# Object Classification and Color Sensing Using YOLO V3

[1]K N V Suresh Varma, [1]*K N V Satyanarayana*, [2]*B Anusha Sai Lakshmi*, [2]*A Ramya Sri*, [2]*CH Vinny*, [2]*B Siva Kumar*

[1] *Assistant Professor*, [2] *B. Tech Students*
[1,2]*Department of ECE, S.R.K.R. Engineering College, Andhra Pradesh, India.*
*Corresponding author mail ID:* [1]

***Abstract***

*The object tracking, recognition and detection is turning into the elemental facet of recent technologies. Alongside the detection of objects, the very next facet typically comes the colour of the object detected which is the main objective of this project. In this paper, we have combined the task of object classification and colour sensing to be done at once which are usually performed separately. The neural network YOLO V3(You Only Look Once, version3) [1] is used for object detection that has been proved to achieve excellent results. In combination with this, for color detection [3], we built an application through which one can automatically get the name of the color by double clicking on the respective area in the obtained output on the screen. This combination would be able to give the output of the image or video with bounding boxes around the objects detected which are classified and trained with their respective names and their respective colours.*

***Keywords*: *YOLO v3, object detection, color detection, image processing, computer vision.***

## 1. Introduction:

Machine Learning has become the most sought of field in data science. It has proved to be a new science of evolution. Machine Learning has a wide range of applications in many fields of science. It has its roots connected to image processing fields also. Apart from image processing, Machine Learning is also widely being used in Biomedical processing for identifying, diagnosing diseases and other medical issues. In video surveillance systems, for detecting criminals, people and desired objects or items. Also, for tracking motion or movement of the objects and people. In day-to-day life, like online recommendations systems, virtual personal assistants and online customer support. In defense applications for detecting mishaps, for maintaining cyber security, for detecting enemies, for target recognition, for tracking missiles. In aerospace, to track airplanes for pilot division systems. Machine Learning is also used in deployment of Radar and its applications.

All of the above-mentioned applications mostly deal with the images or video of the scene in description. Thus, machine learning is a key path in image processing. Machine Learning has made possible the concept of real-time image processing which simply, means applying a machine learning algorithm to a video. These results could be an identified object in the image or video.

The foremost, the basic, the most required and the areas that have always been under research are the object detection recognition and tracking, which is the main focus of this paper. Apart from object identifying its color is also our prime focus and interest also. Being the most challenging problem in computer vision, Object detection [6] has gained a lot of attention by the researchers.

In this paper we used YOLO V3, a neural network for the purpose of Object detection. We used YOLO V3 model to detect the predefined 80 classes of objects in real time. Once the object in the taken image or video frame, are classified accordingly. Next the color of the object is identified, which comes in the picture by double clicking.

Identifying colour for a human eye is an easy task. But this task is not easy for computers. Human eye has light receptors are present, which send the signals to the brain, and then recognizes the color. Likely, we have learnt the names of some colors. In the similar way, we developed a machine learning algorithm by classifying the names of some colors along with their data values i.e., RGB & Hex values. From this we



ISSN: 2233-7857 IJFGCN

will calculate the distance from each color and find the shortest distance. The algorithm which we built for this, is capable to detect the color and give its name. The name is displayed by double clicking at any point on the image or video which is taken as input.

## 2. Related Works:

J. Redmon, S. Divvala, R. Girshick and A. Farhadi., et al. [1], proposed YOLO (You Only Look Once) for object classification and detection using neural networks in C language. It is a single neural network which predicts bounding boxes and class probabilities directly in one evaluation from images. It is extremely fast and outperforms other detection methods, including DPM and R-CNN.

Ramalakshmanna Y, K. N. V. Satyanarayana, Dr P. V. Rama Raju, K. N. V. Suresh Varma., et al. [7], proposed a novel and simple method working on different region of interests present in video or image. It worked on Histogram Orientation of Gradients in image processing events and uses SVM Classification.

Xia Wen Zhang, Zhao Qiu, Ping Huang, Jian Zheng Hu, Jing Yu Luo., et al. [2], explained the use of YOLO v2 for target detection for detecting traffic lights. In combination with it, they proposed the use of HSV color model and designed the ratio of the red and green colours and then, they determined the status of the traffic lights.

A. Ćorović, V. Ilić, S. Đurić, M. Marijan and B. Pavković., et al. [9], provided the demonstration of the usage of the newest YOLOv3 algorithm for the detection of traffic participants. They trained the network for 5 object classes and demonstrated the effectiveness of the approach in the variety of the driving conditions.

Quming Zhou, J.K. Aggarwal., et al [10], worked on Object tracking in Outdoor environment using fusion and multi camera. It defined about the spatial, shape of the object. The results will shown by simple features makes the tracking effective and EKF will improve the tracking accuracy.

H. Altun, R. Sinekli, U. Tekbas, F. Karakaya and M. Peker., et al. [4], introduced a method for an efficient color detection in RGB space using in hierarchical structure, deep neural networks. In this structure, we evaluate R, G, B components of a pixel and then constructs the set of predefined class of colors for the given pixel.

F. S. Khan, R. M. Anwer, J. van de Weijer, A. D. Bagdanov, M. Vanrell and A. M. Lopez., et al [3], proposed the color attributes on object detection. It fuses the shape and color of an image and it will be classified, and significant drop in performance for object detection.

## 3. Methodology:

For classifying and detecting both the object and colour in a video frame at once, we came up with two methods both of which come under supervised learning:

1. Training an algorithm by labelling some objects along with their colour.
2. Considering two different neural networks one to train the objects of different classes and another network which can identify colours individually.

Out of the two solutions the first one was so tedious. And, there was a doubt whether the second solution will work or not. Since the first method proved to be too tedious, we worked on the second method anyway. For object classification, we used Yolo V3 model to detect 80 predefined classes of objects in real time. Hence, the algorithm is divided into two steps in implementation:

Firstly, the object recognition algorithm based on YOLO v3 is used to detect, classify and identify the objects in the image frame and draw a rectangular bounding box around it.



ISSN: 2233-7857 IJFGCN

Secondly, the colour of the object is identified by using a colours dataset [13] containing the RGB and HEX values of 865 colours. The block diagram in figure-3 depicts the process flow of the algorithm in a simple form.

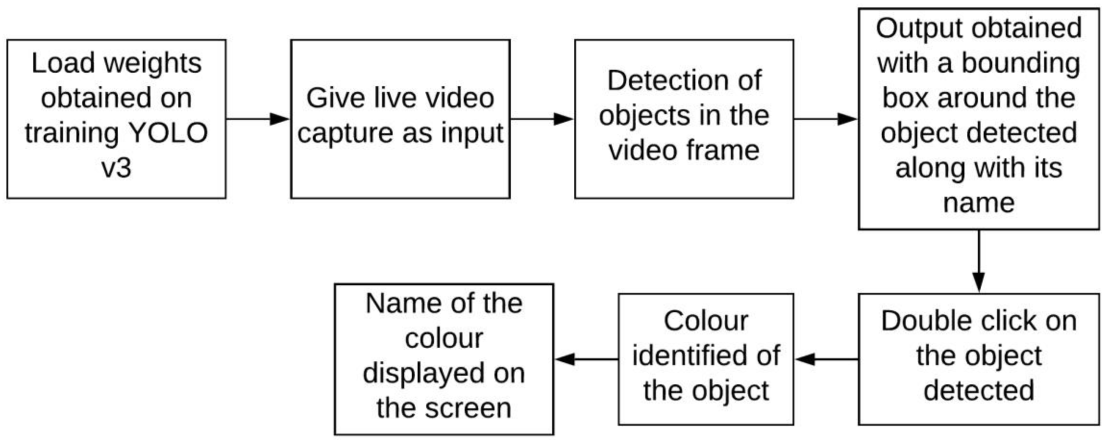


Figure –3: The block diagram representation of the algorithm.

Here is a brief explanation of YOLO-V3 [5]. Darknet-53 is the backbone of yolo v3 architecture. First of all, yolo v3 uses a 53-layer network of darknet which is trained on imageNet. Each layer is followed by batch normalization and leaky ReLU activation. That's why we call it darknet 53. Later it stacks another 53 layers for the task of detection forming a fully convolutional architecture of 106 layers.

Whenever an image is given as input to Yolo V3 model, which takes an image of size 416*416, the network downsamples the image called stride of the network. If the stride of the network is 32, then the image becomes 13*13 giving a total of 169 cells. Then Yolo V3 model predicts 3 bounding boxes for each cell so that the bounding box is responsible for detecting given object if the center of the objet falls in the receptive field of the cell. There may be a possibility of one or more cells in the image detect the same object. To solve this problem, we implemented IOU (Intersection Over Union) function. Each bounding box is represented by a vector of size 5+number of classes. Here Yolo V3 provides 80 classes.

We identify the values that represent normalized coordinates of the centre of the box and object score i.e., probability of the thing detected in the box becoming an object. We set a threshold for the object score. If the object score is greater than the threshold value set, then the algorithm looks at the class probabilities. The remaining 80 values represents the individual probabilities of each class. If the detected object has a value greater than the threshold object score, it outputs the object name as a class which has a greater class probability.

Thus, Yolo V3 model detects the objects [9]. Similarly, Yolo V3 model provides three scales in detecting objects with strides 32,16,8 giving feature maps of size 13*13, 26*26, 52*52, Which generates 507,2028,8112 bounding boxes respectively totally generating 10647 bounding boxes. In which the 13*13 feature maps usually detects large objects in the image, 52*52 feature maps usually detect relatively small objects in the image.

This is working of Yolo [9] in the background. The weights obtained after retraining YOLO V3 [5] likewise are retrieved. After this, the bounding boxes are drawn and name of the object is displayed at the top of the bounding box. Next, the colour of the object was to be determined [2].



ISSN: 2233-7857 IJFGCN

For identifying colour, we used a colours dataset in which the names and RGB [4] values of 865 colours are defined. Firstly, we load the csv file of the colours dataset. Secondly, to get the color name, distance(d) is calculated which tells us how close the color is and the one having minimum distance is chosen. The distance is calculated using the following formula:

d = abs (Red – ithRedColor) + (Green – ithGreenColor) + (Blue – ithBlueColor)

Finally, we set a mouse callback event on the output window so that whenever mouse is double clicked on the window, the colour identification function is called which calculates the minimum distance between the RGB values of the pixel of the object which is clicked and the predefined RGB values of the colours. Finally, the colour of the object is displayed.

## 4. Results:

Figure- 4(a) shows the initial video frame i.e., with the objects to be identified, just before applying the algorithm.

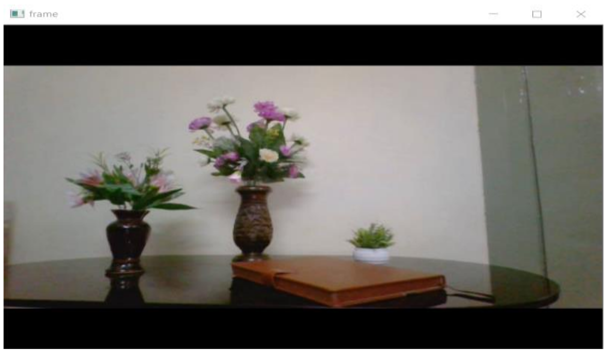


Figure – 4(a): Input video frame before applying the algorithm.

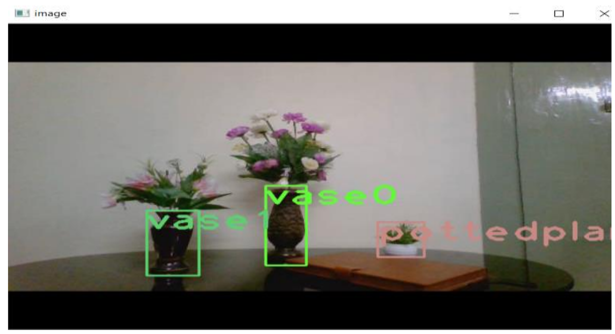


Figure – 4(b): Output obtained by the algorithm with detected objects.

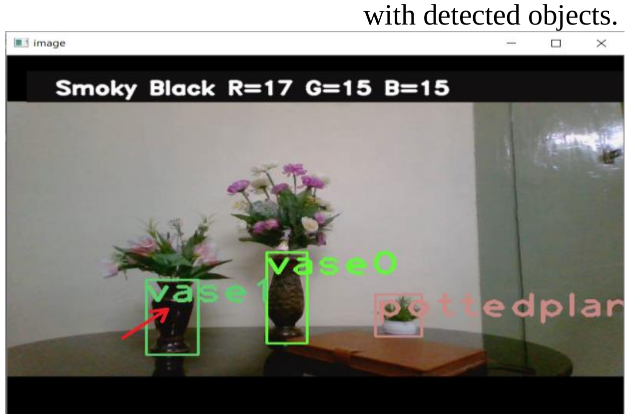


Figure – 4(c): Output with colour sensed, when double clicked at the pointed position shown by a red arrow.

Figure- 4(b) is the output obtained with objects in the video frame classified accordingly with a bounding box around them. There are two vases detected followed by a number beside each one of it, which are of same class. So here, the number is the index of each vase which depicts that 'vase 0' is the first object detected followed by 'vase 2' and the 'pottedplant'.

From figure – 4(c), When double clicked on the video Frame at the point indicated by the red arrow mark in the picture, the colour of the vase 0 is obtained as shown namely, "Smoky Black R= 17 G= 15 B= 15", while the values being the R, G, B values of the colour identified. This is the final output of the algorithm.



ISSN: 2233-7857 IJFGCN

## 5. Conclusion:

In this paper, the combination of YOLO v3 object classification [8] algorithm and color identification model are used to identify an object and its colour in video frame or an image. This design can be extended to also calculate the depth of the object in a video. This can be a benefitting application in robotics for machine intelligence concerned with machine vision. The depth [7], [8] can be found to be handled by a single camera [10] as of monovision. This model can also be further customized for a single object and a single colour or for as many as required objects and colours. Also, some further research in fusion [10] application of the object and colour recognition technologies would be beneficial.

## References:

[1] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 779-788.

[2] Xia Wen Zhang, Zhao Qiu, Ping Huang, Jian Zheng Hu, Jing Yu Luo. "Application Research of YOLO v2 Combined with Color Identification,"

[3] F. S. Khan, R. M. Anwer, J. van de Weijer, A. D. Bagdanov, M. Vanrell and A. M. Lopez, "Color attributes for object detection," 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, 2012, pp. 3306-3313.

[4] H. Altun, R. Sinekli, U. Tekbas, F. Karakaya and M. Peker, "An efficient color detection in RGB space using hierarchical neural network structure," 2011 International Symposium on Innovations in Intelligent Systems and Applications, Istanbul, 2011, pp. 154-158.

[5] Redmon, Joseph & Farhadi, Ali. (2018). YOLOv3: An Incremental Improvement.

[6] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 6517-6525.

[7] Ramalakshmanna Y, K. N. V. Satyanarayana, Dr P. V. Rama Raju, K. N. V. Suresh Varma, "Human Detection and Depth Calculation Using Stereographic Vision with Pixel Disparity and Hog Classifier," journal of computational and theoretical nanoscience, Vol – 16, issue – 4, pp. 1280-1283.

[8] D.Srinivasa Reddy, V.Divya, G V S Padma Rao, K.N.V.Satyanarayana, "Object Classification and Distance Estimation using Deep Constitutional Neural Network and Triangulation Method," Journal of Applied science and Computations, vol- 6, issue- 3, pp. 2820-2828.

[9] A. Ćorović, V. Ilić, S. Ðurić, M. Marijan and B. Pavković, "The Real-Time Detection of Traffic Participants Using YOLO Algorithm," 2018 26th Telecommunications Forum (TELFOR), Belgrade, 2018, pp. 1-4.

[10] Quming Zhou, J.K. Aggarwal, Object tracking in an outdoor environment using fusion of features and cameras, Image and Vision Computing, Volume 24, Issue 11, 2006, Pages 1244-1255

[11] Simon Haykin. 1994. Neural Networks: A Comprehensive Foundation (1st. ed.). Prentice Hall PTR, USA

[12] Szeliski, Richard. Computer Vision: Algorithms and Applications. Ist ed, http://szeliski.org/Book, DOI 10.1007/978-1-84882-935-0

[13] Wikipedia contributers. Lists of colors. Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Lists_of_colors&oldid=942820909

[14] Tijtgat, Nils. (2017, May 16). How to train YOLOv2 to detect custom objects. https://timebutt.github.io/static/how-to-train-yolov2-to-detect-custom-objects/

[15] Gonzalez, Rafael. (2008). Digital Image Processing. (2nd ed.). Pearson Education India.



ISSN: 2233-7857 IJFGCN