# Theft Detection System using Convolutional Neural Network and Object Tracking

Pallav Doshi[1], Shubhankar Punktambekar[2], Niraj Kini[3], Simarjeet Singh Dhami [4]

[1,2,3,4] *Student, Department of Computer Engineering, NBN Sinhgad School of Engineering, Maharashtra, India*

## ABSTRACT

*Today modern world is highly digital and highly connected generating enormous data over the internet with high speed. Such data has led many researchers, scientific professionals to explore this data to enable computing machines to understand the real world as a real human being does. Image processing has been explored as the field to provide machines with a vision of their own of which object detection and its tracing throughout the frame has emerged as the most important and vastly explored topic. In recent years due to the great ability and power with feature learning of the Convolutional Neural network (CNN) it has received an overwhelming interest from the computer vision community, leading through many significant breakthroughs. Firstly the paper will introduce the basic building blocks of CNN. Secondly this paper will dive into the object detection practices which then will lead us to object tracking. Finally the paper will be explain how a theft can be detected by combining object detection and object tracking methodologies.*

**Keyword:** - *Computer vision, Convolutional Neural Network, object detection, object tracking.*

## 1. INTRODUCTION

Due to the highly connected and digital world, we are experiencing a humungous generation of image data at a very rapid pace every single day. This generation of data is very difficult for a human to explore and generate necessary insights manually. Hence we need to deploy computing machines to help us understand this raw data and enable us to implement the findings into an application which will help the human race to solve much complex problems. With greater exploration in field of image processing researchers are not only interested in image classification but are greatly inspired to accurately obtain the semantic category of the object and its precise location in the frame and further to continuously track that location. Object detection and tracking technology works on the target object with the help of various theories and methodologies of image processing and pattern recognition.

But to build a real world object detection tracking technology is much of a complex process and requires a high skill of precision and a clear and structured data. Complex backgrounds, noise disturbance, low-resolution, scale and other factors lead to seriously affecting the performance of our system. Hence a variety of solutions were explored leading us to Convolutional neural network (CNN). CNN as a successful methodology of Deep learning has the ability to extract features hierarchically, further all the research has led to a concrete conclusion that CNN has strong power of discrimination and generalization of features.

## 2. NEURAL NETWORKS

Neural networks are basically set of algorithms stacked up together in a manner analogous to human brain. These networks interpret data through machine perception, labelling of data or by clustering the data. The patterns recognized by these networks are structured in a vector into which the data of any kind be it images, sound, text, must be translated. The layers in the neural network are made up of nodes which mimic the functioning of a neuron in the human brain. These nodes are nothing but the area where computations happen. A node combines the inputs from the data with the weights which amplifies or dampens the input, hence giving the input a significance value with regards to the task which the network wants to learn.

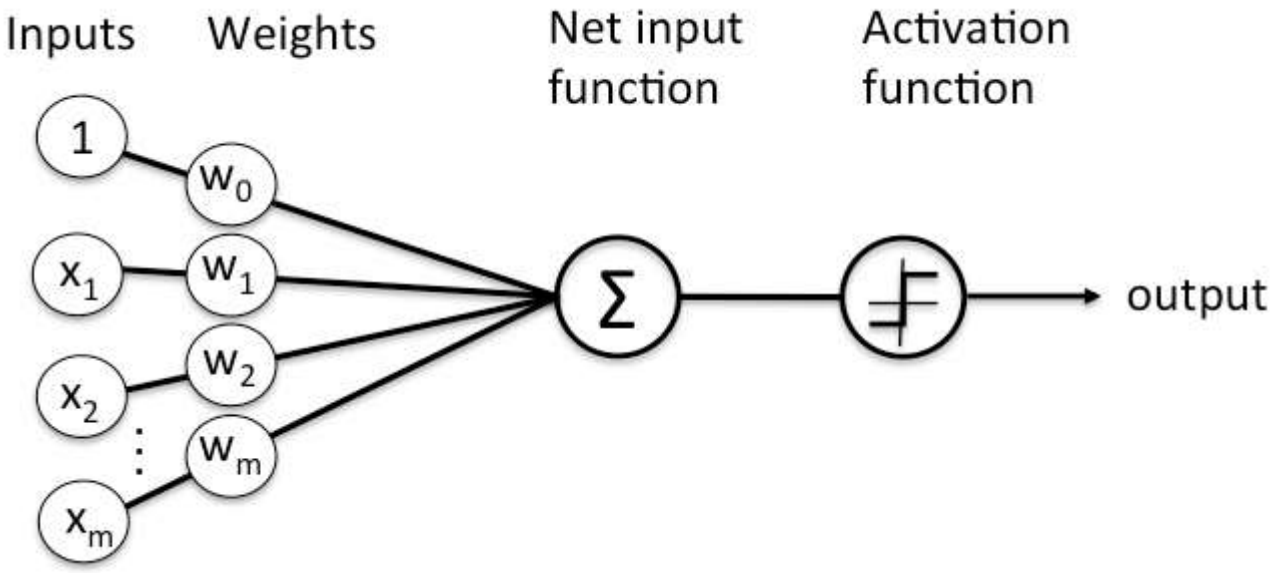


**Fig. 1.** Basic Neural Network Structure

## 3. CONVOLUTIONAL NEURAL NETWORK

Convolutional neural network are deep artificial neural networks which widely used in image classification, image clustering so to perform object recognition within scenes.

Mathematically a convolution is the integral measure of how much two functions overlap each other as one passes over the other. The convolutional nets pass many filters over the image hence each of the nets gathers different signals. At a very early layer one can imagine the convolution nets passing a horizontal line, a vertical line and a diagonal line and hence creating a map of the edges in the image.

Convolutional networks take the slices of the images feature space created by the filters and map them one by one.

Hence by learning different positions of the feature space the convolution networks provide a scalable feature engineering mechanism.

Convolutional neural networks accepts input in the form of volumes that is in the form of three dimensional structure, it does so because every image is defined by three color scales red-blue-green. As a result the convolutional network creates three separate horizons of these colors stacked up one on top of the other. The training of CNN mainly uses forward propagation or back propagation. This training is supervised learning process that requires image data as input and corresponding labels to optimize the network parameters finally obtaining optimized-weight model.

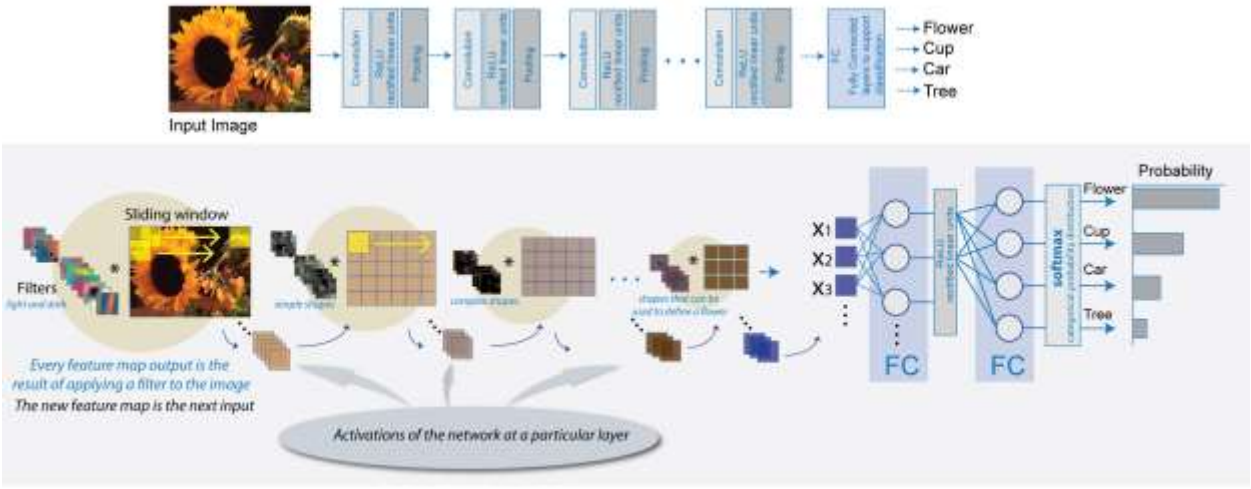


**Fig. 2.** CNN Structure

Applications of CNN include optical character recognition to digitize text and make natural language processing possible, image recognition, video analysis, Checkers game, and many more.

## 4. TRADITIONAL OBJECT DETECTION METHOD

Traditional object detection method is divided into four stages:
• Candidate region generation: This stage is the very first stage which locates the location of the object in the image. This is carried generally using sliding window procedure, as the object position will be anywhere in the image area. The disadvantage of sliding window technique is the large time complexity and redundant windows.
• Feature extraction: Feature extraction decides the performance of the system. However it is very difficult to design a robust feature extraction module as many factors such as object movement, illumination change, noisy background, etc. affect extraction. Scale Invariant Feature Transform (SIFT), Histograms of Oriented Gradients (HOG) and Local Binary Patterns (LBP) are some of the feature extraction mechanisms which are widely used.
• Classification: In this stage Support Vector Machine (SVM) or AdaBoost classifiers are widely used to classify the extracted features.
• Revise Detection Results: After classification there are still many redundant windows, hence it is necessary to remove redundant windows and optimize the detection results by Non-Maximum Suppression (NMS) and combining overlapped Bounding box.

Traditional object detection has two main problems: first, the method of using sliding window is not robust enough,

## 5. ACKNOWLEDGEMENT

Object tracking is the methodology of tracing an object throughout a sequence of images or frames. Object tracking allows us to determine the relative movement of the target object with respect to the other objects present in the images. In a tracking scenario objects can represented by different shapes such as points, primitive geometric shapes, contours, etc. Object tracking involves spatial as well as temporal changes**.**
Steps involved in object tracking are:
• Segmentation This process involves identifying components in the image. It involves operations like boundary detection, thresholding, etc. Thresholding is the process of reducing the grey levels of the image.
• Foreground/background extraction. As the name suggests this method separates the foreground and background of the image depending upon the presence of the concerned object.

• Calculations are the image modifications applied on the image which are required for object tracking.
• Tracking is the last step which involves the last bit of actually utilizing every above mentioned step to actually track the concerned object in the image.

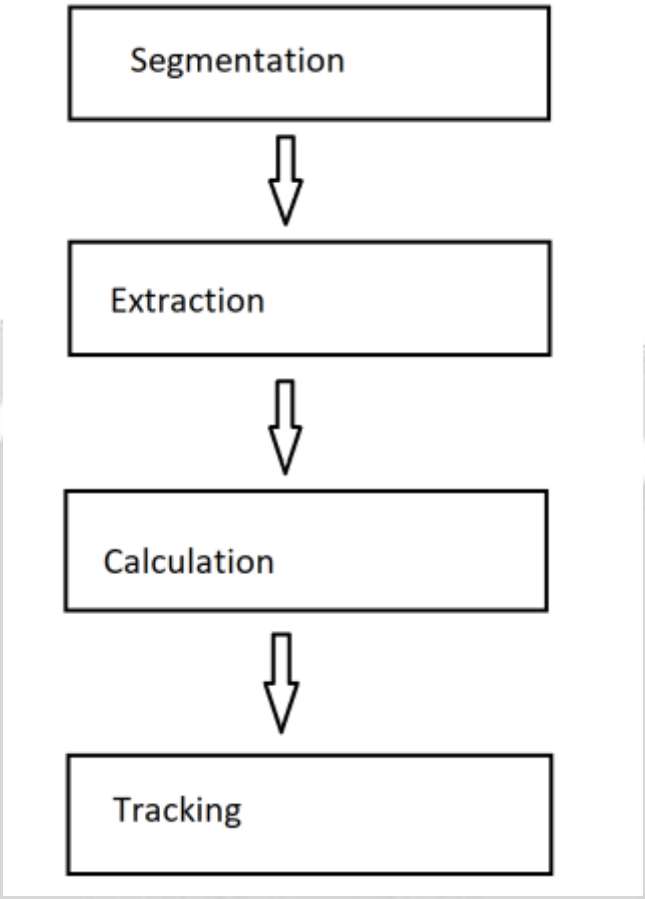


**Fig. 3.** Basic Steps of Object Tracking

## 6. PROPOSED THEFT DETETCTION SYSYTEM

We have built a theft detection system for our college computer lab. The theft detection system identifies the computer screens (monitors) and people in the frame. The computer monitor is our target object to be protected from being stolen and hence it is highly important for us to detect the computer screens with definite accuracy.
The following steps are followed as we move ahead:
1) The program starts with an interface consisting of a button to start the stream.
2) Pressing the button invokes the video stream through the camera.
3) The system then starts detecting the monitors and people in the frame.
4) A boundary is defined within the frame within which the monitors are considered to be safe.
5) If the monitor moves out of the defined boundary, the system assumes that the monitor has been stolen.
6) The detected theft is followed by an alert to the concerned person/administrator.

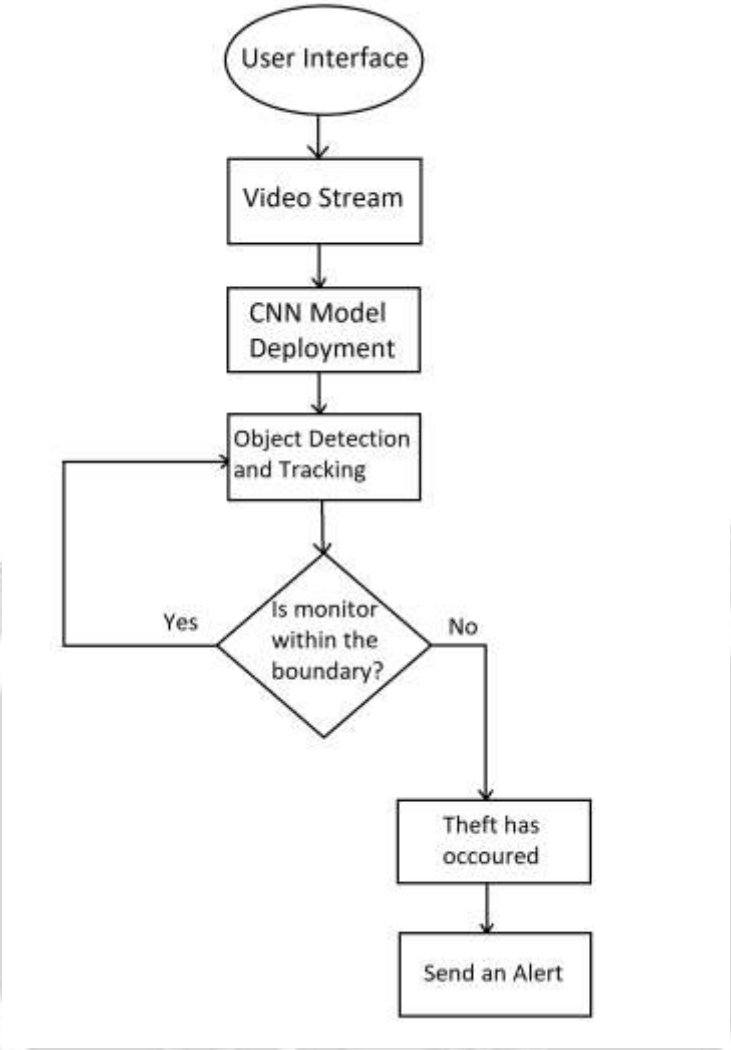


**Fig. 4.** Proposed System

## 7. DATASET

For any Convolutional neural network to be trained we need to have huge amount of data. We sourced our data from ImageNet, which is a world leading website for providing image dataset for experimental and research work. We used about 700 images of monitor to train our neural network. This shows that in order to train a model with good accuracy we need huge amount of data.


**Fig. 5.** Images from the Dataset

## 8. FRAMEWORK

Our model is developed using CAFFE framework utilizing SSD [Single Shot Multi box Detector] as a detector and MobileNet as the neural network. CAFFE is an open source deep learning framework developed by University of California. This framework supports different deep learning architectures which are aimed at image classification and object detection. SSD consists of two steps:

1) Extract feature maps.

2) Apply convolutions to the detect objects. Combining these two result in the detection of the target object.

## 9. RESULT & DISCUSSION

**Condition 1**: When the video stream begins the trained CNN network starts to detect the monitor and verifies whether the detected monitor lies within the predefined boundary. If the above condition satisfies, the model continues to check for the same condition until it fails i.e. now, the monitor does not lie in the predefined boundary.

**Condition 2**: When the monitor is detected outside the predefined boundary, the conclusion is made that the monitor is being stolen. Thus to inform about the same, the system informs the authorized personnel regarding the theft by sending an alert.

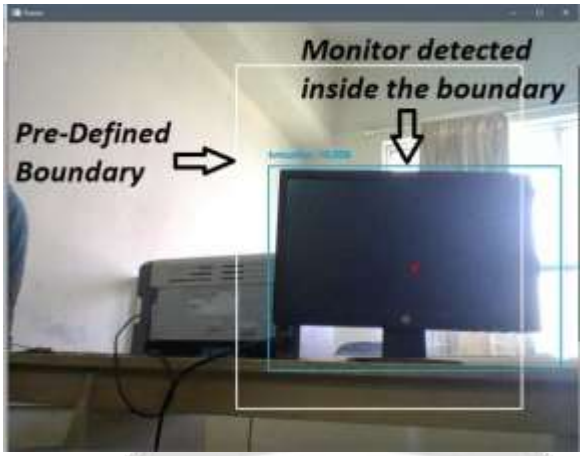


**Fig. 6.** Condition 1: Monitor detected within the boundary.

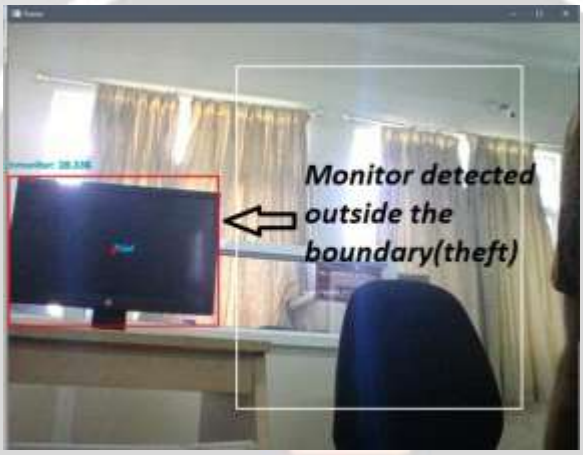


**Fig. 7.** Condition 2: Theft detected

## 10. CONCLUSION

Thus we can arrive to the conclusion that Convolutional neural networks are highly capable of detecting objects and hence providing the computer a sense of vision. In the above proposed system we can clearly see how an object detection and tracking system can potentially avoid theft. Such systems on a large scale can add a new dimension to the camera enabled surveillance system.

## 11. ACKNOWLEDGEMENTS

We would like to thanks the Department of Computer Engineering, NBN Sinhgad School of Engineering for providing assistance and support for this work.

## 12. REFERENCES

[1] Wang Zhiqiang and Liu Jun ”A Review of Object Detection Based on Convolutional Neural Network”,Proceedings of the 36th Chinese Control Conference July 26-28, 2017

[2] Aiswarya S Kumar and Elizabeth Sherly ”A convolutional neural network for visual object recognition in marine sector”,2017 2nd International Conference for Convergence in Technology (I2CT)

[3] Tianmei Guo, Jiwen Dong ,Henjian LiYunxing Gao ”Simple Convolutional Neural Network on Image Classification ”,2017 IEEE 2nd International Conference on Big Data Analysis

[4] Nadia Jmour, Sehla Zayen, Afef Abdelkrim ”Convolutional Neural Networks for image classification”

[5] Fatih Ertam, Galip Aydin ”Data Classification with Deep Learning using Tensorflow”,(UBMK’17)2nd International Conference on Computer Science and Engineering