**Intelligent Intrusion Detection Using TensorFlow**

Satyabrata Maity1 , Adarsh Padhi2 , Gyanaranjan Sahoo2

Computer Science & Engineering12, Siksha ‘O’ Anusandhan University, Odisha, India12,

[satyabratamaity@soa.ac.in](mailto:satyabratamaity@soa.ac.in)1, {gyanaranjan7205,adarshpadhi14}@gmail.com2

**Abstract:** Human Detection is a branch of Object Detection. Object Detection is the task of identifying the presence of predefined types of objects in an image. This task involves both identification of the presence of the objects and identification of the rectangular boundary surrounding each object (i.e. Object Localization). An object detection system which can detect the class “Human” can work as a Human Detection System. To overcome the difficulties of the current scenarios where people are wasting much of time and resources in detecting intrusions; we proposed a solution utilizing Machine Learning which could get us out of this hefty situation by trimming videos containing suspicious activity

The optimizations of video could possibly shorten the time and memory data consumption. It could also help in searching suspicious or malicious activities easily

**Keywords:** TensorFlow, Machine Learning, Object Detection, Optimizations.

**1.Introduction**

Most of us in our daily life, are worried about the thieves, accidents, murders, etc. Everyone is aware of this and have love for their life. So, they appoint Watchmen, Security Guards, Bodyguards in order to protect their lives’. Now a day’s people were much serious with their belonging. So every were a CCTV camera is being installed in order to capture all the things that is a beyond human capability.

Incase,

1. There is a situation in which a security guard misses an hour or two and in that particular situation the incident occur then what will you do?
2. There is a situation in which a security guard misses an hour or two and after that he realize to look after the footage providing he will miss the current footage reading

We have overcome for both the situation, By Motion Detection it could really helpful and can bring a change in our society. We have designed a software that can possibly overcome the problems mentioned above.

***Motion Detection*** is the process of detecting a change in the frame of an object relative to its surroundings or a change in the surroundings relative to an object. It also implements ***Face Detection*** which is a computer technology being used in a variety of applications that identifies human faces in digital images. It also refers to the psychological process by which humans locate and attend to faces in a visual scene.

In this paper, we will be explaining the basics and applicability of Motion Detection and the effectiveness of this technology for taking valid frame footage, sharing of data through messages and online streaming which were required for the user to be alert and can take precaution before dealing with the situation.

**3.PROPOSED SOLUTION**

We have designed a software which will reduce the time as well as the data consumption rate. The project “Intelligent Intrusion Detection” perfectly defines itself. In this project we have used Machine learning which enables to detect between humans and all other species and objects. We have used TensorFlow model with MSCOCO.

TensorFlow is a [free](https://en.wikipedia.org/wiki/Free_software) and [open-source](https://en.wikipedia.org/wiki/Open-source_software) [software library](https://en.wikipedia.org/wiki/Library_(computing)) for [dataflow](https://en.wikipedia.org/wiki/Dataflow_programming) and [differentiable](https://en.wikipedia.org/wiki/Differentiable_programming) programming across a range of tasks. It is a symbolic math library, and is also used for [machine learning](https://en.wikipedia.org/wiki/Machine_learning) applications such as [neural networks](https://en.wikipedia.org/wiki/Neural_networks) and , MSCOCO is a large-scale object detection, segmentation, and captioning dataset. COCO has several features: Object segmentation. Recognition in context.

The detection concept is basically based on the Neural Networks basically RCNN i.e., (Region Based Convolution Neural Networks). It generates bounding boxes and segmentation masks for each instance of an object in the image. It's based on Feature Pyramid Network (FPN) and a ResNet101 backbone.

The application takes a video or webcam stream as an input and writes out the frames containing any human intrusion, thereby making a brief video. This results to shorten of original video and data consumption rate which leads to save of time and storage space.

This project also plays a vital role in protection, due to its double layered security. Firstly, when it finds the intrusion, the first level of security gets activated and it sent a Message or a call to the owner about informing that there is some suspicious in your house. If the application detects weapons like (knife, guns, fire) then the 2nd Level of security gets activated and it sends a message to owner and on confirmation or no response from owner for over 10 minutes it sends location and an alert message to the nearby Police Station.

The application also supports Live Streaming of webcam or video input over Internet to platforms like YouTube and Twitch.

1. **Prototype**

Human detection in a smart surveillance system aims at making distinctions among moving objects in a video sequence. The successful interpretations of higher level human motions greatly rely on the precision of human detection.The detection process occurs in two steps: object detection and object classification by deep learnng techinques. Among the deep learning based techniques, two broad class of methods are prevalent: two stage detection (RCNN [1], Fast RCNN [2], Faster RCNN [3]) and unified detection (Yolo [4], SSD [5]). The major concepts involved in these techniques have been explained below:

The network used in this project is based on Single shot detection (SSD) [5]. The architecture is shown below:



The SSD normally starts with a VGG [6] model, which is converted to a fully convolutional network. Then we attach some extra convolutional layers, that help to handle bigger objects. The output at the VGG network is a 38x38 feature map (conv4 3). The added layers produce 19x19, 10x10, 5x5, 3x3, 1x1 feature maps. All these feature maps are used for predicting bounding boxes at various scales (later layers responsible for larger objects). Thus the overall idea of SSD is shown below. Some of the activations are passed to the sub-network that acts as a classifier and a localizer.



Anchors (collection of boxes overlaid on image at different spatial locations, scales and aspect ratios) act as reference points on ground truth images. A model is trained to make two predictions for each anchor:

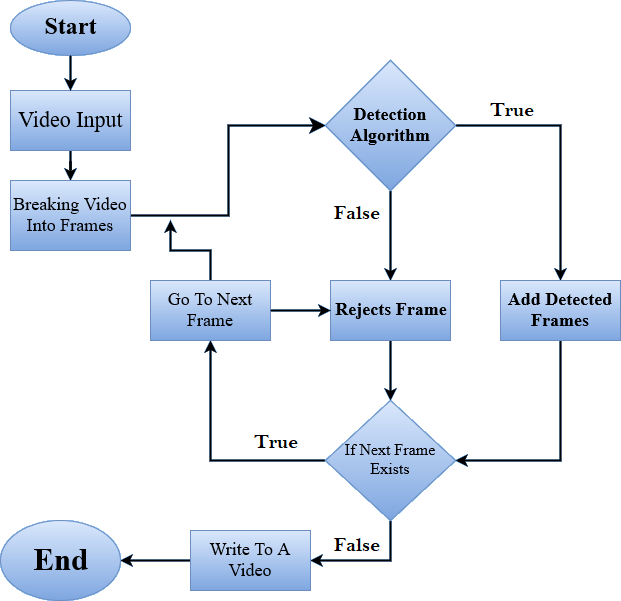
A discrete class

A continuous offset by which the anchor needs to be shifted to fit the ground-truth bounding box.

During training SSD matches ground truth annotations with anchors. Each element of the feature map (cell) has a number of anchors associated with it. Any anchor with an IoU (jaccard distance) greater than 0.5 is considered a match. The loss function used is the multi-box classification and regression loss. The classification loss used is the softmax cross entropy and, for regression the smooth L1 loss is used. The SSD model consists of the base network derived from RESNet and then the modified convolutional layers for fine-tuning and then the classifier and localizer networks. This creates a deep network which is trained end-to-end on the dataset.

The project can be used in 2 different modes exclusive to each other i.e. storage mode, surveillance mode.

Storage Mode: This mode writes only those frames in which any human intrusion is detected to a video file along with current timestamp. The video source can either be a video file or a live webcam. This will help user in several ways like – producing a brief output video such that it can be reviewed fast thereby saving time, taking a much smaller storage space than the original video should have taken thereby saving resources and storage space.



Surveillance Mode: Under this mode, the application actively takes part in security of user. It has 2 levels of security. In 1st level, it informs the user when it finds any human locomotion in its area of vision by either a SMS or a call. On 2nd level, when found any weapons like knife, daggers, guns or fire it informs the user by a message with a link to deactivate the alarm if the alarm is not deactivated in 10 minutes the location along with an alert message is sent to nearest Police Station. Live Stream of video to websites like YouTube and Twitch was also made possible by using ffmpeg library.

Future Work

The Machine Learning Model can be used to detect organisms other than humans

It can be used in research purposes (in places where u wait for an event to happen) to inform you when that event happening like monitoring tigers near forest areas, etc.

The application can be used to detect any causalities like fire, robbery and put on alarm or inform the owner so that necessary actions can be taken.