(\vartheta) - l\delta_{i,k}(\vartheta) * \sqrt{180\left(l(\vartheta - 1)\right)} \quad (9)$$

And

$$b(\vartheta) = b(\vartheta) - l\sum_x \sum_y \partial_{x,y}(\vartheta); (\vartheta = 1,2, \ldots, Lc) \quad (10)$$

3. Faster Regional Convolutional Neural Network

Faster R-CNN belongs to the family of Regional convolutional neural network and it is becoming a replacement for fast R-CNN. The evolution within the different versions of R-CNN was usually in terms of computational efficiency (integrating the different training stages), reduction in test time, and improvement in performance (mAP). These networks usually consist of
1) A region proposal algorithm to generate “bounding boxes” or locations of possible objects in the image;
2) A feature generation stage to obtain features of these objects, usually using a CNN;
3) A classification layer to predict which class this object belongs to; and
4) A regression layer to make the coordinates of the object bounding box more precise.

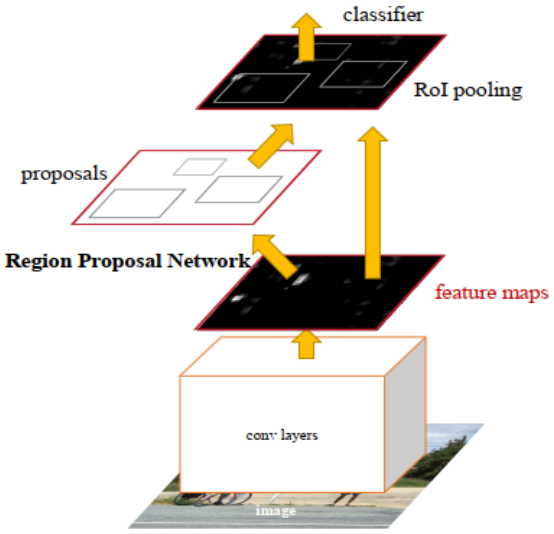

Fig. 3. Architecture of Faster R-CNN [38]

The Faster R-CNN architecture is illustrated in Figure 3. It has commonly used two-step approach to object detection. The first stage generates a set of class agnostic region of interest (RoI) or regional proposal network, where each RoI is defined by a bounding box location and an abjectness score. The second step then classifies each RoI and predicts a refined limit box position. From a general point of view, Faster R-CNN is composed of two subnetworks:

- Regional proposal Network (RPN)
- Region of interest box network (RoI)

The algorithm for simulating a multi-human object detection is presented in Figure 4.

ALGORITHM 1

1. Take input image from camera.
2. Pass the image to the ConvNe layer and returns feature maps for the image.
3. Apply Region Proposal Network on the feature maps and gets object proposals.
4. Apply RoI pooling layer to bring down all the proposals to the same size.
5. Pass these proposals to a fully connected lay.
6. Classify and predict the bounding boxes for the image.
7. Print the predicted object.

Fig. 4. Simulation on multi human object

4. *Regional Proposal Network*

The region proposal network (RPN) starts with the input image being fed into the backbone CNN. The input image is first resized such that its shortest side is 600px with the longer side not exceeding 1000px. It takes an image (of any size) as input and outputs a set of rectangular object proposals, each with an objective score. To generate region proposals, we slide a small network over the convolutional feature map output by the last shared convolutional layer. This small network takes as input an $n \times n$ spatial window of the input convolutional feature map as shown in Figure 5.

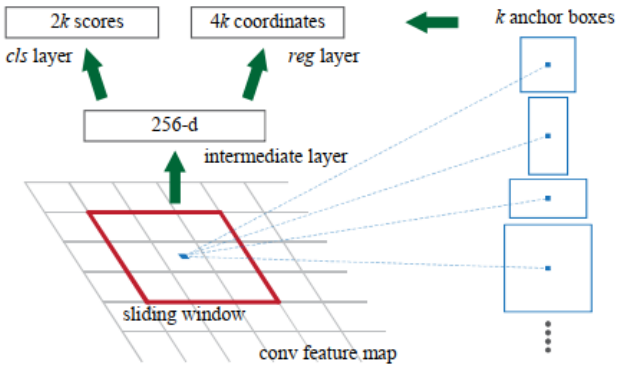

Fig. 5. Regional proposal network [39]

ALGORITHM 2

1. Initialise the input camera
   a. Get image/video frames
   b. Send the frame to raspberry pi
2. Raspberry pi loads the model to begin object detection for intelligent surveillance
3. Pass the object through faster RCNN
4. Is the object detected is a human or home pet

If home pet,
   a. No intruder found go to

If human intruder detected,
   b. Intruder detected
5. Send SMS using Twilio API
6. Save the footage to SD card
7. Send the footage to server

Fig. 6. Algorithm for the proposed video surveillance system

Each sliding window is mapped to a lower-dimensional feature (256-d for ZF and 512-d for VGG, with ReLU following). This feature is fed into two sibling fully connected layers—a box-regression layer (reg) and a box-classification layer (cls). We use n = 3 in this research, noting that the effective receptive field on the input image is large (171 and 228 pixels for ZF and VGG, respectively). The algorithm that explains the flow diagram of the overall system is shown in Figure 6.

5. *System Implementation*

The system simulation was executed on a personal computer having TensorFlow installed in a virtual environment of an Anaconda programming environment. All the codings were done using Python. The system webcam was used to capture images which were later transformed into frames for processing. Also, the same computer was used to process the frames, results of the implemented Faster R_CNN and the results from the integration of the software and hardware part of the developed system. Also, the performance and accuracy of the system using precision and recall for faster R-CNN in intelligent surveillance were conducted as part system evaluation. Thus, the overall system algorithm was implemented using python programming language, TensorFlow and OpenCV APIs to detect different object of humans, cat and dog (i.e., dog and cat as home pet). The training of the model was done using TensorFlow API and faster R-CNN algorithm.

A working prototype of the developed system is shown in Figure 7. Also, a sample screenshot of the system's user interface that showed an alert message when intruder was detected is presented in Figure 8.

Fig. 7. Prototype of intelligent surveillance system using Faster RCNN

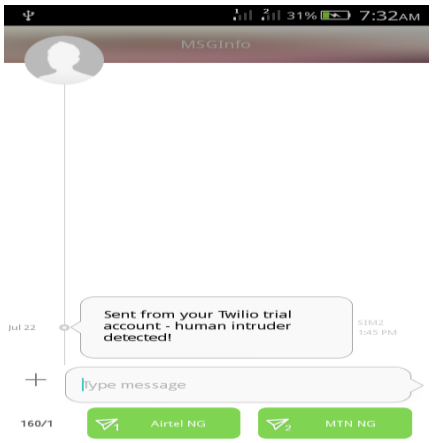

Fig. 8. User interface showing intrusion message

Fig. 9. Simulation on multi-human objects

IV. RESULTS AND DISCUSSION

The trained model was tested with images that produced outputs that contained the classes ID number, the detection confidence and the anchor boxes. This is used to identify and classify the object and determine the location of the object using *ymax*, *xmax* and *ymin*, *xmin* coordinates. Using the input frames, the model locates and recognizes the category in which the object belongs, i.e., human, cat or dog. The tasks of locating and categorizing objects is achieved with principles of detection, localization and classification. The model was also tested on multi-human objects as presented in Figure 9, as well as cat and dog as home pet as shown in Figure 10.

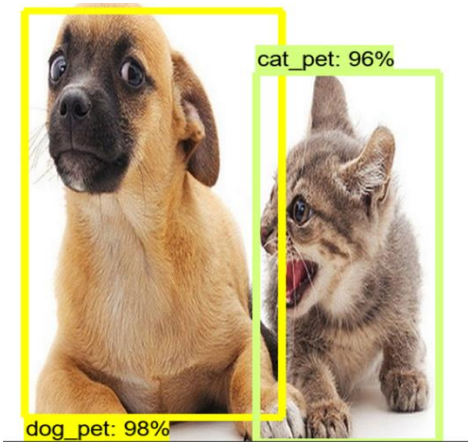

Fig. 10. Simulation on home pet

The result of the performance evaluation conducted on the implemented system is shown in Table 1. The system achieved highest precision ratio and recall ratio of 94% and 98% respectively. These results demonstrated the appropriateness of the image characteristics such as quality, type and size for study. Furthermore, the results revealed the suitability of the algorithms and APIs utilized for the system implementation.

TABLE I
DETECTION RESULT FOR INTELLIGENT VIDEO SURVEILLANCE SYSTEM

S/N	Precision (%)	Recall (%)
1	90	98
2	88	96
3	91	97
4	92	94
5	89	95
6	93	97
7	94	96
8	90	98
9	94	97
10	89	91

V. CONCLUSION

Security remains a major concern to everyone. Everybody wants to be protected from being attacked, and the means to prevent this has been a challenge over the years. A lot of solution has already been put in place to tackle insecurity. Thus, this study provided an additional approach to already existing motion and object detection techniques. The was implemented with Faster-RCNN and it intelligently detected people, dogs, and cats in a bid to raise intrusion alerts via SMS when human intrusion is detected within an environment. Faster R-CNN was adopted as an object detection model because it utilizes a regional proposal network (RPN) for faster and more accurate detection. Also, the intelligent surveillance system is portable and requires minimum expertise to operate. Overall, it provides another opportunity to enhance physical security at home and place of work against human intruders with a good performance.

REFERENCES

[1] Y. Kurylyak, "A Real-Time Motion Detection for Video Surveillance System,". IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, Rende, Italy, 2009, pp. 386-389, doi: 10.1109/IDAACS.2009.5342954.

[2] S. W. Ibrahim, "A comprehensive review on intelligent surveillance systems", CST, vol. 1, no. 1, pp 7-14, May 2016

[3] O.M. Olaniyi, S. Ganiyu and S. J. Akam. Intelligent Video Surveillance Systems: A Survey. Balkan Journal of Electrical and Computer Engineering (BAJECE).1(1).pp 57-53

[4] A. A. Shafie, F. Hafizhelmi, and K. Zaman, "Smart Video Surveillance System for Vehicle Detection and Traffic Flow Control". Journal of Engineering Science & Technology (JESTEC). Vol.13 no 7. 2195-2210

[5] B. Benjdira, T. Khursheed, A. Koubaa, A. Ammar, and K. Ouni, "Car Detection using Unmanned Aerial Vehicles: Comparison between Faster R-CNN and YOLOv3," 2019 1st Int. Conf. Unmanned Veh. Syst., pp. 1–6, 2019

[6] W. Tan, "Object Detection with Multi-RCNN Detectors," pp. 193–197.

[7] A. H. Sanoob, J. Roselin, and P. Latha, "Smartphone Enabled Intelligent Surveillance System," no. c, pp. 1–7, 2015, doi: 10.1109/JSEN.2015.2501407.

[8] L. W. Yang and C. Y. Su, "Low-cost CNN Design for Intelligent Surveillance System," 2018 Int. Conf. Syst. Sci. Eng., pp. 1–4, doi: 10.1109/ICSSE.2018.8520133.

[9] . M. Olaniyi, J. A. Bala, S. O. Ganiyu, and P. E. Wisdom, "A Systematic Review of Background Subtraction Algorithms for Smart Surveillance System," vol. 8, no. 1, pp. 35–54, 2020

[10] C. Jin, S. Li, and H. Kim, "Real-Time Action Detection in Video Surveillance using Sub-Action Descriptor with Multi-CNN," pp. 1–29.

[11] A. Ullah, K. Muhammad, J. Del Ser, S. W. Baik, and V. Albuquerque, "Activity Recognition using Temporal Optical Flow Convolutional Features and Multi-Layer LSTM," IEEE Trans. Ind. Electron., vol. PP, no. c, p. 1, 2018, doi: 10.1109/TIE.2018.2881943.

[12] H. Kaya, H. Dibeklio, and A. A. Salah, "Kernel ELM and CNN based Facial Age Estimation," pp. 80–86.

[13] A. Antoniou, "A General Purpose Intelligent Surveillance System For Mobile Devices using Deep Learning," pp. 2879–2886, 2016

[14] . Muhammad, S. Khan, S. Member, and V. Palade, "Edge Intelligence-Assisted Smoke Detection in," IEEE Trans. Ind. Informatics, vol. PP, no. c, p. 1, 2019, doi: 10.1109/TII.2019.2915592

[15] . Hargude and M. T. It, "i-surveillance: Intelligent Surveillance System Using Background Subtraction Technique," vol. 1.

[16] C. Gao, P. Li, Y. Zhang, J. Liu, and L. Wang, "Author ' s Accepted Manuscript People counting based on head detection combining environment Reference : To appear in : Neurocomputing," Neurocomputing, 2016, doi: 10.1016/j.neucom.2016.01.097.

[17] . Y. Nikouei, Y. Chen, S. Song, R. Xu, B. Y. Choi, and T. Faughnan, "Smart surveillance as an edge network service: From harr-cascade, SVM to a Lightweight CNN," Proc. - 4th IEEE Int. Conf. Collab. Internet Comput. CIC 2018, pp. 256–265, 2018, doi: 10.1109/CIC.2018.00042.

[18] Z. Xu, C. Hu, and L. Mei, "Video structured description technology based intelligence analysis of surveillance videos for public security applications," 2015, doi: 10.1007/s11042-015-3112-5.

[19] T. Hussain, K. Muhammad, A. Ullah, Z. Cao, S. W. Baik, and V. H. C. De Albuquerque, "Cloud-assisted multiview video summarization using CNN and bidirectional LSTM," IEEE Trans. Ind. Informatics, vol. 16, no. 1, pp. 77–86, 2020, doi: 10.1109/TII.2019.2929228.

[20] H. Kaya, H. Dibeklio, and A. A. Salah, "Kernel ELM and CNN based Facial Age Estimation," pp. 80–86.

[21] Y. Byeon and S. Pan, "A Surveillance System Using CNN for Face Recognition with Object, Human and Face Detection," pp. 975–984, doi: 10.1007/978-981-10-0557-2.

[22] Nogay, H.S. T.C. Akinci, and M. Yilmaz. "Detection of invisible cracks in ceramic materials using by pre-trained deep convolutional neural network." Neural Computing and Applications 34.2 (2022): 1423-1432.

[23] A. N. Shuaibu, A. S. Malik, and I. Faye, "Adaptive Feature Learning CNN for Behavior Recognition in Crowd Scene," pp. 357–361, 2017

[24] H. Ahamed, I. Alam, and M. Islam, "HOG-CNNBasedRealTimeFaceRecognition," 2018 Int. Conf. Adv. Electr. Electron. Eng., pp. 1–4, 2018.

[25] X. Xiang, N. Lv, X. Guo, S. Wang, and A. El Saddik, "Engineering vehicles detection based on modified faster R-CNN for power grid surveillance," Sensors (Switzerland), vol. 18, no. 7, 2018, doi: 10.3390/s18072258.

[26] M. H. Gauswami, "Implementation of Machine Learning for Gender Detection using CNN on Raspberry Pi Platform," 2018 2nd Int. Conf. Inven. Syst. Control, no. Icisc, pp. 608–613, 2018.

[27] D. Chahyati, M. I. Fanany, and A. M. Arymurthy, "Tracking People by Detection Using CNN Features," Procedia Comput. Sci., vol. 124, pp. 167–172, 2018, doi: 10.1016/j.procs.2017.12.143.

[28] D. Chahyati, M. I. Fanany, and A. M. Arymurthy, "Tracking People by Detection Using CNN Features," Procedia Comput. Sci., vol. 124, pp. 167–172, 2018, doi: 10.1016/j.procs.2017.12.143.

[29] H. C. Shin and J. Y. Lee, "Pedestrian Video Data Abstraction and Classification for Surveillance System," 9th Int. Conf. Inf. Commun. Technol. Converg. ICT Converg. Powered by Smart Intell. ICTC 2018, pp. 1476–1478, 2018, doi: 10.1109/ICTC.2018.8539426.

[30] A. Ullah, K. Muhammad, J. Del Ser, S. W. Baik, and V. Albuquerque, "Activity Recognition using Temporal Optical Flow Convolutional Features and Multi-Layer LSTM," IEEE Trans. Ind. Electron., vol. PP, no. c, p. 1, 2018, doi: 10.1109/TIE.2018.2881943.

[31] L. Du, R. Zhang, and X. Wang, "Overview of two-stage detection algorithms," 2020, doi:10.1088/1742-6596/1544/1/012033
R. Arti, "Animal Detection Using Deep Learning Algorithm," vol. 7, no. 1, pp. 434–439, 2020

[32] O., Türk, A. Çalışkan, Acar, E. and B. Ergen. "Palmprint recognition system based on deep region of interest features with the aid of hybrid approach. SIViP " 17, 3837–3845. 2023.

BIOGRAPHIES

Olayemi Mikail Olaniyi is a Professor in the Department of Computer Science at National Open University of Nigeria, Abuja, Nigeria. He obtained his B.Tech. and M.Sc. in Computer Engineering and Electronic and Computer Engineering respectively. He had his Ph.D. in Computer Science and Engineering (Computer Security) from Ladoke Akintola University of Technology, Ogbomosho, Oyo State, Nigeria. He has published in reputable journals and learned conferences. His areas of research include Computer Security, Intelligent/Embedded Systems design and Applied Medical Informatics.

Shefiu Olusegun Ganiyu is a Senior Lecturer in the Department of Computer Science, Kampala International University, Uganda. He holds a Ph.D. in Computer Science, which focused on risk-aware access control for pervasive environments. Similarly, he obtained Bachelor degree in Mathematics/Computer Science and Master degree in Information Science from Federal University Minna and the University of Ibadan respectively. Prior to joining the academic environment, he acquired valuable work experience as a programmer/information system developer. His research interests include information security risk management, dynamic access control, user behaviour analytics, security of pervasive computing including Bring Your Own Device (BYOD) strategy, and physical security. Also, he has participated in several projects involving information systems security and development**.**

Efenedo Gabriel Ilori is a Senior Lecturer in the Department of Electrical and Electronics Engineering, Delta State University, Abraka, Oleh Campus, Delta State, Nigeria. He obtained his BEng

degree in Electrical and Electronics Engineering as well MEng and PhD in Electronics and Telecommunication Engineering from the University of Benin, Benin-City, Edo State, Nigeria. He has publications in reputable journals and conferences. His area of research includes: Intelligent, cellular communication and information system.

Sunday J Akam obtained his Bachelor's Degree in Computer Engineering from the Federal University of Technology, Minna, Niger State, Nigeria. He is a promising computer vision and embedded systems developer. He has keen passion for AI, machine vision and developing advanced vision models.