# Recognition of Handwritten Digits/Char using Machine Learning

Kumbhar Rutuja Chandrakant[1*], Gaikwad Shraddha Dhanaji[2], Kenjale Sanjana Ashok[3], Kadam Ujwal Namdeo[4]

[1,2,3,4]*Department of Computer Engineering, SVPM's College of Engineering, Malegaon, India*

*Abstract*: **Handwritten Digits/Characters an idea of recognize digit which is used for the computers. Object detection technique is used for detecting the digits and characters in English language i.e., digits from 0-9 and characters from (A-Z), (a-z) Capital letters and small letters. Object detection is performed by Cascade classifier. Cascade classifier class to detect objects in a video stream. The cascades are a bunch of XML files that contain OpenCV data used to detect objects. There are a unit 2 rule area unit employed in this technology i.e., create rule and YOLO rule. Pose estimation refers to laptop vision techniques that discover human figures in pictures and videos. The rule is just estimating wherever key body joints area unit. Human create estimation is the method of estimating the configuration of the body (pose) from one, usually monocular, image. YOLO is associate abbreviation for the term 'You solely Look Once'. this is often associate rule that detects and acknowledges numerous objects in a very image (in real-time). Object detection in YOLO is completed as a regression downside and provides the category possibilities of the detected pictures. YOLO is an associate rule that uses neural networks to supply time period object detection. This rule is widespread attributable to its speed and accuracy. It has been employed in numerous applications to discover traffic signals, people, parking meters, and animals.**

*Keywords*: **Machine Learning, Image processing, YOLO, Classification, TensorFlow, Nodejs, Pose estimation.**

## 1. Introduction

It is a computer technology related to computer vision and image processing that deals with detecting objects of a certain class (such as humans, buildings, or cars) in digital images and videos. Object detection has been widely used for Number detection, Letter detection, Face detection, vehicle detection, Fire detection, Weapon detection. object detection indicates that the location of the objects inside the image. Object detection is performed by the Cascade classifier. Cascade classifier class to detect objects in a video stream, the word "cascade" is consisting of several simpler classifiers or stages that are applied subsequently to a region of interest until at some stage the candidate is rejected or all the stages are passed. The cascades are a bunch of XML files that contain OpenCV data used to detect objects The most popular and probably the simplest way to detect Object using Python is by using the OpenCV package. The algorithm may have 30 to 50 of these stages or cascades, and it will only detect a object if all stages pass.one more library file use alternatives to OpenCV, that is

dlib — that come with Deep Learning based Detection and Recognition models. Computer vision is the field of Artificial Intelligence that studies how computers can gain high-level understanding from digital images or videos in order to produce numerical or symbolic information. Every object class has its own special features that helps in classifying the class — for eg., all circles are round. Object class detection uses these special features. For eg., when looking for circles, objects that are at a particular distance from a point are sought. when looking for squares & objects that are perpendicular at corners and have both side lengths are needed. A similar approach is used for face Recognition where eyes, nose, and lips can be found and features like skin color & distance between eyes can be found.

## 2. Methodology

Object detection could be an elementary drawback in pc vision analysis. In recent years, exceptional progress has been created in object detection, arguably benefited from the fast development of deep neural networks based mostly on strategies. Among them, one of the foremost influential strategies is that the R-CNN framework that performs CNN-based classification on the item proposals made by varied strategies. Object detection may be a basic downside in laptop vision analysis. In recent years, exceptional progress has been created in object detection, arguably benefited from the speedy development of deep neural network primarily based ways. Among them, one of the foremost important ways is that the R-CNN framework that performs CNN-based classification on the article proposals created in varied ways.

There are various techniques employed for the process of Digits and Characters recognition.

- *Conversion of video into image frame:* Convert the video into frame by using Cascade classifier for detecting the object like weapon, fire, face etc. which can be achieved by cv2. Cv2 library can be used to execute different operations on videos. Take a video as input & break the video into frame.
- *Image enhancement for object detection:* Image enhancement techniques are increasingly needed for improving object detection.
- *Feature extraction:* Feature extraction includes

---

reducing the amount of resources required to brief a large set of data. feature extraction is a general term for methods of constructing combinations of the variables to get around these situations while still briefly the data with correct accuracy.

- *Object detection:* Object detection has been widely used for digit detection, vehicle detection, pedestrian counting, web images, security systems and driverless cars.

### A. Pose Algorithm

Pose estimation refers to laptop vision techniques that find human figures in pictures and videos. The algorithmic rule is just estimating wherever key body joints square measure.

Human cause estimation is that the type of estimating the configuration of the human pose from one, usually monocular, image. Human cause estimation and chase may be a laptop vision task that features police investigation, associating, and chasing linguistics key points. samples of linguistics key points square measure "right shoulders," "left knees," or the "left brake lights of vehicles."

The performance of linguistics keypoint chase in live video footage needs high procedure resources which has been limiting the accuracy of create estimation. With the newest advances, new applications with period of time needs become doable, like self-driving cars and last-mile delivery robots. Given that the image sensing element (camera) is graduated and therefore the mapping from 3D points within the scene and 2nd points within the image is thought. If conjointly the pure mathematics of the thing is thought, it means the projected image of the thing on the camera image may be a well-known perform of the object's cause. Once a group of management points on the thing, generally corners or alternative feature points, has been known, it's then potential to unravel the cause transformation from a group of equations that relate the 3D coordinates of the points with their 2nd image coordinates. Algorithms that confirm the cause of purpose—to some extent—a degree cloud with regard to another purpose cloud are referred to as point set registration algorithms, if the correspondences between points don't seem to be already celebrated.

### B. YOLO Algorithm

YOLO is an associate degree formula that uses neural networks to supply time period object detection. This formula is in style due to its speed and accuracy. it's been utilized in varied applications to sight traffic signals, people, parking meters, and animals. This article introduces readers to the YOLO formula for object detection and explains however I t works. It additionally highlights a number of its real-life applications. YOLO is associate degree abbreviation for the term 'You solely Look Once'. this can be an associate degree formula that detects and acknowledges varied objects in a very image (in real-time). Object detection in YOLO is completed as a regression drawback and provides the category *chances* of the detected pictures.

YOLO algorithmic program employs convolutional neural networks (CNN) to sight objects in a period of time. Because

the name suggests, the algorithmic program needs solely one forward propagation through a neural network to sight objects.This means that prediction within the entire image is finished in a very single algorithmic program run. The CNN is employed to predict numerous catego ry possibilities and bounding boxes at the same time.The yolo algorithmic program consists of varied variants. a number of the common ones embody small yolo and yolov3.

YOLO algorithm works using the following three techniques:

### 1) Residual blocks

First, the image is divided into various grids. Each grid has a dimension of s x s. The below image display how an choose image is divided into grids. In the image there are multiple grid cells of same dimension. Every grid cell will recognize objects that appear within them. e.g., if an object mid performs within a certain grid cell, this cell will be responsible for recognize it.

### 2) Bounding box regression

A bounding box is a line that high spot an object in an image. Each bounding box in the image consists of the below points:

1. Width (bw)
2. Height (bh).
3. Class (for example, person, car, traffic light, etc.): This is represented by the letter c.
4. Bounding box center (bx, by) The image shows an example of a bounding box. The bounding box has been acting for by a yellow line.

### 3) Intersection over union (IOU)

Intersection over union (IOU) is a phenomenon in object detection that describes how boxes overlap. yolo uses IOU to provide an o/p box that surrounds the objects correctly.
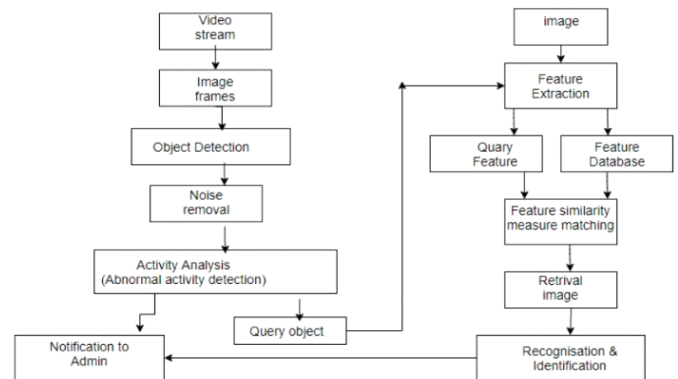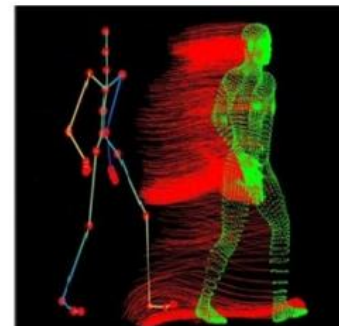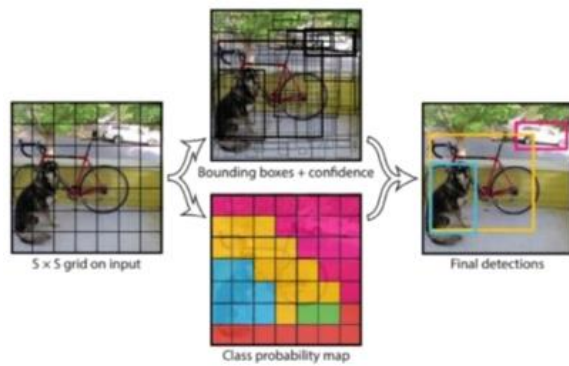


Fig. 1.  Flow chart



Fig. 2.

Fig. 3.

## 3. Conclusion

In this paper, we tend to propose a multi-stage network cascades framework with cluster algorithms learning for object detection. Specially, the projected framework effectively utilizes linguistics segmentation options to help object detection by incorporating the linguistics segmentation network, proposal generation eight network and algorithmic detection network into a unified design. Additionally, a gaggle algorithmic learning theme is projected to recursively score object proposals and regress their bounding boxes considering the locations of the encompassing proposals of a similar object. We tend to show that the projected framework is especially effective in object localization.

## References

[1]  Daniel Bernhardt, Peter Robinson, "Detecting Affect from Non-stylised Body Motions," (2017).
[2]  Guanbin Li, Yukang Gan, Hejun Wu, "Cross-Modal Attentional Context Learning for RGB-D Object Detection (2019)".
[3]  Ke Zhao, "Handwritten Digit Recognition and Classification Using Machine Learning", Technological University Dublin, 2018.
[4]  Peng Chen and Guang-Da Hu, "A Simple Algorithm for Camera Pose Estimation," IEEE International Conference on Automation.
[5]  Shikhar Tandon, Shadab Akhter, Vaibhav Pratap Singh, Nizam Uddin Khan, "Handwritten Digits Recognition Using Machine learning", International Journal of Information Sciences and Application (IJISA), 2019.
[6]  Bapu Chendage, Rajivkumar Mente, "A Study of Digital Image Processing Techniques for Character Recognition," 2020.
[7]  Qichang Hu, Sakrapee Paisitkriangkrai, Chunhua Shen, Anton Van Den Hengel, and Fatih Porikli, "Fast Detection of Multiple Objects in Traffic Scenes with a standard detection framework," IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 4, April 2016.
[8]  Hanxi Li, Yi Li, and Fatih Porikli "DeepTrack: Learning Discriminative Feature Representations Online for Robust Visual Tracking," IEEE transactions on image processing, vol. 25, no. 4, April 2016.
[9]  Jaebum Choi and Markus Maurer" Local Volumetric Hybrid-Map-Based Simultaneous Localization and Mapping with Moving Object Tracking," IEEE Transactions on intelligent transportation systems, 2016.
[10] Stefan Duffner and Christophe Garcia, "Using discriminative motion context for on-line visual object tracking," IEEE Transaction on Circuit and System for Video Technology, 2015.
[11] Shan Gao, Zhenjun Han dynasty, Ce Li, Qixiang Ye, and Jianbin Jiao, "Period of time Multipedestrian pursuit in Traffic Scenes via associate RGB-D-Based stratified Graph Model," IEEE transactions on intelligent transportation systems one, 2015.