Khushi Sinha, 2026, 14:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

YOLOv5 and EfficientDet based Object Detection and Classification for Visually Impaired

Khushi Sinha, Esha Garg, Gaurav Singal, Preeti Kaur
(Computer Science & AI) NSUT, Delhi, India

Abstract- This research presents a novel approach that utilizes the EfficientDet model to address the challenges faced by people with visual impairments in recognizing their environment. The project focuses on the integration of state-of-the-art computer vision technology into assistive devices, specifically blind sticks, to improve the perception of the environment. The EfficientDet model, known for its superior efficiency- accuracy tradeoff, is trained to detect and classify different objects in real time providing valuable insights to users. The project contributes to the field of assistive technology by using state-of-the-art machine learning models to improve accessibility and autonomy for people with visual impairments.

Keywords— EfficientDet Model, Object Detection, Visual Impairment, Assistive Technology

I. INTRODUCTION

In the field of assistive technology, advancements in machine learning and computer vision have opened up opportunities to address the unique challenges faced by visually impaired people. This research seeks to contribute to this evolving landscape by harnessing the power of state-of-the-art object detection models, namely YOLO (You Only Look Once) and EfficientDet. The intersection of artificial intelligence with assistive devices holds promise for improving the autonomy and safety of people with visual impairments, particularly in recognizing terrains and objects in real time.

1. Domain Overview

Visual impairment poses a profound impact on a person's ability to move independently and perceive their surroundings. While traditional assistive devices provide valuable support, the need for advanced solutions that can distinguish between different terrains and objects is paramount. The use of machine learning models in assistive technologies offers the potential to improve users' awareness of their surroundings and enable them to navigate more confidently and safely.

2. Problem Description

The central challenge of this research work is to develop a robust and efficient system that is able to recognize terrain and objects in order to help visually impaired people. The aim is not only to detect obstacles and landmarks, but also to distinguish between different terrains to provide a more comprehensive understanding of the user's environment.

3. Objectives

The main objective of this research is to train and use the YOLO and EfficientDet models to develop a specialized system for visually impaired people.

4. Contributions

This research makes significant contributions to the fields of assistive technology and computer vision by integrating YOLO and EfficientDet models and addressing terrain recognition.

5. Paper Organization

This paper unfolds as follows: Section 2 provides a comprehensive review of existing literature in the domain of assistive technology, object detection, and the specific contributions of YOLO and EfficientDet models. Section 3 outlines the methodology, detailing the dataset used, model training processes, and system integration. Section 4 presents the results of the implemented system, followed by a discussion of findings in Section 5. The paper concludes with insights into potential future work in Section 6.

II. LITERARY REVIEW

1. Assistive Technology for the Visually Impaired

The integration of assistive technology has made a decisive contribution to enabling people with visual impairments to overcome obstacles and move independently in their environment. Conventional aids such as canes and guide dogs provide valuable assistance in recognising physical obstacles close to the ground. However, a comprehensive solution that extends beyond obstacle detection to include terrain and object detection is essential. These systems aim to improve the quality of life of visually impaired people by fostering a more comprehensive understanding of their surroundings.

2. Object Detection in Assistive Systems

The integration of object detection models within assistive systems marks a paradigm shift in improving the user's perception of the environment. Object detection techniques play a crucial role in the identification and categorization of elements in the visual field. While earlier approaches focused on simple obstacle detection, more recent advancements explore the integration of sophisticated models capable of recognizing a wider range of objects and terrains. Such models contribute to a more holistic understanding of the environment, enabling the user to make informed decisions in real time[1].

3. YOLO (You Only Look Once) in Object Detection

A remarkable advance in real-time object detection is the YOLO model (You Only Look Once). YOLO is characterized by its ability to process images in a single forward pass, enabling real-time detection with impressive speed and accuracy. The model divides an image into a grid and simultaneously predicts bounding boxes and class probabilities. YOLO's efficiency makes it particularly suitable for applications where low latency is critical, meeting the real-time requirements of assistive systems for visually impaired people. Existing studies demonstrate the success of YOLO in various computer vision tasks, from pedestrian detection to general object recognition[2]

4. EfIcientDet Model

Complementing the strengths of YOLO, the EfficientDet model offers a compelling solution by optimizing the efficiency- accuracy tradeoff. As part of the EfficientNet family, this model introduces compound scaling to balance depth and width, resulting in models that achieve high accuracy while maintaining computational efficiency[3].

EfficientDet has shown remarkable performance in object detection in various datasets, making it an attractive candidate for applications in assistive technology[4]. The model's ability to maintain accuracy with reduced computational overhead fits the resource constraints commonly encountered in devices for visually impaired users[5]

5. Comparative Analysis of YOLO and EfIcientDet

While YOLO and EfficientDet share the overarching goal of real-time object detection, their architectural nuances and performance characteristics distinguish them. Comparative analyses in the literature have explored the tradeoffs between speed and accuracy, resource efficiency, and adaptability to different tasks. Understanding these distinctions is crucial for selecting the most suitable model for the specific requirements of an assistive system for visually impaired individuals.

YOLOv5 achieves state-of-the-art results in the object detection field. In comparison to other Deep Learning architectures, YOLOv5 is simple and reliable. It needs much less computational power than other architectures, while keeping comparable results

6. Research Gaps and Opportunities

Despite significant progress, there remain research gaps in the application of advanced object detection models, particularly YOLO and EfficientDet, in the context of assistive technology for the visually impaired. Existing studies have predominantly focused on generic object detection tasks and there is a need for targeted research into the usability, adaptability and real-world effectiveness of these models in improving the autonomy of visually impaired people[6].

III. METHODOLOGY

1. Dataset Preparation

The foundation of our research lies in the creation of a robust and diverse dataset, which is crucial for training and evaluating the YOLO and EfficientDet models. It was important to include images from different environments, including urban and suburban environments, indoor environments and areas with varying terrain. Special care was taken to include different lighting conditions and weather scenarios.

The dataset that was utilized came from Kaggle. A pre-existing dataset with a variety of terrain types, including grassy, rocky, marshy, and sandy, was utilized. The dataset includes around 32,000 photos based on the classification project type. The photos are 256 by 256 in size.

An additional dataset obtained from Roboflow has been utilized. We made advantage of an already-existing dataset that included several kinds of environmental obstacles, including signs, trees, tracks, crash barriers, and terrain.The dataset comprised 2476 photos based on object detection. The auto orient dataset preprocessing was used, and the photos were resized to stretch to 720x720.

2. Model Selection and Training

YOLO (You Only Look Once)

The YOLO model, known for its ability to recognize objects in real time, was chosen for its efficiency and speed. The YOLOv5 architecture, an improvement over its predecessors, was chosen for its enhanced accuracy and its ability to process different categories of objects. The model was trained using transfer learning on an existing object recognition dataset and fine-tuned on our curated dataset. The training parameters, including the learning rate, the batch size, and the number of epochs, were optimized for the specific requirements of our assistive system.

YOLOv5 uses a single deep neural network to simultaneously predict bounding boxes and class probabilities for multiple objects in an image.

YOLOv5 is typically trained on large datasets with labeled bounding boxes around objects of interest. It can be fine-tuned on specific datasets to improve performance on particular types of objects or scenes. Inference with YOLOv5 is efficient and well-suited for real-time applications.

YOLOv5 is implemented using the PyTorch deep learning framework.

EfficientDet

Complementing YOLO, the EfficientDet model was chosen for its efficiency-accuracy balance. The model was trained using the TensorFlow Object Detection API. Fine-tuning occurred on our curated dataset, with special attention given to adaptability to the assistive system's real-time constraints. Hyperparameters were adjusted to strike a balance between model accuracy and computational efficiency.

EfficientDet introduces a compound scaling method that uniformly scales the resolution, depth, and width of the backbone network.

EfficientDet uses a new type of feature pyramid network called BiFPN to efficiently aggregate features at different scales.

The EfficientDet backbone is based on the EfficientNet architecture, which is designed to balance model size and performance by optimizing the depth, width, and resolution.

EfficientNet has shown success in image classification, and EfficientDet extends it for object detection.

Like many object detection algorithms, EfficientDet uses anchor boxes to predict bounding boxes.

The model predicts class probabilities and adjusts the coordinates of anchor boxes to localize objects in the image.

The model is scalable, meaning it can be adapted to different resource constraints by adjusting the scaling coefficients.

EfficientDet is typically implemented using the TensorFlow deep learning framework.

Software Implementation
The software stack for our assistive system involved the integration of several components. The EfficientDet models was deployed using TensorFlow Lite for compatibility with the Raspberry Pi's resource constraints. A Python-based implementation, leveraging OpenCV for image processing, facilitated real-time object detection.

Real-Time Object Detection Algorithm
The real-time object detection algorithm utilized by EfficientDet involves dividing the input image into a grid and predicting bounding boxes and class probabilities for each grid cell. The models employ anchor boxes to improve localization accuracy and apply non-maximum suppression to filter redundant detections.

IV. RESULT ANALYSIS

1. YOLOv5

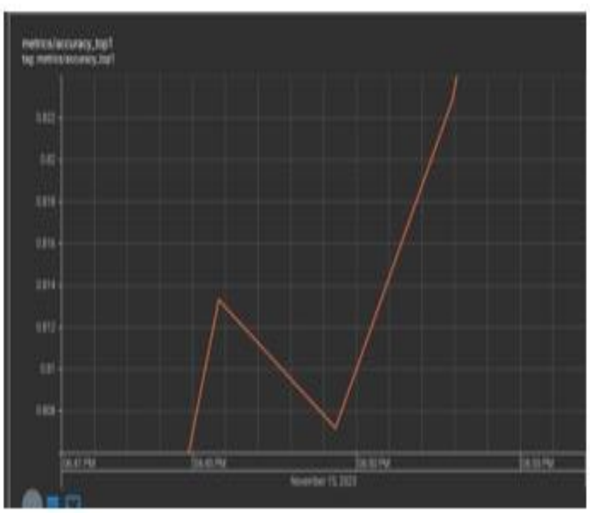

Fig 1: The YOLO model reached an accuracy of 0.89 after 5 epochs across all tested categories.

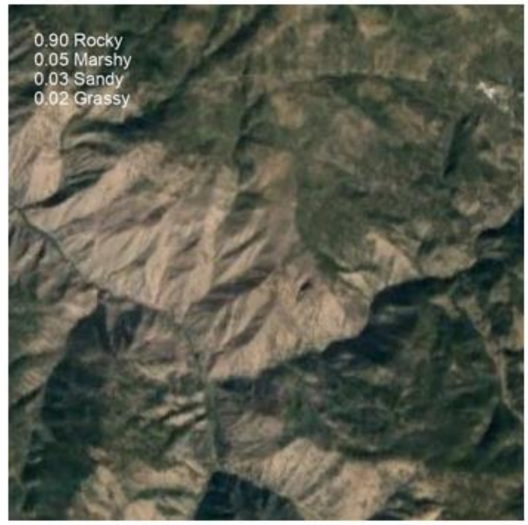

Fig 2: For individual categories, such as rocky, marshy, sandy, grassy The confidence level peaked 0.9 which further highlighted the model's effectiveness in accurate object classification.

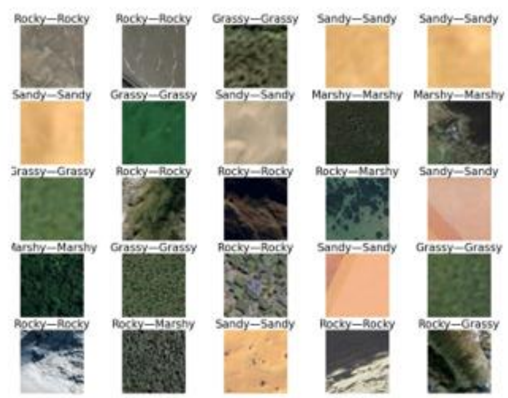

Fig 3: Overall 92% images were classified correctly.

2. EfIcientDet

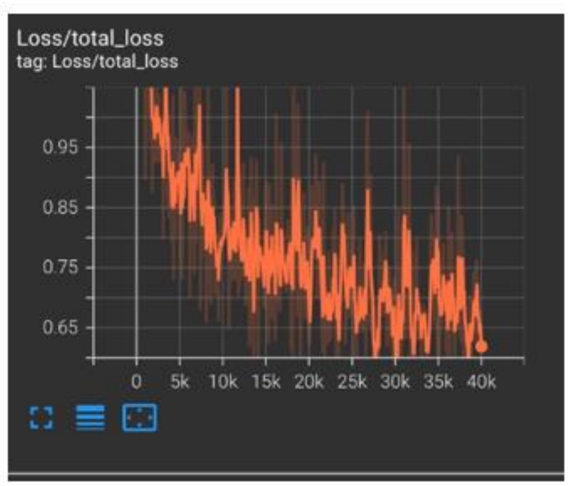

Fig 4: The total loss graph for EfIcientDet reveals an initial rapid decline, showcasing swift model adaptation to essential features. Subsequent stabilization mitigates overfitting, while consistent reductions indicate improved object classification proficiency. Individual loss components' balanced evolution contributes to a well-rounded model.

Regular monitoring guides hyperparameter optimization for optimal performance.

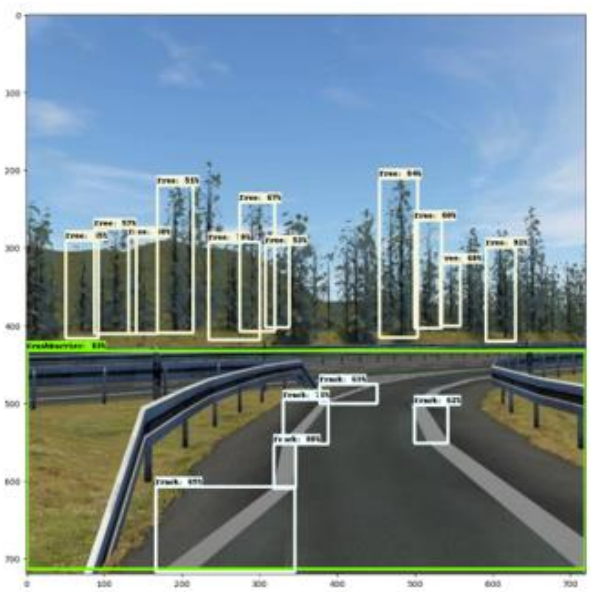

Fig 4: Object are recognized with a maximum accuracy of 91%

The effective use of bounding boxes is crucial in object detection models, serving as precise frames that encapsulate identified objects within an image. Represented by coordinates specifying the box's top-left corner and dimensions, bounding boxes not only facilitate accurate object localization but also contribute to object classification.

Limitations and Challenges

While our models demonstrate promising results, certain limitations and challenges were identified during the evaluation. Challenges include occasional misclassifications in complex environments, sensitivity to extreme lighting conditions, and constraints in processing speed on resource-constrained devices. These limitations underscore the need for ongoing refinement and adaptation to dynamic scenarios.

In summary, the YOLO and EfficientDet models prove to be effective solutions for real-time object detection in assistive systems for the visually impaired. The choice between these models depends on specific use-case requirements, considering factors such as processing speed, precision, and adaptability to dynamic environments.

V. CONCLUSION

To summarize, our research aims to address the challenges that visually impaired people face when moving independently in their environment. The integration of YOLO and EfficientDet models into an assistive system represents a significant advance in improving environmental perception. Our experiments and evaluations demonstrate the feasibility of real-time object recognition, covering different terrains and objects, using a portable and accessible hardware device.

The YOLO model, with its impressive speed and accuracy, and the EfficientDet model, which balances efficiency and precision, contribute complementary strengths to our supporting solution. By using transfer learning and fine-tuning a carefully curated dataset, we achieve models that match the real-world needs of visually impaired users.

However, there are further challenges and the success of our models in controlled environments requires further validation in diverse and dynamic environments.

Future Work

The promising outcomes of this research pave the way for future endeavors and improvements:

Continuous refinement of the YOLO and EfficientDet models is essential for broader applicability. Further fine-tuning on more extensive and diverse datasets and exploring model compression techniques can enhance both accuracy and efficiency.

Testing in dynamic and unpredictable environments is crucial for validating the robustness of the assistive system. Future work will focus on addressing challenges such as real-time adaptation to changing lighting conditions, diverse terrains, and the presence of moving objects.

Data can be improvised by adding more versatile images including aerial shots which will improve the quality of data and therefore increase the performance of the model. Data can be further annotated and the number of labels can be increased to make the model more robust.

As we move forward, the convergence of cutting-edge technologies and user-centered design will drive the evolution of assistive systems, ensuring that visually impaired individuals can navigate the world with greater independence and confidence.

Acknowledgment
The research was supported by Netaji Subhas University Technology, New Delhi.

We thank our Computer Science department for providing us with this exemplary task and motivation.

REFERENCES

1. Arduino based home automation using the internet of things (IOT). Embedded Assistive Stick for VisuallyImpaired.
2. Serrano, Kanny & Bandala, Argel. (2023). YOLO-Based Terrain Classification for UAV Safe Landing Zone Detection. 1-5. 10.1109/TENSYMP55890.2023.10223656.
3. Song S, Jing J, Huang Y, Shi M. EfficientDet for fabric defect detection based on edge computing. Journal of Engineered Fibers and Fabrics. 2021;16. doi:10.1177/15589250211008346
4. Mingxing Tan Ruoming Pang Quoc V. Le Google Research, Brain Team {tanmingxing, rpang, qvl}@google.com.arxiv.org/pdf/1911.09070.pdf
5. Naddaf-Sh, Sadra & Naddaf, Mahdi & Zargarzadeh, Hassan & Kashanipour, Amir Reza. (2020). An Efficient and Scalable Deep Learning Approach for Road Damage Detection. 10.1109/BigData50022.2020.9377751.
6. Mehla, N., Ishita, Talukdar, R., Sharma, D.K. (2024). Object Detection in Autonomous Maritime Vehicles: Comparison Between YOLO V8 and EfficientDet. In: Namasudra, S., Trivedi, M.C., Crespo, R.G., Lorenz, P. (eds) Data Science and Network Engineering. ICDSNE 2023. Lecture Notes in Networks and Systems, vol 791. Springer, Singapore. https://doi.org/10.1007/978-981-99-6755-1_10
7. Models/research/object_detection at master · tensorflow/models (no date)GitHub Available at: https://github.com/tensorflow/models/tree/master/research/object_detection?ref=blog.roboflow.com
8. Dataset: @misc{ unitysemseg_dataset,title = { UnitySemSeg Dataset },type = { Open Source Dataset },author = { Hochschule Karlsruhe },howpublished = { \url{ https://universe.roboflow.com/hochschule-karlsruhe/unitysemseg }}, url = { https://universe.roboflow.com/hochschule-karlsruhe/unitysemseg },journal = { Roboflow Universe },publisher = { Roboflow },year = { 2022 }, month = { oct }, note = { visited on 2023-11-15 },}