

Intelligent Visual Recognition for Real-Time Object and Color Detection

BVN Prasad Paruchuri^{1,*}, Degala Mohana ManiVenkatesh¹, Vankdoth Srinivasa Rao¹, Koppagiri Jyothsna Devi¹, Pulipati Nagaraju², and P. Nagabhushanam³

¹*Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, India*

²*Department of Computer Science and Engineering, Vishnu Institute of Technology, Bhimavaram, India*

³*Department of Computer Science and Engineering, Sri Vasavi Engineering College, Tadepalligudem, India*

*Corresponding author: bvnprasadparuchuri@yahoo.com

Abstract. Object and color detection is an important part of visual recognition and scene understanding, because recognizing an object together with its color provides the contextual cues needed to interpret real-world situations. Accurate detection is difficult, however, because of changes in lighting, complex backgrounds, and differences in object shape, size, and orientation. The proposed system is designed to address these problems by using deep learning to identify objects and their dominant colors in still images, video streams, and live camera feeds. It uses a pretrained You Only Look Once (YOLO) detector, YOLOv8, trained on a standard benchmark dataset such as Common Objects in Context (COCO). For color analysis, the system extracts the pixel values of the detected object regions and applies preprocessing (resizing, pixel value normalization, noise reduction, and conversion from the RGB to the hue, saturation, and value (HSV) color space) to reduce sensitivity to illumination changes. The system outputs object labels with confidence scores and displays the detections with text annotations and bounding boxes. An easy-to-use graphical user interface (GUI) supports single-image, batch-image, and real-time detection. The system was demonstrated qualitatively on 15 test images; quantitative evaluation metrics were not computed.

Keywords. Object detection; color detection; convolutional neural network; YOLO; HSV; visual recognition; scene understanding

I. INTRODUCTION

Object and color detection is a basic research area in visual recognition and scene understanding. Humans distinguish objects by features such as shape, size, texture, and color, and one of the goals of computer vision is to give machines the same ability. Object detection deals with detecting and localizing objects in an image or a video frame, whereas color detection identifies and classifies the dominant color of each detected object.

Combining these techniques enables intelligent systems to perceive visual scenes more efficiently and to make meaningful decisions based on visual information. In recent years, advances in artificial intelligence (AI) and deep learning have raised the accuracy and efficiency of object detection systems to unprecedented levels. Traditional computer vision relied on handcrafted feature extraction algorithms such as the histogram of oriented gradients (HOG), the scale-invariant feature transform (SIFT), and edge detection.

Object detection advanced considerably with the introduction of convolutional neural networks (CNNs), which extract features directly from images. Earlier methods relied on manually designed features and often failed to capture complex patterns in the data. In contrast, modern ap-

proaches such as YOLO, Faster R-CNN, and SSD identify objects accurately while maintaining high processing speed, which makes real-time detection feasible.

II. LITERATURE REVIEW

The YOLO framework [1] reformulated object detection as a single regression problem, which allows it to run in real time at very high speed. YOLO showed that deep learning models can be highly accurate while offering real-time performance. The Single Shot MultiBox Detector (SSD) [2] balances detection speed and accuracy by performing object localization and classification in a single pass, and it has been used widely for real-time object detection because of its efficient processing.

Different color spaces, such as the RGB, HSV, and Lab models, have been compared for color analysis, and the HSV color space has been found to be less sensitive to changes in illumination [3]. That work underlined the effect of color space transformation on the accuracy of color detection. Color-based object tracking methods using HSV thresholding and histogram back-projection have also been proposed, showing that adding color information to object detection makes tracking more stable in dynamic environments.

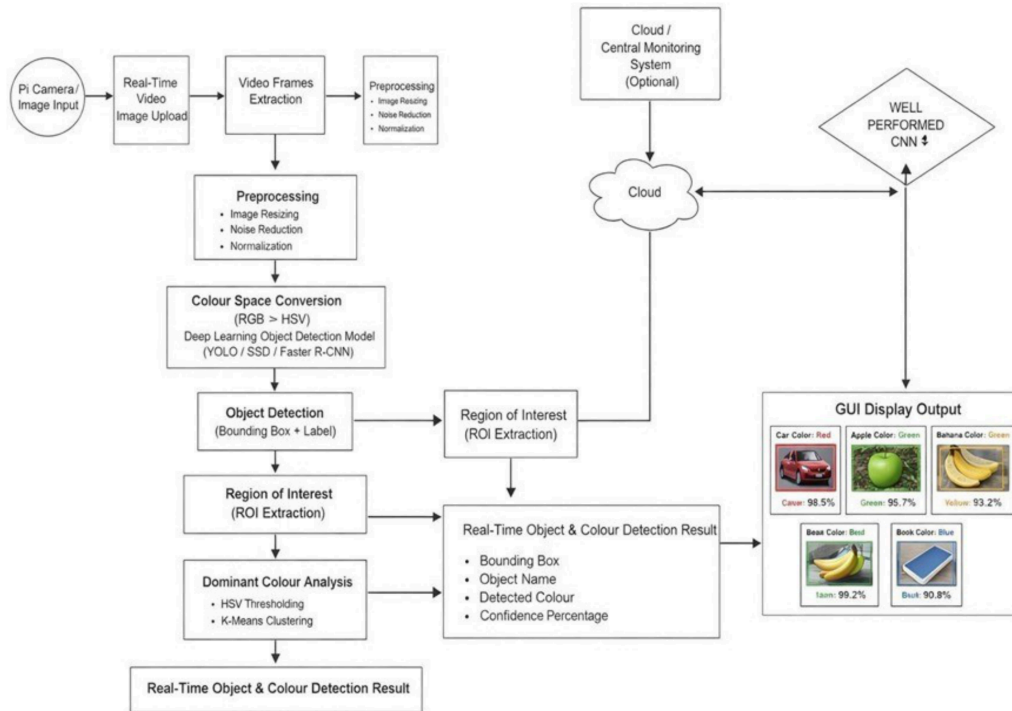


Fig. 1. Proposed block diagram for integrated object and color recognition.

The Faster R-CNN architecture of Ren and colleagues [4] builds on earlier region-based detectors and adds a region proposal network (RPN) that generates object proposals efficiently [5]. It achieved excellent results on standard datasets such as PASCAL Visual Object Classes (VOC) and MS COCO and became a foundation for contemporary object detection frameworks, improving detection accuracy substantially over earlier methods such as R-CNN and Fast R-CNN.

III. METHODOLOGY

The proposed object and color detection system was designed and developed step by step using deep learning and image processing techniques, as shown in Fig. 1. The object detection model is trained on a standard object detection dataset, such as MS COCO or PASCAL VOC, which consists of images of many object types annotated with bounding boxes. These datasets cover a variety of real-world scenarios, which allows the model to learn strong and generalized feature representations.

To improve domain-specific performance, a dataset of images related to the target application was also curated. The collected data were preprocessed by image resizing and normalization and augmented by rotation, flipping, and brightness adjustment, which add variety to the data and reduce overfitting. A pretrained YOLOv8 model was then fine-tuned on the preprocessed data by transfer learning to adapt it to the specific object classes.

After object detection, a color analysis module based on the HSV color space was applied. The detected object regions were extracted with the bounding box coordinates, and HSV thresholding was used to identify the dominant color of each object. The HSV color space has been reported to be less sensitive to lighting changes than RGB-

based techniques [3]. Combining deep learning object detection with classical image processing for color identification is intended to keep the system computationally efficient.

Standard metrics such as precision, recall, F1-score, and mean average precision (mAP) for object detection, and accuracy for color classification, apply to such a system; in this study they were not computed, and the system was demonstrated qualitatively on 15 test images.

A. Problem Statement

Reliable recognition of objects and their colors in real scenes is hindered by variations in lighting, background, object orientation, and distance. An automated system is therefore needed that can reliably detect multiple objects and determine their colors, either in real time or from static images.

B. Dataset

A self-developed dataset of 15 color images was used for the object and color measurements. The photographs were collected manually and show everyday objects such as bottles, books, chairs, and electronic devices under various lighting conditions. To allow object and color analysis, each image contains one or more objects with visible color variation, and the images vary in background, object orientation, and distance to represent real-life conditions. The images were taken with a standard camera in JPEG format, and simple annotation was performed to specify the object classes to be evaluated. Because the dataset is small, it was used mainly for testing and demonstration rather than for extensive training; it is an experimental dataset for checking how the implemented object and color detection system works.

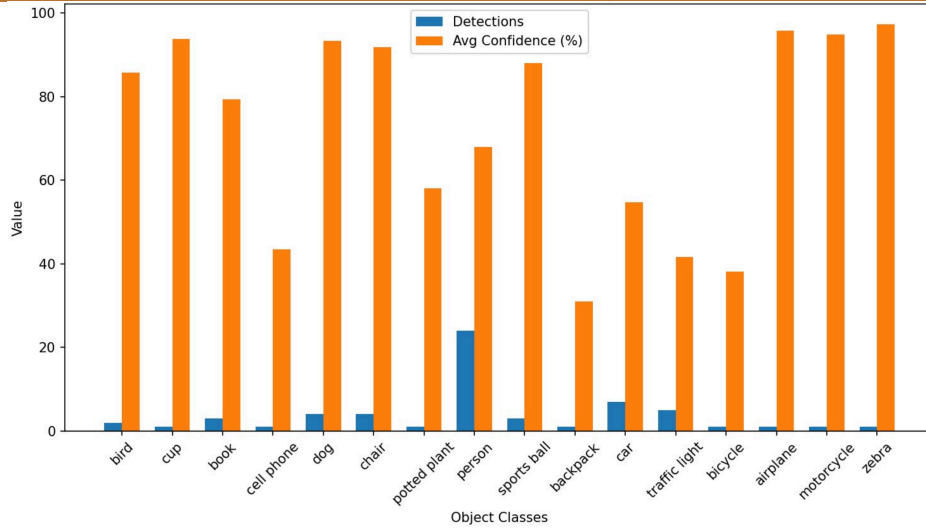


Fig. 3. Number of detections and average confidence score for each object class.

C. Training

The object detection model was built on a pretrained deep learning architecture, YOLO, that had already been trained on benchmark datasets, and it was then fine-tuned to recognize the objects in the custom dataset. This transfer learning approach made training faster and required much less computing power than training from scratch. It also allowed the model to reuse features it had already learned, such as edges and shapes.

To prepare the dataset, all images were resized to the same input size, their brightness was adjusted so that they were neither too bright nor too dark, and noise was removed to make them clearer. These steps made the images in the dataset consistent.

To identify the colors of the detected objects, the HSV color space was used. HSV is less sensitive to varying lighting conditions than the red, green, and blue (RGB) representation [3]. Used together, the object detection model and the color analysis extract both what an object is and what color it is, which is useful in applications such as security cameras, inventory tracking, and assistive tools. The system combines deep learning with classical image processing methods.

D. Mathematical Representation of Object and Color Detection

The convolution operation used for feature extraction in the object detection process is given in (1):

$$F(i, j) = \sum_m \sum_n I(i + m, j + n) \times K(m, n), \quad (1)$$

where I is the input image and K is the convolution kernel. By applying this operation, the model captures meaningful visual characteristics, including edges, textures, and other distinctive patterns in the image.

E. RGB to HSV Color Conversion

The system converts the RGB image to the HSV color space for color detection. The hue component can be expressed as in (2):

$$H = \tan^{-1} \left(\frac{\sqrt{3} (G - B)}{2R - G - B} \right), \quad (2)$$

where R , G , and B are the red, green, and blue intensity values of a pixel. Because this conversion separates chromaticity from luminance, it helps the system identify the dominant colors of the detected objects more reliably.

F. Intersection over Union

The localization performance of an object detection model can be assessed with the intersection over union (IoU) metric; IoU values were not computed in this study. The IoU is the ratio of the intersection to the union of the predicted and ground-truth bounding boxes, as given in (3):

$$\text{IoU} = \frac{\text{Area}(B_{\text{pred}} \cap B_{\text{truth}})}{\text{Area}(B_{\text{pred}} \cup B_{\text{truth}})}, \quad (3)$$

where B_{pred} and B_{truth} are the predicted and ground-truth bounding boxes. The IoU increases as object localization becomes more accurate, and it is commonly used to evaluate detection performance.

IV. RESULTS AND DISCUSSION

The real-time object and color detection system combines image-based, video-based, and live-camera object detection and color recognition in a single system with an interactive GUI. Fig. 2 shows a real-time output with bounding boxes and detected colors. In the examples shown, the model detected multiple objects and reported the dominant color and a confidence score for each. Fig. 3 shows the number of detections and the average confidence score for each object class.

These results were demonstrated qualitatively on 15 test images. The bounding box visualization shows where each detected object was localized and makes the outputs easy to inspect; because the outputs were not compared with ground-truth labels, they illustrate the behavior of the system rather than measure its accuracy.



Fig. 2. Real-time output showing bounding boxes and detected colors (faces blurred for privacy).

As Fig. 4 shows, the system detected various objects, such as birds, dogs, cars, and airplanes, and drew a bounding box around each. Each detected object is labeled with its class name, its predicted dominant color, and a confidence score. The examples illustrate multi-object detection across diverse scenes.



Fig. 4. Batch processing results for complex backgrounds (faces blurred for privacy).

The system is modular, so it can easily be scaled and integrated with other intelligent systems, and edge computing devices and cloud-based processing could further improve its performance and accessibility. Future work will incorporate instance segmentation to extract pixel-accurate colors and will extend the dataset to more challenging environments. As dataset diversity and model optimization continue to improve, the system could develop into a more adaptive and context-driven visual recognition system.

V. CONCLUSION

This study developed an intelligent visual recognition system that detects objects and their colors in real time. The YOLOv8 deep learning model was combined with HSV color space analysis for scene understanding. The system was demonstrated qualitatively on 15 test images, on which

it produced object labels, bounding boxes, dominant colors, and confidence scores; accuracy and timing were not measured quantitatively. Subject to quantitative evaluation, the method could be considered for applications such as smart surveillance, autonomous navigation, and assistive technologies, where fast decisions are essential.

ACKNOWLEDGMENT

The authors thank the faculty and staff of the Department of Computer Science and Engineering at Koneru Lakshmaiah Education Foundation, Vaddeswaram, for their technical support and for providing the computational resources required for this study.

AUTHOR CONTRIBUTIONS

All authors contributed to the conception and preparation of this work and approved the final version of the manuscript.

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

DECLARATION OF GENERATIVE AI USE

The authors declare that no generative artificial intelligence tools were used in the preparation of this manuscript. The authors take full responsibility for the content of the publication.

DATA AVAILABILITY

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- [1] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.
- [2] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg, "SSD: Single shot multibox detector," in *Comput. Vis., ECCV 2016, Lect. Notes Comput. Sci.*, vol. 9905, 2016, pp. 21-37, doi: 10.1007/978-3-319-46448-0_2.
- [3] H. D. Cheng, X. H. Jiang, Y. Sun, and Jingli Wang, "Color image segmentation: Advances and prospects," *Pattern Recognit.*, vol. 34, no. 12, pp. 2259-2281, 2001, doi: 10.1016/S0031-3203(00)00149-7.
- [4] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137-1149, 2017, doi: 10.1109/TPAMI.2016.2577031.
- [5] Srinivasa Rao Vankdoth and Michael Arock, "End-to-end deep learning pipeline for scalable, deployable object detection engine in the traffic system," *Signal Image Video Process.*, vol. 18, no. 2, pp. 1589-1600, 2024, doi: 10.1007/s11760-023-02869-5.