# DETECTION AND CLASSIFICATION OF MOTION SENSOR IMAGES USING MACHINE LEARNING

**Mr.Kevin Joy D'Souza[1],Mr. Deepak[2], Mr. Sai Prakash[3],Mr. Thejesh S Poojary[4]**

[1]Assistant Professor,Dept. of CSE,Yenepoya Institute of Technology,Moodbidri,India-574227

Students, Dept. of CSE,Yenepoya Institute of Technology, Moodbidri, India-574225

---

***Abstract***— Image detection and classification is a fundamental unit in the field of Image processing. A novel approach for the detection and classification of human movements in video scenes is presented in this application. It consists of detecting, segmenting and tracking foreground objects in video scenes to further classify their movements as conventional or non-conventional. From each tracked object in the scene, features such as position, speed, changes in direction and temporal consistency of the bounding box dimension are extracted. These features make up feature vectors that are stored together with labels that categorize the movement and which are assigned by human supervisors.

***Keywords—Image Detection and classification; YOLO algorithm; conventional or non-conventional;)***

## I. INTRODUCTION

The segmentation of events into video scenes is a new area of research in computer science and has been steadily growing as a result of broader real-world applications. The first major reason is the rising interest and the use of video based security systems, known as CCTV. However, most of the CCTV systems currently available in the market have limited functionality that involves capturing, storing and monitoring the video collected from one or more cameras. Some CCTV systems already incorporate motion detection algorithms and are able to compress video recordings only when differences are detected at the scene of the scene. The main use of such systems is the recording of regular and non-standard events for consultation and analysis. Acquisition and Classification of Human Motions in Video Actions In other words, such systems do not have a fixed intelligence capable of rendering object events. There are no ways to alert the operator when unusual events occur. Such an attribute can be very helpful in preventing and effectively detecting non-conventional events.

Human use gait, represented by a histogram, was used to classify non-dividing conditions within the house. Classification was performed using a standard histogram. Besides this method is based on the characteristics of the object and not on the movement of the object, the author argues that distance differences between objects and cameras are a serious way that can produce errors in the histogram estimation. Therefore, one of the challenges in automated analysis of video scenes is the ability to adapt to different locations and to the actual characterization of the scene. In this paper we present a novel approach with adaptive switching capabilities in different application environments and able to detect abnormal human movements in video scenes. Such an approach has time to measure and recover and extract objects from the original objects moving to the scene. A non-parametric learning algorithm is used to classify an object's motion as conventional or nonconventional.

## II. LITERATURE SURVEY

The purpose of finding an object is to find all the conditions of the objects from a known category, such as people, cars or faces in an image. Usually, only a small number of objects are present in the image, but there are a large number of locations and scales that may occur and that need to be explored in some way. Each detection of the image is reported with some form of pose information [1]. This is as simple as the location of the object, a location and scale, or the extent of the object defined in terms of a bounding box. In some other situations, the pose information is more detailed and contains the parameters of a linear or non-linear transformation. For example for face detection in a face detector may compute the locations of the eyes, nose and mouth, in addition to the bounding box of the face. An example of a bicycle detection in an image that specifies the locations of certain parts is shown in Figure 1. The pose can also be defined by a three dimensional transformation specifying the location of the object relative to the camera. Object detection systems always construct a model for an object class from a set of training examples. In the case of a fixed rigid object in an image, only one example may be needed, but more generally multiple training examples are necessary to capture certain aspects of class variability

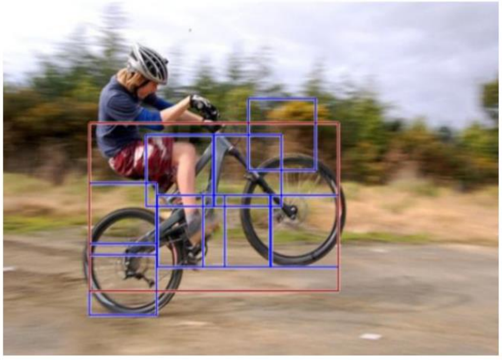


Figure 1

Convolutional implementation of the sliding windows Before we discuss the implementation of the sliding window using convents, let us analyze how we can convert the fully connected layers of the network into convolutional layers. Fig. 2 shows a simple convolutional network with two fully connected layers each of shape

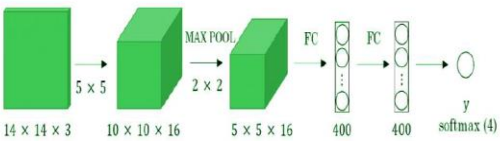



Figure 2

A fully connected layer can be converted to a convolutional layer with the help of a 1D convolutional layer [2]. The width and height of this layer is equal to one and the number of filters are equal to the shape of the fully connected layer. An example of this is shown in Fig 3.

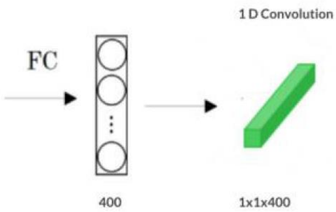



Figure 3

We can apply the concept of conversion of a fully connected layer into a convolutional layer to the model by replacing the fully connected layer with a 1-D convolutional layer. The number of filters of the 1D convolutional layer is equal to the shape of the fully connected layer [3]. This representation is shown in Fig 4. Also, the output softmax layer is also a convolutional layer of shape (1, 1, 4), where 4 is the number of classes to predict.

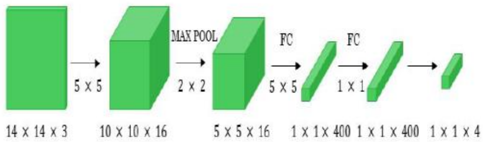



Figure 4

Now, let's extend the above approach to implement a convolutional version of the sliding window [4]. First, let us consider the ConvNet that we have trained to be in the following representation (no fully connected layers).

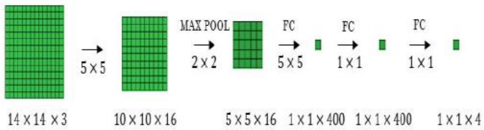



Figure 5

Let's assume the size of the input image to be 16 × 16 × 3. If we are using the sliding window approach, then we would have passed this image to the above ConvNet four times [5], where each time the sliding window crops the part of the input image matrix of size 14 × 14 × 3 and pass it through the ConvNet. But instead of this, we feed the full image (with shape 16 × 16 × 3) directly into the trained ConvNet (see Fig. 6). This results will give an output matrix of shape 2 × 2 × 4. Each cell in the output matrix represents the result of the possible crop and the classified value of the cropped image. For example, the left cell of the output matrix (the green one) in Fig. 6 represents the result of the first sliding window. The other cells in the matrix represent the results of the remaining sliding window operations.

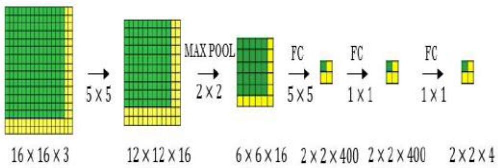



Figure 6

The stride of the sliding window is decided by the number of filters used in the Max Pool layer. In the example above, the Max Pool layer has two filters, and for the result, the sliding window moves with a stride of two resulting in four possible outputs to the given input. The main advantage of using this technique is that the sliding window runs and computes all values simultaneously.

Consequently, this technique is really fast. The weakness of this technique is that the position of the bounding boxes is not very accurate.

A better algorithm that tackles the issue of predicting accurate bounding boxes while using the convolutional sliding window technique is the YOLO algorithm [6]. YOLO stands for you only look once which was developed in 2015 by Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. It is popular because it achieves high accuracy while running in real-time. This algorithm requires only one forward propagation pass through the network to make the predictions. This algorithm divides the image into grids and then runs the image classification and localization algorithm (discussed under object localization) on each of the grid cells. For example, we can give an input image of size 256 × 256. We place a 3 × 3 grid on the image (see Fig. 7).

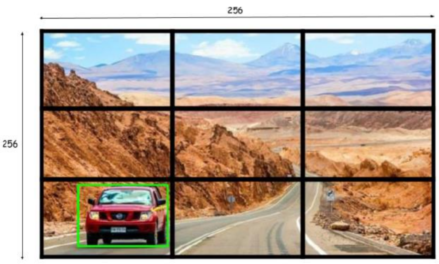



Figure 7

Next, we shall apply the image classification and localization algorithm on each grid cell. In the image each grid cell, the target variable is defined as Yi, j= [pcbxbybhbwc1c2c3c4] T (6) Do everything once with the convolution sliding window. Since the shape of the target variable for each grid cell in the image is 1 × 9 and there are 9 (3 × 3) grid cells, the final output of the model will be:

$$Final\ Output = \underbrace{3 \times 3}_{\text{Number of grid cells}} \times \underbrace{9}_{\text{Output label for each grid cell}}$$

The advantages of the YOLO algorithm is that it is very fast and predicts much more accurate bounding boxes. Also, in practice to get the more accurate predictions, we use a much finer grid, say 19 × 19, in which case the target output is of the shape 19 × 19 × 9 [7].

Identify applicable sponsor/s here. If no sponsors, delete this text box (*sponsors).*

## III. BACKGROUND

Current neural systems are often referred to as large neural systems. Aside from the fact that various neural systems have been in existence since the 1980s, a few reasons left a good adaptation of the system with various hidden layers. One of the basic issues is the scourge of greatness. As the number of objects increases, the number of different elements of a multiplicity becomes larger. As the number of combinations builds, the number of preparation tests should increase linearly. Gathering preparation data of sufficient size is appropriate and large or inaccessible. Fortunately, the actual information is not separately distributed and often involves a building, where the information available lies in the basement. The complex theory expects that most aggregated information is invalid or unusual. We can reduce the size by finding out how to communicate information using structural indicators. Another way to guess is to expect an existing consensus. This means expecting the amount of neural system that detects how it is made does not change much in a small region. Over the past decade, neural systems have undergone a renaissance, in large part due to the accessibility of all the impressive PCs and the massive storage data. In the mid-2000s, it was found that neural systems could be well adapted using auxiliary processing units. GPUs are more efficient at doing this than standard CPUs and they offer a modest choice unlike professional machines. Today, analysts routinely use consumer reality cards, for example, the NVIDIA Tesla K40. Other possibilities for advancement in further understanding include the capabilities of standardized horizontal and horizontal joint implant capabilities and detailed crossing capabilities and sigmoidal input capabilities. With great learning, there is no predetermined need for handmade AI preparations already in use. The classic style design framework, for example, incorporates a handmade component that puts the recognition phase ahead of the AI phase. Comprehensive deep learning is embedded in a single neural system. The lower layers of the neural system see how they can identify the basic points, and then pay attention to the higher parts of the system.

## COMPUTER VISION

Computer vision extracts meaningful information from the content of digital images or video. This is different from just image processing, which involves changing visual information at the pixel level. Applications of computer vision include image classification, visual detection, 3D scene reconstruction from 2D images, image retrieval, augmented reality, machine vision and traffic automation. Today, machine learning is an essential part of many computer vision algorithms (44). Such algorithms can be described as a combination of image processing and machine learning. Effective solutions require algorithms that can handle large amounts of information in visual images, and for most applications, computationally, computation can be performed in real time (28).

## OBJECT DETECTION

Object discovery is one of the classic problems of computer vision and is often described as a difficult task. In many ways, it is similar to other computer viewing tasks, in that it involves creating a solution that is not connected to deformation and changes in brightness and spectra. What makes object detection a unique problem is that it includes detection and image classification [20]. Part of the area is not required, for example, for the whole picture. To find an object, we need to have an idea of where the object may be and how the image is separated. This creates a kind of chicken and egg, in which, to see the composition (and category) of an item, we need to know its location, and recognize the location of something, we need to know its composition. [53] Some of the different physical objects, such as clothing and the human face, can be parts of the same thing, but it has been difficult to know this without first seeing the original object. On the other hand, some objects appear only slightly from the background, requiring separation before recognition. [51

## IV. METHODOLOGY

The following is the method for visual object tracking in videos, which includes object detection and tracking using Yolo. The whole implementation is done in Python.

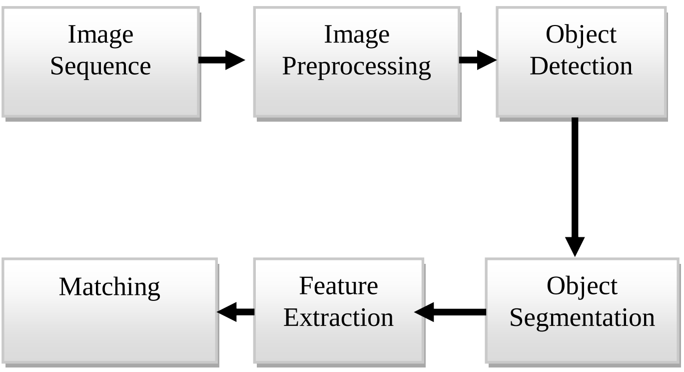



Fig-1 Basic flow diagram of object tracking

The YOLO framework, on the other hand, treats object detection differently. It takes the whole picture in one case and predicts the bounding box coordinates and class probabilities for these boxes. It is very fast and can process up to 45 frames per second. The algorithm applies a single neural network to the full image, and then divides the image into regions and predicts bounding boxes and probabilities for each region. These bounding boxes are weighted by the predicted probabilities.

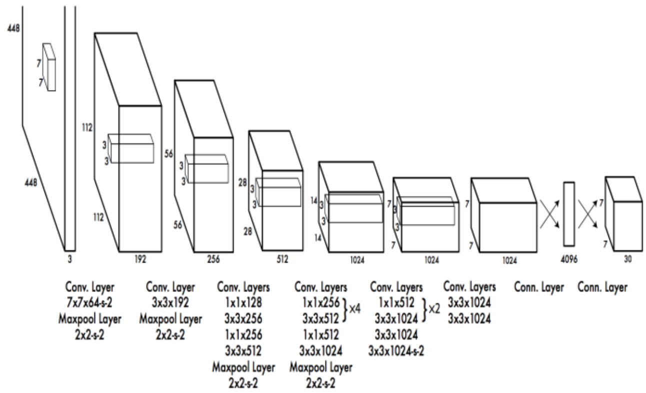



Fig -2 Network design

Steps involved in YOLO object detection:

- YOLO first takes an input image.
- The framework then divides the input image into grids.
- Image classification and localization are applied on each grid. YOLO then predicts the bounding boxes and their corresponding class probability for objects.

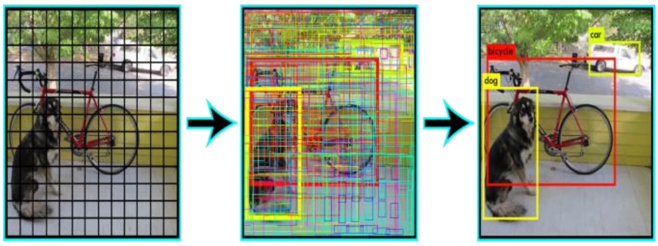
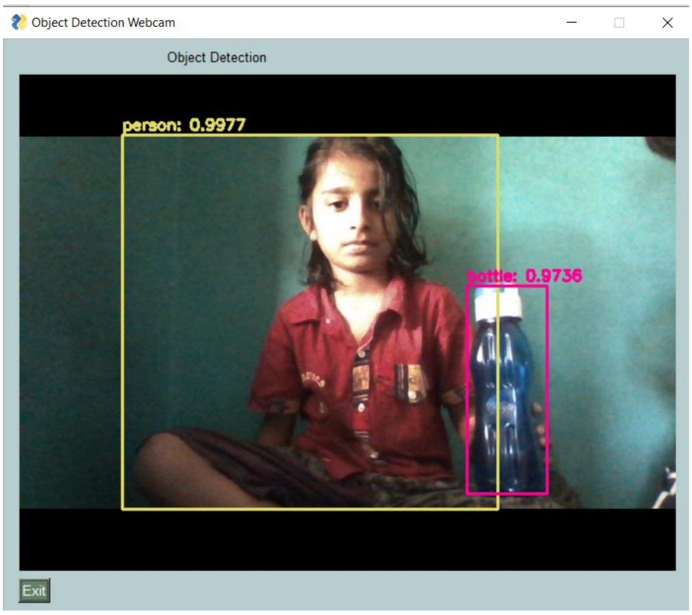


Fig-3

## RESULTS

Validation of orientation results is performed on the top of correct object detections. So, in this part, we focus on the orientation of objects that are correctly detected. We got orientation precision of 84.89% with processing speed of 0.031 seconds per image.




## CONCLUSION

In this paper, we proposed that, YOLO algorithm for the purpose of identifying objects in real time. This algorithm is generalized, which performs different strategies for normalizing different domains from natural images. The algorithm is easy to build and can be trained directly on the full image. YOLO has access to the whole picture in assessing boundaries. And it predicts fewer false positives in background areas. This algorithm is a very efficient and fast algorithm to use in real time compared to other classification algorithms.

## REFERENCES

[1] Niu, W., Long, J., Han, D., Wang, and Y.F.: Human activity detection and recognition for video surveillance. In: IEEE Intl Conf Multimedia and Expo, pp. 719–722. IEEE Computer Society Press, Los Alamitos (2015).

[2] 1.A.G.Hochuli, L.E.S.Oliveria,A.S.Britto Jr, A.L.Koerich: Detecton and Classification of human movements in video scenes. In: IEEE Annual conf on pervasive computing and communications 416-421. IEEE Computer society press (2015).

[3] Mukesh Tiwari,Dr.Rakesh Singhai: A review of detection and tracking of object from image and video sequences.In: IEEE Computer society,DC,pp,142-149(2016).

[4] Onkar R.Kirpan,Pooja I.Baviskar,Shivani D.Khawase,Anjali S. Mankar, Karishma A. Ramteke:Object detection on raspberry pi.In: IEEE intl Trans PP, Computer society 2277-6370(2016).

[5] Vijayalaxmi, K.Anjali: Object detection and tracking using image processing. In: Global Journal of Advanced Engineering Technologies, ISSN: 2277-6370 (2017).

[6] Aggarwal A., Biswas S, Sural and Majumdar: Object tracking using background substraction and motion estimation in MPEG videos In Asian conference on computer vision 121-130 (2017).

[7] Amandeep, Goyal M: Moving object detection techniques Int.J Computer science society,345-349 (2018).