IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

SMART ALARMING SYSTEM USING OBJECT DETECTION

AUTHOR NAMES (**BABU AMAN RAI SAXENA, MADHURI NIKHIL, GODHA VARUN DEV, UPPALA SREE CHANDER**) Department (Information Technology) University (Jawaharlal Nehru Technological University)

Address (Ashok Nagar, Kukatpally, Hyderabad, Telangana 500085)

COUNTRY (India)

Abstract: - Security is important to live life peacefully but, it's very costly in a country like India, adopting a CCTV camera may not be helpful as it may only identify faces of criminals, and will not help in stopping a criminal activity, to tackle this we need a person to monitor the CCTV camera, to eliminate the cost of employing a person, we are introducing a security structure that can perceive things using object detection and alarm the user, with less graphics and memory requirements. Object Detection is the method of detection and shaping real-world objects. There are many detection methods in existence but the accuracy, rapidity, and efficiency of detection are not good enough. The Objective is to detect of objects using You Only Look Once (YOLO) approach. Proposed method has several advantages while comparing to other object detection algorithms. In other algorithms like Convolutional Neural Network, Fast Convolutional Neural Network the algorithm will not guess at the image completely but in YOLO the algorithm looks at the image by predicting all the bounding boxes using convolutional network and class probabilities for these boxes and detects the image faster as compared to other conventional algorithms. Although newer algorithms are more accurate, they require high graphics which will make this detection process expensive. This algorithm performs efficient object detection while not compromising on the performance. Using this algorithm, we are implementing an invasion finding system so that if an object gets detected at a fully secured area the user and the security personnel can be informed, we also have an alarm system that is used to alert the user with the voice (name of the object detected). We have used google text to speech conversion API to make the software speak the name of the object detected. In the previous works object detection was represented in the field of security nevertheless did not use its capacity to detect many objects (Knives, guns, etc.), instead they focused only on harmful object. Apart from this we used text to speech conversion technique to alert the user which has not been used commonly before in object detection.

Key-Words: **-**Convolutional Neural Network, Fast-Convolutional Neural Network, Bounding Boxes, YOLO, Google text to speech

1 Introduction

In computer vision, the neural convolution networks are contrasted in the classification of images. In this paper a secure surveillance method is adopted using YOLO algorithm, which is used in detection and tracking of objects in python-based environment. Object detection involves detecting region of interest of object from a given class of image. Different methods are –Frame differencing, Optical flow, Background subtraction and object alarming [9]. It is a method used for detection and locates an object which is in motion with the help of a camera and alarming the user of intrusion. Detection algorithms and tracking algorithms are described by extracting the features of image and video for security applications. Object detection is a domain that has been profited from the recent advances in deep learning methodology. In Recent years people developed many algorithms for object detection, some of which include YOLO, SSD, Mask RCNN and Retina Net. SSD is a object detection algorithm that was created in Google Inc. This method is grounded on the VGG-16 architecture [1]. YOLO is simple and easier to tool. set of default boxes is builtto pass over numerous feature maps in a convolutional way. If an object detected is one among the object classifiers during prediction, then a score is produced. The object shape is adjusted to match the localization box. For each box the shape offsets, confidence level is being predicted. During training, default boxes are

tested to the ground truth boxes. The model loss is gauged as a weighted sum of confidence loss and localization loss. Measure of the variance of the predicted box from the ground truth box is localization loss. Confidence is measure of in which manner confidence the system is that a foreseen object is the actual object. Security is a big worry in today's world. Hiring someone to protect ourselves while we are busy is costly and many people cannot be able to afford it, so to protect people there is police. By using object detection, we alarm the user of any intrusion in their home, this can be done by installing a camera at their home, and running this project on that live video, if there is, any activity of intrusion the users can inform police and be safe.

2 Related Work:

In this section we first outline research related to addressing object detection for the use of security. We have referred from the following sources, You Only Look Once: Unified, ,Real-Time Object Detection, presented by Joseph Redmon [12]. Their previous work was on detecting objects using a regression algorithm. To achieve high accuracy and confident predictions they have proposed YOLO algorithm is based approach in the paper. Understanding of the Object Detection Based on CNN Family and YOLO, by Juan Du [15]. Is the paper that we have referred from, they generally explained about the object detection families like CNN, R-CNN and equated their efficacy and introduced YOLO algorithm to increase the efficiency [2,9]. Erudition to Localize Objects with Structured Output Regression. This paper is about Object Localization. In this paper, they used the Bounding box methods for localization of objects to overcome drawbacks of the sliding window method [4].

The above matter was about the algorithm in general but considering the entire methodology here we are using a approach that could detect various objects like knives, guns and even people when specified, this system is not dependent on a single harmful object they comprise of many, In the previous papers that worked on object detection in the field of security they didn't include many elements[7] but, in here we are including many objects that are harmful, in addition to this we are providing an alarm that might remain used as a boon to alert the user.

3 Problem Formulation:

For the need of security, we are using a webcam in the paper but for the need of quality performance we prefer the program to be attached to the CCTV camera, The YOLO algorithm runs through taking inputs from camera it takes the video and converts it into the frames[9].while doing this the process of evaluating the images is done an in this process images are identified and the objects are recolonized, we are taking the names of the objects detected and verifying, if any object is harmful and any person is detected which can be suspicious then the class name of the object detection is sent to the Google Text to Speech and this API raises an alarm with the name of the object detected. This will continue for a while and this will make definite that the user is known about any intrusion activity.

3.1 Existing System

In Existing system motion sensors and thermal sensors were used or simply CCTV cameras were placed to monitor the environment at a place such as banks, borders and safe houses, the person would monitor and notify or inform the police about any possible intrusion or threat, this would be a costly affair as there are many cameras at the borders and security places, it requires many people to monitor through CCTV cameras.

The cost is the important factor to be a hindrance for the implementation, although we can get accurate results using these models, we can also achieve similar accuracy while depending on object detection and alarming methodology. Apart from cost this system can be attached to the existing systems with minimum requirements thereby, giving an advantage while selecting this model,

4 Problem Solution:

In this paper, we are present a surveillance system that can be beneficial for detecting, tracking and alarming the user of any obscure activity, the main problem that we face in the modern observation of surveillance is that it uses conventional technology and relies on a person checking or monitoring a pace through CCTV, using modern technology people may also rely on the motion sensors and other tracking devices, although it may be feasible but using these technologies might increase the charge of our security at a place.

4.1 Proposed System:

YOLO algorithm is used as YOLO is trained on PASCAL VOC dataset, it is capable of detecting objects which can be seen in everyday life [9,10]. and will alarm the user if the objects detected are harmful. This will diminish the work done by the normal person to monitor the camera constantly.
To implement this algorithm, we need four modules.

5 Implementation:

This involves four step process, although the first three are common for YOLO we have an additional step which involves object alarming

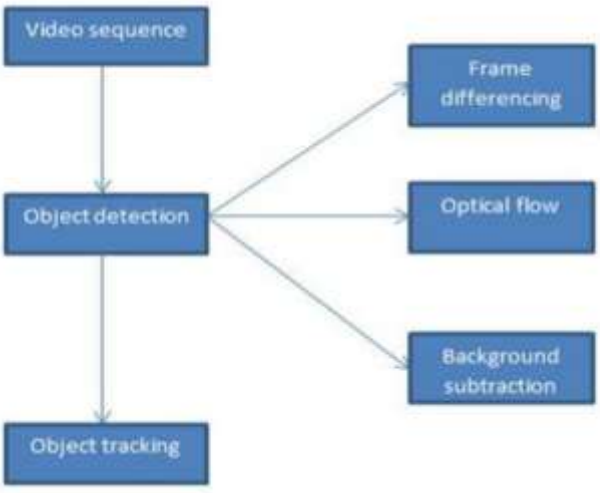


Figure 1:Methods used for Object tracking algorithm

5.1 Frame Differencing:

Frames are taken from camera at consistent intervals of the period. Variance is projected from consecutive frames. Optical Flow is a technique that guesses and computes the visual flow field with algorithm used for ocular stream. local mean algorithm is used to enhance this. To filter noise algorithm takes place. It comprises a wide revision to figure and size of the objects and supportive in avoiding time consuming and complex preprocessing approaches.

5.2 Backgroud Subtraction:

Background subtraction (BS) technique is a hasty method of confining objects in motion from a video captured by immobile camera. It forms the chief step of a multi-stage system. It is a type of process that which separates background from the foreground object in sequence of images.

5.3 Object Tracking:

Object tracking which is done in video sequences like security cameras and CCTV surveillance system,the objective is to track the path followed, speed of an object [7]. The rate of real time detection can be augmented by engaging object tracking and running classification [3,6] some frames are captured in a static interval of time. Object detection can work on slow frame rate observing for objects to catch onto and once those objects are noticed and locked, then object tracking, can progress in faster frame speed [4].

5.4 Object Alarming:

After the detection of object, the detected frame name is fetched from the dataset and is sent to Google Text to Speech API commonly known as the gTTS API, after sending object name the voice of the object detected will be spelled out loudly so the user can know of any intruder in his place. gTTS is simple to use as means which is been used to convert the text on the bounding box to speech and raises an alarm. The gTTS API supports many languages including English, Hindi, Tamil, French, German and many more. The speech can be delivered in any one of the two available audio speeds, fast or slow.

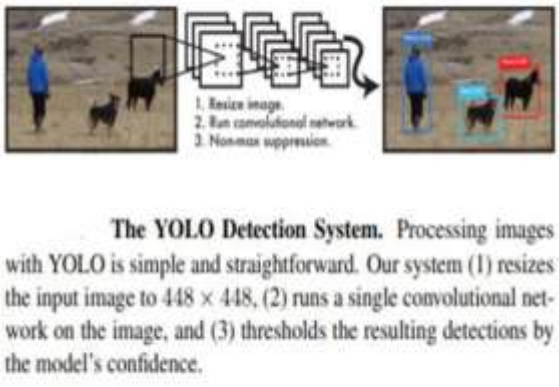
5.4.1 Convert Text to Speech in Python:

Several APIs are available to translate text to speech in python. One of such API's is the Google Text to Speech API commonly known as the gTTS API. gTTS is a easy tool which changes the text pass into audio which will be saved as mp3 files. The gTTS API supports numerous languages as well as English, Hindi, Tamil, French, German and many more. The speech can be delivered in any one of the two available audio speeds, fast or slow. Though, as of the latest update, it will not be possible to change the vocal sound of the generated audio.

6 Design Methodology:

YOLO Algorithm:

There are number of object detection algorithms such as RCNN, SSD, Faster-RCNN, but YOLO uses a different approach. YOLO is a shrewd convolutional neural network (CNN) for object detection in real-time. The algorithm implies a only neural network to the full image, then divides the image in regions and will predict the bounding boxes, probabilities for every region [9,1,2]. These leaping boxes are weighted by the probabilities. YOLO is good because it achieves high in accuracy while also being able to run in real-time. The algorithm "only looks once" at the image and requires only one forward spread through neural network to make predictions [8]. After completion of non-max suppression, then the outputs are recognized as objects together along with bounding boxes. By using a YOLO algorithm, a single CNN can simultaneously predicts several bounding boxes and class probabilities for that boxes. YOLO trains on full images and optimizes detection performances [9].



The YOLO Detection System. Processing images with YOLO is simple and straightforward. Our system (1) resizes the input image to 448 × 448, (2) runs a single convolutional network on the image, and (3) thresholds the resulting detections by the model's confidence.

Figure 2: YOLO detection system method of resizing the images.

Unified detection

We unite the distinct components of object detection into solo neural network. This network uses features from entire image to predict every box. It also predicts bounding boxes in all classes for the image paralelly. it means our network reasons universally about complete image and the objects in the image. The design of YOLO can able to complete end-to-end training and real time speeds with maintaining high precision [9]. This system divides input image into an S × S grid. If middle of the object falls into grid's cell, that particular cell is accountable for detecting that specific object. All grid cell predicts B boxes and confidence scores for boxes.. Properly we define confidence as Pr(Object) ∗ IOU truth prediction .If not any object are present in that grid cell, the confidence scores will be zero or the confidence score to match the juncture over union (IOU) between the projected box including the ground truth. Every bounding box will have 5 predictions: x, y, w, h, confidence. The width and height will be predicted relative to the full image. At last, the confidence calculation signifies the IOU among the projected box and ground truth box. Every grid cell will predict C conditional class probabilities, Pr(Classification |Object).

These probabilities are placed on the grid cell containing an object. We will predict class probabilities per each grid cell, nevertheless of the no.of boxes B. At the test time we increase the conditional class probabilities and individual box assurance predictions, which will give us class-particular confidence scores for every individual box.

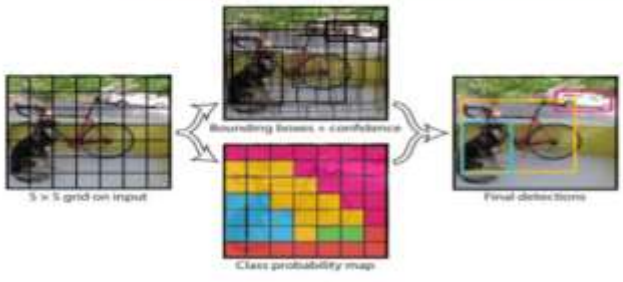
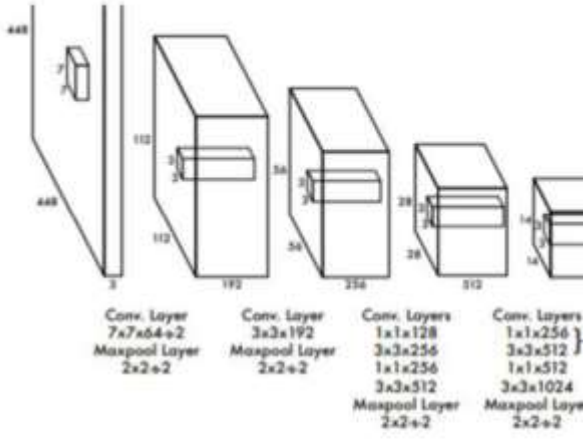


Figure 3: Bounding boxes and class probability map representation

For assessing YOLO on dataset PASCAL VOC, we use the S = 7, B = 2. PASCAL VOC that has 20 labelled classes and C = 20. final estimate is a 7 × 7 × 30 tensor.

6.1 **Algorithm Architecture**

We have implemented this model as convolutional neural network and evaluated it on PASCAL VOC detection dataset. The previous convolutional layers of the network extract feature, from image while the fully connected layers predict output probabilities, coordinates [11,14] Our network architecture is inspired by the Google Net model. Our network has 24 convolutional layers and 2 fully connected layers. we use 1 × 1 reduction layers trailed by 3 × 3 convolutional layers. The full network details is shown in Fig. We also train rapid version of YOLO designed to advance the boundaries of object detection.



The Architecture. Our detection network has 24 convolution convolutional layers reduce the features space from preceding layers. W task at half the resolution (224 × 224 input image) and then double the r

Figure 4: YOLO Architecture Representation

6.2 **System Architecture**:

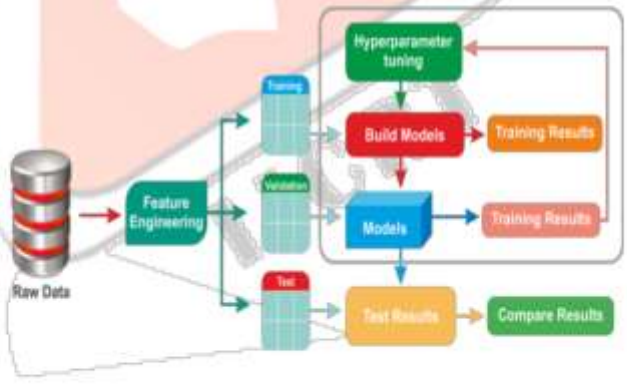


Figure 4: System architecture of the YOLO.

The above system architecture shows the actual skeleton of the project, the model is trained on the dataset and then the required data is given to the trained model for further detection of the object based on the names in the dataset.

The Raw data represented here is the video being formulated by the algorithm using a camera, this enters the algorithm frame by frame, next part is the feature engineering, this will divide data into three parts, Testing, validation and training, the frames. Models are generated and in the validation section checked for any outliers. The data then again trained and sent to testing if the data is feasible, this process will continue on a loop. The tested data is compared with trained data to give t accuracy of the model. This accuracy will be helpful for determining the object validity, if object gets detected. The

6.3 Harmful Object Detection

Detection of harmful object is done through training the images on PASCAL VOC dataset [10,15]. In this process the algorithm detects from the live video and the harmful objects get detected and this sends alarm notice to google text to speech thereby, making the user aware of the activity. The figure below shows the person with a harmful weapon, this type of image when detected by video will inform the user with a alarm containing name of both gun and person, so that user gets to know a person with a gun may try to enter his house.

Figure 5:Representaion of the Weapon.

7 Results:

The results of this paper are recorded using a webcam, making sure that algorithm and the entire implementation works on the low-quality camera as well.

Fig 6: Detection of the object.

The sample figures are shown above as the result of starting the program, here we see the yolo algorithm is been run and is using the data of video through webcam of a laptop, here the algorithm is detecting the objects.

Fig 7: Detection of the object person and remote.

Here we are demonstrating algorithms ability to detect multiple objects at a same time (remote and person are detected), and accuracy is also specified.

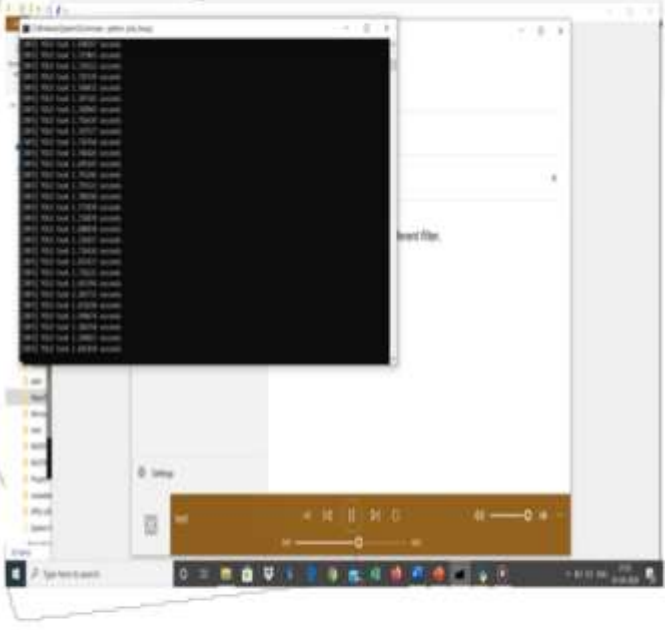


Figure 8: time to detect objects (value in command prompt)

Here we can see the background information in the command prompt, the time of the taken to detect the object is specified and will be visible in the prompt.

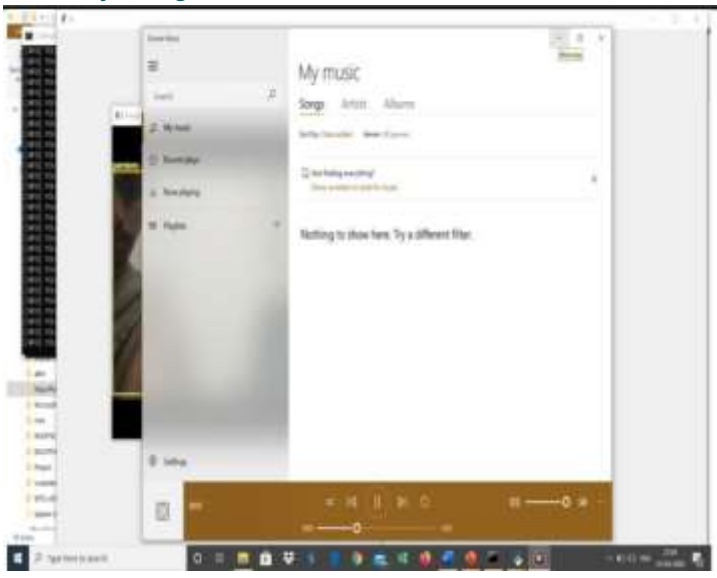


Figure 9: usage of goggle text to speech to alarm the user.

As specified earlier that if the object is harmful, then we are given an alarm, this alarm comprises on the name of object detected and is given through the laptop in built voice system.

8 Conclusion:

In the present paper we have successfully presented a security system using YOLO algorithm as YOLO is fast and accurate object detector algorithm[9]. This model is simple to build and be trained on full images. Unlike the classifier-based approaches, YOLO is trained on loss function that will directly correspond to the detection performance or efficiency and the entire model is trained. This has been only place where the object detection is combined with text to speech methodology to present a different outcome. In this project we have successfully used YOLO algorithm of object detection to help security users identify intruders using an alarm(name of the object).This will help people from theft and robbery by helping them to know about any intruder activity happening at their home so that they can inform security personnel, this can be used in military also as it can detect objects and alarm the army of any intruders at the borders, another use of this application can be made on bank's CCTV cameras (as this can detect the harmful objects(knives..etc)) and the police can be informed.

References:

[1] Wei Liu and Alexander C. Berg, "SSD: Single Shot Multi Box Detector", Google Inc., Dec 2016.

[2] Ishaq Azhar Mohammed, "RISK-BASED ACCESS CONTROL MODEL: A SYSTEMATIC LITERATURE REVIEW", International Journal of Creative Research Thoughts (IJCRT), ISSN:2320-2882, Volume.7, Issue 2, pp.794-797, May 2019, Available at :http://www.ijcrt.org/papers/IJCRT1134133.pdf

[3] Andrew G. Howard, and Hartwig Adam, "Mobile Nets: Efficient Convolutional Neural Networks for Mobile Vision Applications", Google Inc., 17 Apr 2017.

[4] Justin Lai, Sydney Maples, "Ammunition Detection: Developing a Real Time Gun Detection Classifier", Stanford University, Feb 2017.

[5] Shreyamsh Kamate, "UAV: Application of Object Detection and Tracking Techniques for Unmanned Aerial Vehicles", Texas A&M University, 2015.

[6] Adrian Rosebrock, "Object detection with deep learning and OpenCV", pyimagesearch.

[7] Mohana and H. V. R. Aradhya, "Elegant and efficient algorithms for real time object detection, counting and classification for video surveillance applications from single fixedcamera,"2016International 00Conference on Circuits, Controls, Communications and Computing (I4C), Bangalore, 2016, pp. 1-7.

[8] Ishaq Azhar Mohammed. (2020). ARTIFICIAL INTELLIGENCE FOR CYBERSECURITY: A SYSTEMATIC MAPPING OF LITERATURE. International Journal of Innovations in Engineering Research and Technology, 7(09), 172–176. Retrieved from https://repo.ijiert.org/index.php/ijiert/article/view/2797

[9] Apoorva Raghunandan, Mohana, Pakala Raghav and H. V. Ravish Aradhya, "Object Detection Algorithms for video surveillance applications" International conference on communication and signal processing (ICCSP), 2018, pp. 0570-0575.

[10] Manjunath Jogin, Mohana, "Feature extraction using Convolution Neural Networks (CNN) and Deep Learning" 2018 IEEE International Conference On the Recent Trends In Electronics Information Communication Technology,(RTEICT) 2018, India.

[11] Arka Prava Jana, Abhiraj Biswas and Mohana, "YOLO based Detection and Classification of Objects in video records" 2018 IEEE International Conference On the Recent Trends In Electroniks Information and Communication Technology,(RTEICT) 2018, India.

[12] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, andA. Zisserman." The pascal visual object classes (voc) challenge.International journal of computer vision", 88(2):303–

[13] Ishaq Azhar Mohammed, "SECURITY, PRIVACY AND RISKS WITHIN SMART CITIES: LITERATURE REVIEW AND DEVELOPMENT OF A SMART CITY INTERACTION FRAMEWORK", International Journal of Creative Research Thoughts (IJCRT), ISSN:2320-2882, Volume.8, Issue 1, pp.55-59, January 2020, Available at :http://www.ijcrt.org/papers/IJCRT2001009.pdf

[14] C.-Y. Fu, W. Liu, A. Ranga, A. Tyagi,andA.C.Berg.Dssd: "Deconvolutional single shot detector". arXiv preprintarXiv:1701.06659, 2017.

[15] D. Gordon, A. Kembhavi, M. Rastegari, J. Redmon, D. Fox,and A. Farhadi. Iqa: "Visual question answering in interactiveenvironments." arXiv preprint arXiv:1712.03316, 2017.

[16] K. He, X. Zhang, S. Ren, and J. Sun." Deep residual learninfor image recognition". In Proceedings of the IEEE conferenceon computer vision and pattern recognition, pages 770–778, 2016.

[17] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara,A. Fathi, I. Fischer, Z.Wojna, Y. Song, S. Guadarrama, et al."Speed/accuracy trade-offs for modern convolutional object detectors."

[18] Ishaq Azhar Mohammed. (2020). USABILITY AND PRIVACY IN ACADEMIC LIBRARIES: REGAINING A FOOTHOLD THROUGH IDENTITY AND ACCESS MANAGEMENT. International Journal of Innovations in Engineering Research and Technology, 7(3), 43–48. Retrieved from https://repo.ijiert.org/index.php/ijiert/article/view/2782

[19] I. Krasin, T. Duerig, N. Alldrin, V. Ferrari, S. Abu-El-Haija,A. Kuznetsova, H. Rom, J. Uijlings, S. Popov, A. Veit,S. Belongie, V. Gomes, A. Gupta, C. Sun, G. Chechik,D. Cai, Z. Feng, D. Narayanan, and

[20] K. Murphy." Openimages:A public dataset for large-scale multi- label andmulti-class image classification." Dataset fromhttps://github.com/openimages, 2017.